AL- Rafidain University College

PISSN: (1681-6870); EISSN: (2790-2293)
Journal of AL-Rafidain University College for Sciences

Available online at: https://www.jrucs.iq

JRUCS
Journal of AL-Rafidain University College for Sciences

Robust Modeling for Factors Affecting in Relative Humidity in Basra Governorate

Nazik J. Sadik	Nabaa N. Mahdi
dr.nazik@coadec.uobaghdad.edu.iq	nabaanaeemmahdi@uomustansiriyah.edu.iq
Department of Statistics - College of Administration and Economic - Baghdad University, Baghdad, Iraq	Department of Statistics - College of Administration and Economic - Mustansiriyah University, Baghdad, Iraq

Husam A. Rasheed
husamstat@uomustansiriyah.edu.iq
Department of Statistics - College of Administration and Economic - Mustansiriyah University, Baghdad, Iraq

Article Information

Article History:
Received: December, 22, 2022
Accepted: March, 3, 2023
Available Online: December, 31, 2023

Keywords:
Robust methods, least squares, relative humidity

Correspondence:
Nazik J. Sadik
dr.nazik@coadec.uobaghdad.edu.iq

Abstract

Basra Governorate is considered among the Iraqi governorates, with high levels of relative humidity sometimes, which makes the weather uncomfortable during the year. It was necessary to building in this research a robust regression model to study the effect of some factors such as the maximum and minimum temperatures, the amount of rainfall, atmospheric pressure, and the average amount of dust falling. When building linear regression models, many problems occur in which failure some of the hypotheses model. In this research, least squares estimates were compared with robust and found the preference for the method of S.

https://doi.org/10.55562/jrucs.v54i1.618

1. Introduction

One of the most important statistical methods is modeling different phenomena with models that are closer to reality, and these models must have efficient capabilities that correctly express the studied phenomenon so that we can predict and develop future plans and build strategies, Either they are probabilistic, such as time series models, Markov chains, and reliability models, or they are causal, meaning that the construction of these models is a cause and effect of the cause, and this is what is called regression models that are built on a relationship between the explanatory variables or what is called explanatory and the dependent variable or the response variable.

The German mathematician (Carl Friedrich Gauss) introduced the Ordinary Least Squares method which is the best unbiased linear estimation (BLUE) under the basic assumptions, Where the random errors are independent and normally distributed with the least variance among all other linear estimators, and the presence of any problem such as the presence of outliers or extreme values would make them less accurate for the deviation of the distribution and make it twisted, and

this makes these estimators inefficient, so we resort to other estimation methods that are more flexible, which are the robust estimators. Due to the existence of such variables in our practical life, this problem has received great attention, and the reason is the realization that the use of traditional methods in estimating parameters leads to the emergence of the problem and the data is not fully subject to the assumptions made. It leads to incorrect and unsatisfactory results in estimation and prediction.

2. Problem of Search

Climate change in the world in general and in Iraq in particular is caused by the presence of observations or data that contain abnormal values, or that the distribution of errors does not follow the normal distribution, and thus there will be a breach in the assumptions of the model that was built for the purpose of predictive estimation, and therefore correct data must be dealt with and analyzed In order to obtain forecasts or estimates.

3. Research objective

The research aims to find the best robust estimator for data that penetrates the method of least squares in finding the estimators of the model under study. These methods are the method of maximum likelihood or what is called the (M) method, the (MM) method, and the (S) method. A comparison was made between the three methods using the mean error standard (MSE).

4. Robust Regression [1,2,3]

The robust regression is used to estimate the parameters of the regression model when the data suffers from the problem of the presence of outliers or extreme values, so the traditional estimation methods become useless, and we resort to estimating the model parameters by alternative methods that are resistant to addressing these problems. The efficiency of the traditional methods is equivalent to the usual least squares method, and it is more efficient than it in the event of a problem.

Robust Method

There are many robust methods that have been suggested by researchers, and among those methods we will present the following:

M Method

This method is considered one of the most widely used methods of robust regression. It was proposed by the scientist Hooper in 1973. It is not affected by outliers, in addition to its ease of theoretical use. It is built on the basis of:

1- Minimize the objective function instead of decreasing the sum of squares of the error of the objective function [1, 4], as in the following formula:

$$\min \sum_{i=1}^{n} p\left(\frac{e_i}{s}\right) \tag{1}$$

2- The error can be found through the following formula:

$$e_i = y_i - x_i'\beta \tag{2}$$

3- Find the standard deviation according to the following formula:

$$s = 1.483 MAD \tag{3}$$

Where, the median of the absolute deviation is calculated according to the formula:

$$MAD = median\left|e_i - median(e_i)\right| \tag{4}$$

4- Obtain the parameters of the model, it is by taking the partial derivatives with respect to the parameters and equating them to zero, and we get this equation:

$$\sum_{i=1}^{n} \psi\left(\frac{y_i - x'\hat{\beta}}{s}\right) x_i = 0 \tag{5}$$

The selection of the function ψ depends on the preference of the amount of weight to determine the outliers, and this function is characterized by the following characteristics:

$$p(e) \geq 0,\ p(0) = 0,\ p(e) = p(-e),\ p(e_i) \geq p(e_i')\ for |e_i| \geq |e_i'|$$

This means that the probability of the zero value of the error is equal to zero, and that the positive and negative values have the same probability. The probability of the original function of error is greater than the probability of the derivative of the error, and by defining the weight function, the equation can be solved: [5, 7]

$$w(e_i) = \frac{\psi\left(\frac{y_i - x_i'\hat{\beta}}{s}\right)}{\left(\frac{y_i - x_i'\hat{\beta}}{s}\right)} \quad (6)$$

Equation (5) can be written in the following form:

$$\sum_{i=1}^{n} w\left(\frac{y_i - x'\hat{\beta}}{s}\right) x_i = 0 \quad (7)$$

We can write the above formula in matrix form:

$$X'WX\hat{\beta}_M = X'WY \quad (8)$$

The estimator for M can be obtained according to the following formula:

$$\hat{\beta}_M = (X'WX)^{-1} X'WY \quad (9)$$

Where W is the matrix of weights

MM Method [4]

This method depends on the use of the robust (S) estimation by minimizing the residual measure of the (M) method, so it is considered a follower of the (M) estimation method. It has high efficiency in the case of normal distribution of errors with a high breaking point. This estimator can be obtained by following the following steps:[9,8]

1- Determine an initial estimator with a high breakdown point, and it is not necessarily efficient, and through it we calculate the initial residuals according to the following formula:

$$r_i(\hat{\beta}_s) = y_i - x_i'\hat{\beta}_s \qquad ,1 < i < n \quad (10)$$

2- Calculate the estimator (M) of the (measurement (Sn) for the residuals depending on the estimated (M) equation for the measurement parameter and according to the following formula:

$$\left(\sum_{i=1}^{n} P\ (r_i(\hat{\beta}_s)/s)\right)/n = 1/2 \quad (11)$$

Whereas:

$$S_n = S\left(r_1(\hat{\beta}_s), \ldots\ldots, r_n(\hat{\beta}_s)\right) \quad (12)$$

3- The estimator MM can be defined as the estimator M of β using a function (re-descending)

$$\psi(u) = \frac{\partial p_1(u)}{\partial u} \quad (13)$$

4- By solving the following equation, we will get the MM estimator.

$$\sum_{i=1}^{n} X_{ij}\, \psi_1\left(\frac{y_i - x'\beta}{s_n}\right) = 0 \quad (14)$$

And the measurement estimate can be obtained from the second step, as:

$$r_i(\beta) = y_i - x_i'\beta \quad (15)$$

S method

One of the disadvantages of the M method is the loss of the data distribution as well as the probability function of the data due to the use of the median only as a weight value. The S method is based on the estimation of the residuals of the M method, and the standard deviation of the residuals is used in order to get rid of the weaknesses. The S estimator can be defined according to the following formula [8][10]:

$$\hat{\beta}_s = \widehat{min_\beta}\, \sigma_s\,(e_1, e_2, \ldots\ldots, e_n) \tag{16}$$

and is determined $\hat{\sigma}_s$ minimum robust scale estimator according to the following equation:

$$min_\beta \sum_{i=1}^{n} P\left(\frac{y_i - \sum_{i=0}^{k} x_{ij}\,\beta_j}{\hat{\sigma}_s}\right) \tag{17}$$

$\hat{\sigma}_s$ can be calculated according to the following formula:

$$\hat{\sigma}_s = \sqrt{\frac{1}{0.199\ n}\sum_{i=1}^{n} w_i e_i^2} \quad , \qquad w_i = w_s = \frac{p(u_i)}{{u_i}^2} \tag{18}$$

As for how to obtain the initial estimate, we depend on equation (4) and divide by (0.6745). Thus, the solution can be obtained by substituting the parameters according to the following:

$$\sum_{i-1}^{n} x_{ij}\,\psi\left(\frac{y_i - \sum_{i-0}^{k} x_{ij}\beta_j}{\hat{\sigma}_s}\right) = 0 \qquad j = 0,1,\ldots,k \tag{19}$$

Since ψ is a function derived from P:

$$\psi(u_i) = P'(u_i) = \begin{cases} u_i(1-(u_i/c)^2)^2 & |u_i| \le c \\ 0 & |u_i| > c \end{cases}$$

5. The Practical side

Data on the weather conditions of Basra Governorate for the period from (1-2015 to 12-2020) were collected from the Central Statistical Organization website, where the independent variables represent, Monthly Average Precipitation (MAP), General Average Atmospheric Pressure (GAP), and General Average Normal Temperature (GANOT), while the dependent variable represents the General Average Relative Humidity (GARH).

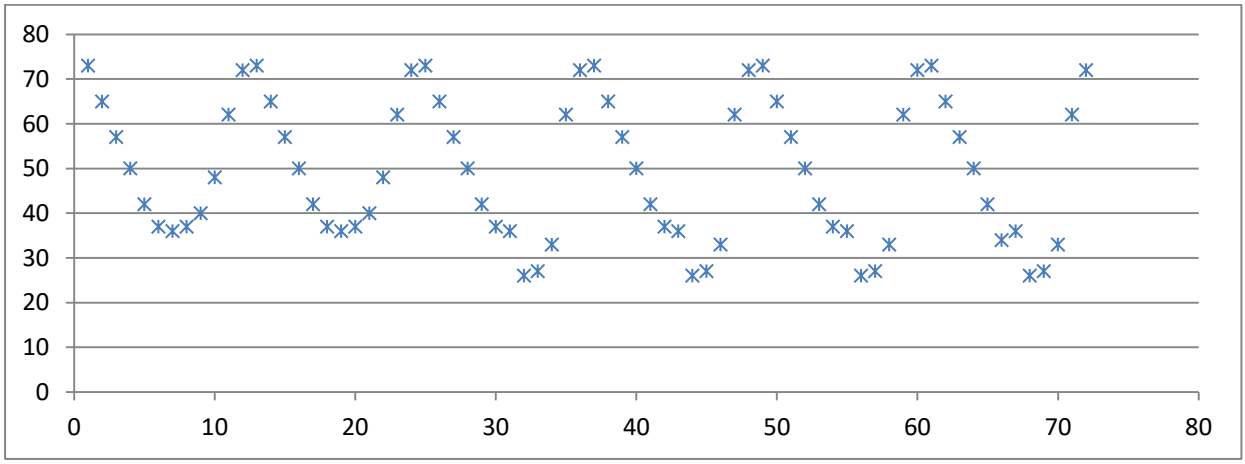

Fig. (1): General Average relative humidity (2015-2020)

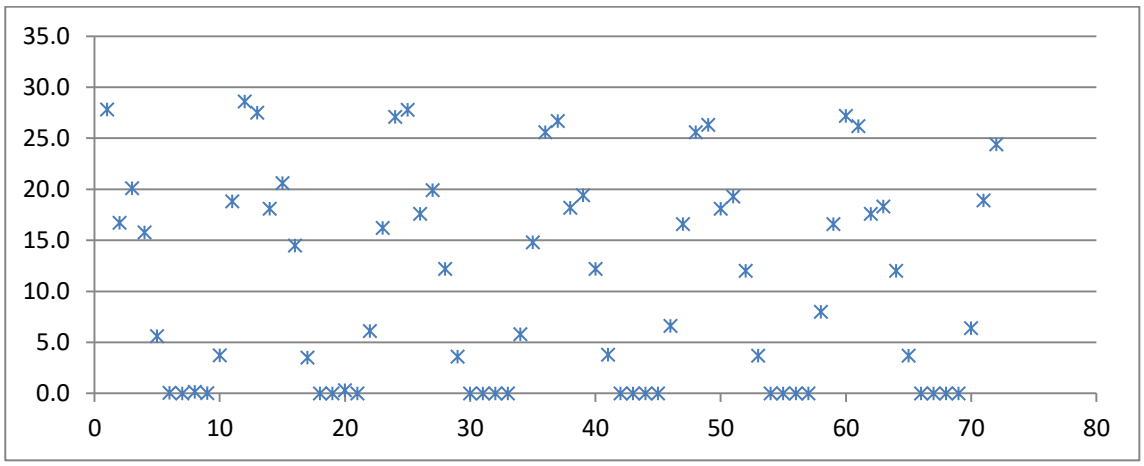

Fig. (2): Monthly Average precipitation (millimeters) (2015-2020)

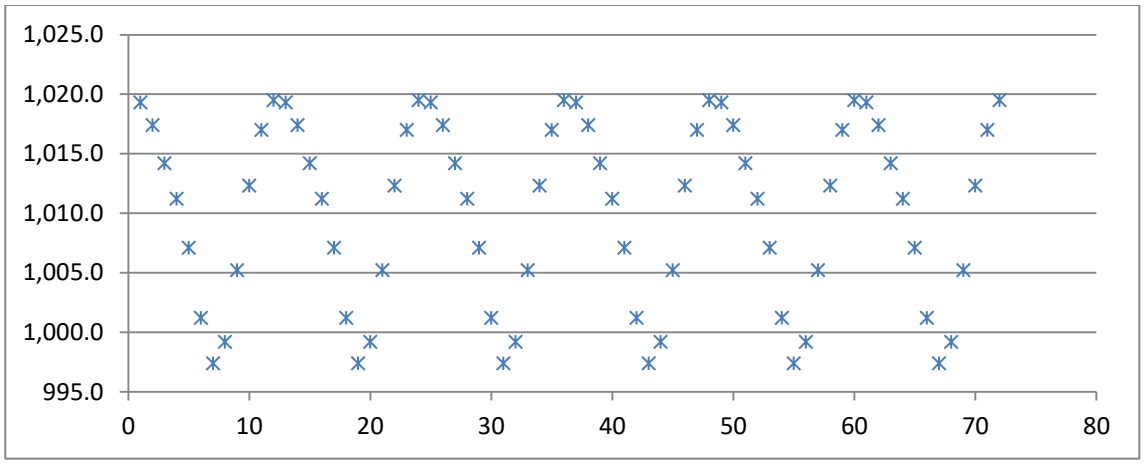

Fig. (3): General Average atmospheric pressure (millibars) (2015-2020)

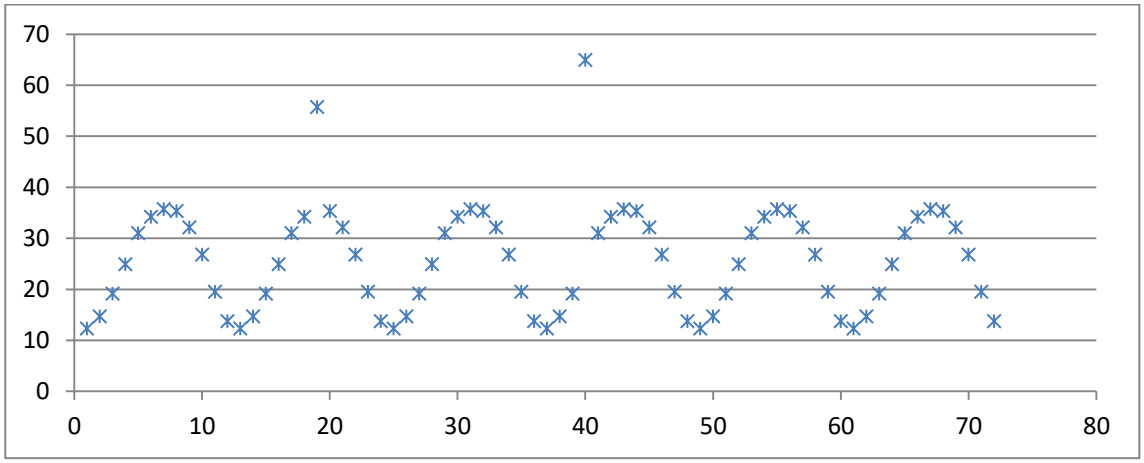

Fig. (4): General Average normal temperatures (2015-2020)

• Testing outliers in the data

The data was tested using the box-plot method, and it was proved that there are abnormal values in the model, as shown in the following figures:

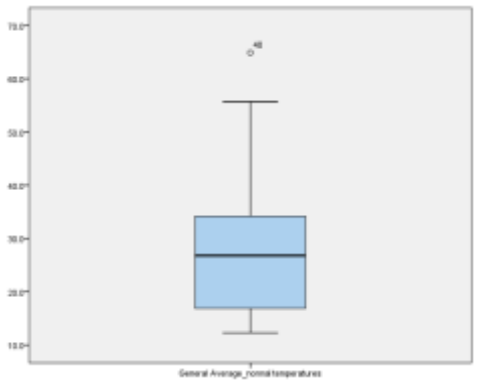

Fig. (5): box-plot General Average normal temperatures

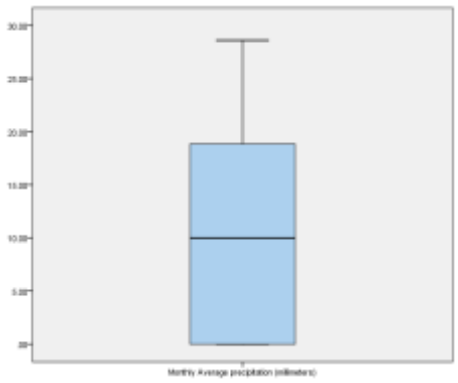

Fig. (6): box-plot of Monthly Average precipitation (millimeters)

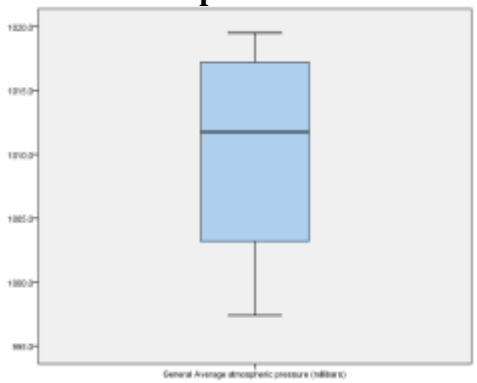

Fig. (7): box-plot General Average atmospheric pressure (millibars)

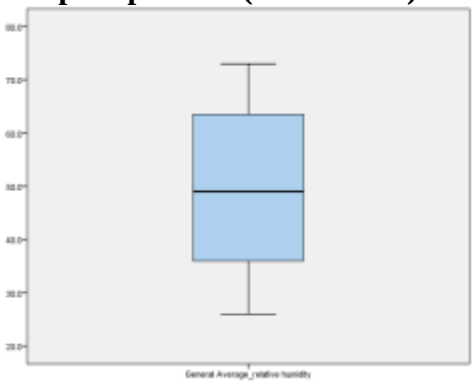

Fig. (8): box-plot General Average relative humidity

• Results of using the robust M method the weight function (Bisquare) and the function (Huber) were used, and the results were as shown in Table (1).

Table (1): estimations of the robust M method

Variable	Coefficient	Std. Error	z-Statistic	Prob.
MAP	0.314595	0.114946	2.736894	0.0062
GAP	0.076080	0.004518	16.83982	0.0000
GANOT	-1.153488	0.133406	-8.646418	0.0000

We note from the above table that there is a direct effect of each of the precipitation (MAP) and atmospheric pressure (GAP) on the general average relative humidity (GARH), while there was an inverse effect of temperature (GANOT) on the relative humidity (GARH), i.e. the greater (GANOT) increased, so (GARH) decreased, for Basra Governorate, where we note from the value of R2, that the explanatory variables explain 65.9% of the dependent variable (GARH).

• Results of using the robust S method

Through the application of the robust S method and as shown in Table (2), the results showed that there is a significant effect of the studied variables on the variable (GARH) of Basra Governorate, where we notice that the higher the MAP, GAP and the lower the GANOT value, the higher the GARH value of Basra Governorate. We notice from the value of R2 that the independent variables explain 91.6% of the effect of the dependent variable, GARH.

Table (2): Parameter of the robust S method

Variable	Coefficient	Std. Error	z-Statistic	Prob.
MAP	0.609024	0.046005	13.23821	0.0000
GAP	0.065559	0.001808	36.25653	0.0000
GANOT	-0.827979	0.053393	-15.50713	0.0000

• Results of the robust MM method The method proved a significant effect of the study variables on the relative humidity (GARH) in Basra Governorate, and the results were similar to the two robust

methods (M, S), that is, there is a direct effect of the two variables (MAP, GAP) and an inverse effect of the variable (GANOT), and the value of R^2 shows, That all the independent variables explain 66.4% of the dependent variable, as shown in Table (3).

Table (3): Estimations of the robust (MM) method

Variable	Coefficient	Std. Error	z-Statistic	Prob.
MAP	0.389142	0.102638	3.791419	0.0001
GAP	0.073504	0.004034	18.22075	0.0000
GANOT	-1.073507	0.119121	-9.011886	0.0000

Accordingly, we can choose the best way to represent the impeccable linear regression model by choosing the largest value of the coefficient of determination R^2, where the S method was the best way to represent the model, because it has the largest coefficient of determination R^2, as shown in Table (4).

Table (4): Comparison criteria

Methods	R^2
M	65.9%
S	91.6%
MM	66.4%

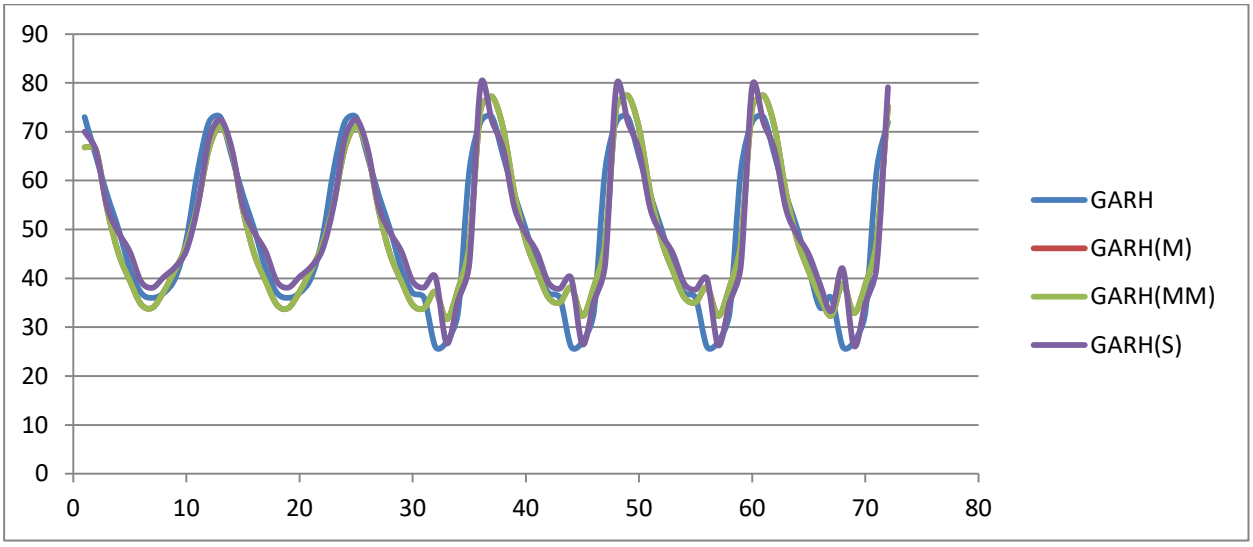

Fig. (9): Predicted Values for General Average relative humidity (2015-2020)

6. Conclusions and recommendations

1. There is a strong linear regression relationship between MAP, GAP and the dependent variable GARH.
2. There is a strong and inverse linear regression relationship between the variable GANOT and the dependent variable GARH in Basra Governorate.
3. The best way to estimate the model in the presence of outliers is the robust S method.
4. We recommend studying other variables that affect the relative humidity (GARH) of Basra Governorate.
5. We recommend resorting to the use of renewable energy (solar energy) in the generation of electrical energy to address the increase in humidity in Basra Governorate.

References

[1] Hassan, Arshad Hameed and Ahmed, Shaimaa Muhammad, "Comparison of some robust estimates for estimating the parameters of the multiple linear regression models in the presence of high inflection points using simulation." Journal of Business Economics Issue (Special - Part 1) September, 2021 Pages: 103-89

[2] Hassan, Sabah Haseeb, "Choosing a robust model in linear regression," College of Science - University of Kirkuk, Scientific Studies, Volume (2) - Issue (1), 2007.

[3] Hassan, Tara Ahmed and Reda, Mahdi Saber, "Using robust regression to find the most appropriate model to represent meteorological data in the city of Erbil during the period (1998-2010)", Journal of the College of Administration and Economics, 34th year, Issue (89), 2011.

[4] Ali, Saif Hussein, "Choice the best estimations of parameters robust regression model with practical application", M. Sc. Thesis, College of Administration and Economics, University of Karbala, 2019.

[5] Shaker, Salih Moayed, "Improvement of the M-Robust Method in Estimating the Parameters of the Linear Regression Model", Iraqi Journal of Statistical Sciences (16) 2009 P.P [242-219]

[6] N. R. Draper and H. Smith, Applied Regression Analysis, 3rd Edition, Wiley Interscience Publication, United States, 1998

[7] Y. Susanti and H. Pratiwi, "Robust Regression Model for Predicting the Soybean Production in Indonesia", Canadian journal on Scientific and Industrial Research, 2, No.9 (2011) , 318-328.

[8] Yuliana, S. Hasih, P. Sri Sulistijowati. H & Twenty, L."M Estimation, S Estimation and MM Estimation in Robust Regression", International Journal of Pure and Applied Mathematics, 91(3), 2014.

[9] C. Chen, "Robust Regression and Outlier Detection with the ROBUSTREG Procedure", Statistics and Data Analysis, paper265-27. SAS Institute Inc., Cary. NC.

[10] David M. Rocke, (1996),"Robustness Properties of S-estimators of Multivariate Location and shape in high dimension", Ann. Statist, Vol. 24, No. 3, (1996).

AL- Rafidain University College

PISSN: (1681-6870); EISSN: (2790-2293)

مجلة كلية الرافدين الجامعة للعلوم

Available online at: https://www.jrucs.iq

JRUCS

Journal of AL-Rafidain University College for Sciences

النمذجة الحصينة للعوامل المؤثرة في الرطوبة النسبية في محافظة البصرة

نازك جعفر صادق	نبأ نعيم مهدي
dr.nazik@coadec.uobaghdad.edu.iq	nabaanaeemmahdi@uomustansiriyah.edu.iq
قسم الاحصاء - كلية الادارة والاقتصاد - جامعة بغداد، بغداد، العراق	قسم الاحصاء - كلية الادارة والاقتصاد - الجامعة المستنصرية، بغداد، العراق

حسام عبد الرزاق رشيد

husamstat@uomustansiriyah.edu.iq

قسم الاحصاء - كلية الادارة والاقتصاد - الجامعة المستنصرية، بغداد، العراق

المستخلص

تعتبر محافظة البصرة من المحافظات العراقية ذات معدلات الرطوبة النسبية العالية ، مما يجعل الطقس في بعض الأحيان غير مريح خلال العام. تم في هذا البحث بناء نموذج انحداري حصصين لدراسة تأثير بعض العوامل مثل درجات الحرارة العظمى والصغرى، وكمية الأمطار، والضغط الجوي، ومتوسط كمية الغبار المتساقط. عند بناء نماذج الانحدار الخطي تحدث العديد من المشاكل التي تؤدي الى فشل بعض نماذج الفرضيات. في هذا البحث تمت مقارنة تقديرات المربعات الصغرى مع الحصينة ووجدت الأفضلية لطريقة S.

معلومات البحث

تواريخ البحث:

تاريخ تقديم البحث: 2022/12/22

تاريخ قبول البحث: 2023/3/3

تاريخ رفع البحث على الموقع: 2023/12/31

الكلمات المفتاحية:

طرق قوية، المربعات الصغرى، الرطوبة النسبية

للمراسلة:

نازك جعفر صادق

dr.nazik@coadec.uobaghdad.edu.iq

https://doi.org/10.55562/jrucs.v54i1.618