A Try To Building The Best Linear Regression Model For Prediction And Controlling The Consumption Of Electric Energy In Baghdad

Assistant prof. Dr. Sabah F.Abdul Hussein

Lecturer Manar M.Rashid

Abstract:

This paper tries to build the best linear regression model (BLRM) using Least square method (L.S.M) for data taken from random sample of families in Baghdad to predict and control the local consumption of electric energy. To achieve that aim it has depended on the examination of residuals of linear models. It's used "SPSS system" for the following:

- Detect the outliers and the influential observations of them and also the multicollinearity problem.
- Meet the usual assumptions about the errors (UAE)
- Find the mean square errors (MSE) and the mean square predicted errors (MSPE) as criteria to arrive at BLRM.

Key-words: BLRM, UAE, Outliers, multicollinearity, MSE, MSPE

المستخلص:

حاول الباحثان في هذه الورقة بناء افضل نموذج إنحدار خطي (BLRM) بطريقة المربعات الصغرى للتنبؤ والسيطرة على الاستهلاك المحلي للطاقة الكهربائية باستخدام بيانات عينه عشوائية من العوائل العراقية في احدى مناطق بغداد. ولتحقيق هذا الهدف اعتمد البحث على فحص البواقي لنماذج الانحدار الخطية باستخدام نظام التحليل الاحصائيِ SPSS وذلك للتأكد مما يلي:

- الكشف عن القيم المتطرفة والمشاهدات المؤثرة منها وكذلك مشكلة التعدد الخطي.
- تلبية الفرضيات الاعتيادية حول حد الخطأ في النموذج (UAE).
- حساب قيمة كل من متوسط مربعات البواقي "MSE" ومتوسط مربعات اخطاء التنبؤ MSPE وعدهما معيارين للوصول الى النموذج الخطي الافضل.

الكلمات المفتاحية: نموذج الانحدار الخطي الافضل، الفرضيات الاعتيادية حول حد الخطأ، المشاهدات المتطرفه والمؤثرة، التعدد الخطي، متوسط مربعات البواقي، متوسط مربعات اخطاء التنبؤ.

1- The preface and the aim:

Iraqis suffer from the lack of electric energy. Although much money is spent to improve it after 2003, there is no real difference. The problem is that there is no plan to determine the factors which control and predict the need of electric energy. So the Iraqi minister of oil said "There is always a gab between the produced electric energy and the local consumption".[عبد المهدي، 2015] This paper has determined two real factors which are: the sizes of the families and their incomes. It has also added the number of rooms of the family's house as a third factor but it is dropped from the final linear model because it is high correlated with the family size and it provides the same information such as the size of family variable. In fact the family size and the number of rooms are two faces of one coin, so they are regarded as one factor. As we shall see later the linear regression model has improved after dropping the third factor and according to the Mallow's statistic "CP", the bias of dropping that factor is very small. In spite of increasing the total produced electric energy from (34670328) to (46064647) MW/H in the period (2002-2009) [الجهاز المركزي للاحصاء 2010-2011], it doesn't meet the need of the citizens because there is no plan to connect between the demand of electric

energy consumption and related significant factors that can be displayed by regression equation to show the marginal propensity consumption (MPC) of the electric energy.
The aim of this paper is to build the (BLRM) using (L.S.M) for data taken from random sample of families in Baghdad to predict and control the local consumption of electric energy. To accomplish this aim, the paper is divided into two main parts; the theoretical part which involves the detection of outliers and multicollinearity, the "UAE" and the criteria "MSE and MSPE" and the applied part which involves the data of simple random sample and using "SPSS system" for creating linear regression models and examining their errors (Residuals) in order to accomplish the (BLRM).

2- **The theoretical part:** Steps of building BLRM using L.S.M

2-1. Define the problem: the problem is that how the increasing need of electricity in Iraq can be met and determined (controlled) such that no gab will be between the produced electricity and the need of it.
2-2. Determine the aim: the aim is to building the BLRM that can predict and control the increasing need of electric energy in Baghdad.
2-3. Choose the variables: The dependent variable (Y) and the predictors (X_i), i=1,2,3, …… that are basic and available.
2-4. Collect data about the variables using astatistical method.
2-5. Regress (Y) on (X_i) using astatistical package like "SPSS" and notice the following:

2-5-1. The outliers and the influential observations.

Since we usually assume that $ei \sim N(o, \sigma^2)$ and $S^2 = \frac{\Sigma e_i^2}{n-p}$ where $E(e_i)= \bar{e}$=o, Then $e_i/s \sim N(o,1)$ and also since 95% of N(0,1) distribution lies between (-1.96, 1.96) then we can expect approximately that 95% of ei/s were between the limit (-2,2) and that is, out of this limit are regarded outliers. But the outliers are not necessary to be influential observations in fitting the chosen model, so they must be tested by cook statistic (D) where:[Draper & Smith,1999]

$$Di=\left\{ \frac{ei}{s(1-rii)^{\frac{1}{2}}} \right\}^2 \left\{ \frac{rii}{1-rii} \right\}^{\frac{1}{P}}$$

ei: The i th residual when the full data set is used (i= 1, 2, ….n).

S^2: The estimate of the variance σ^2 provided by MSE.
rii: The i th diagonal entry of the hat matrix; $H=X(\acute{X}X)^{-1}\acute{X}$
p: The no. of the parameters to be estimated.
Then
$Di > F[p, n-p, 1-\propto]$ ⇒ The ith observation is influential
O.W ⇒ The outlier can be omitted

2-5-2. The linear relation between (Y) and each (x_i) and the multi collinearity problem by making sure of the following indicators:

2-5-2-1. The augmented simple linear correlation symmetric matrix

$$\begin{pmatrix} 1 & r12 & r13 & . & . & r1k & . & r1y \\ r21 & . & r23 & . & . & r2k & . & r2y \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ rk1 & rk2 & rk3 & . & . & 1 & . & rky \end{pmatrix}$$

each riy should approach to one not to zero

each rij should approach to zero not to one (j=1,2,….k)

2-5-2-2. The variance inflation factor (VIF) should not be more than five[Haan,2002], That is R_i^2 (The coefficient of X_i determination) should not be more than (0.80) because:

$$\text{VIFi} = \frac{1}{1-R_i^2} \Rightarrow R_i^2 = 1 - \frac{1}{VIF(i)} \quad , \quad Then\ for\ VIF \leq 5$$

$$R_i^2 \leq 1 - \frac{1}{5} \quad \Rightarrow \quad R_i^2 \leq 0.80$$

2-5-2-3. The stability and the reason of the coefficients. That is, the simple change in data should not make a dramatic changes in coefficients and also do not have the incorrect signs.

Another way to determine the severity of the multicollinearity and diagnose the causing variable is by using the tests of Farrar-Glauber:[بخيت وسمر ,2009]

First - X^2 test for showing the existence and the severity of multicollinearity:

$$X^2 = -\,[n-1-\tfrac{1}{6}(2k+5)].\ \text{Ln}\,|R| \quad \text{where:}$$

n: sample size

k: no. of predictors

Ln $|R|$ = Logarithm of the determination of simple correlation matrix among predictors.

X^2 computed > X^2 tabled $[\propto, k(k-1)/2] \Rightarrow$ multicollinearity exists

O.W $\Rightarrow$ no multicollinearity

The severity of the multicollinearity depends on howmuch X^2 comp. is bigger than X^2 tab., That is, if X^2 comp. is not so bigger than X^2 tab, the researcher can ignore it.

Second- F and T tests to determine the causing variable of the problem:

$$Fj = \frac{\left(R^2_{xi.x1,x2,....xk}\right) \ / \ (k-1)}{(1-R^2_{xi}, x1, x2, \dots xk)/(n-k)} \quad ; j = 1, 2, \dots \dots, k$$

Fj (comp.) > Ftab. ($\propto$, k-2, n-k-1) ⇒ Xj correlated with other variables
O.W ⇒ Xj not correlated

$$tij = \frac{(r_{xixj.x1x2.....,xk})\sqrt{n-k}}{\sqrt{1-(r^2_{xixj.x1x2,.....,xk})}}$$

Where rx_ix_j denotes to partial correlation coefficient between X_i & x_j

tij (comp.) > tab. ($\propto$, n-k) ⇒ The partial correlation is significant
O.W ⇒ Partial correlation is not

If xi is the causing variable of the problem and there is another variable in the model provides the same information then it is better to omit the causing variable. Otherwise we should use alternative method to estimate the parameters of the model

2-5-3- Make sure of UAE

The UAE as Ostrom has determined, besides the linear relation are the following:[Ostrom, 1990]

- Non stochastic (X_i); that is $E(e_ix_i)=0$
- $E(e_i)=0$. That is the linear regression equation provides the expected value of the dependent variable ($\hat{Y}$).
- $V(e_i)= E(e_i^2) = \sigma^2$ for each value ⇒ i.e Non heteroscedasticity.
- $Cov(e_i,e_j) = E(e_i,e_j) = o\forall\ i \neq j$ i.e Non autoregression .
- Normality of error-term distribution in order to make F & T- Tests.

Usually we indicate the above UAE by $e_i \sim iid\ N(o, \sigma^2)$ and the third and fourth conditions above by $V(\text{ee}) = \sigma^2 In$ because:

$$V(\text{ee}) = E(\text{ee}) = E \begin{bmatrix} e_1 \\ e_2 \\ . \\ . \\ e_n \end{bmatrix} [e_1, e_2, \dots\dots, e_x] = E \begin{bmatrix} e_i^2 & e_1\ e_2 & e_1\ e_n \\ e_2\ e_1 & e_2^2 & e_2\ e_n \\ \vdots & \ddots & \vdots \\ e_n\ e_1 & e_n\ e_2\ \cdots & e_n^2 \end{bmatrix}$$

$$= \begin{bmatrix} Ee_i^2 & E(e_1\ e_2) \dots E(e_1\ e_n) \\ E(e_2\ e_1) & Ee_2^2 \dots . E(e_2\ e_n) \\ & . \\ & . \\ E(e_n\ e_1) & E(e_n\ e_2) \ \dots \dots E(e_n^2) \end{bmatrix} = \begin{bmatrix} var(e_1) & cov(e_1\ e_2) \dots cov(e_1\ e_n) \\ cov(e_2\ e_1) & v(e2) \dots . cov(e_2\ e_n) \\ & . \\ & . \\ cov(e_n\ e_1) & cov(e_n\ e_2) \ \dots var(e_n^2) \end{bmatrix}$$

- Var (ei) = $E(e_i^2) = \sigma^2$ and cov(eiej) = E(eiej) =0

- $V(eé) = \begin{bmatrix} \sigma^2 & o & . & . & o \\ o & \sigma^2 & . & . & o \\ . & . & . & . & . \\ . & . & . & . & . \\ o & o & . & . & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & o & . & . & o \\ o & 1 & . & . & o \\ . & . & . & . & . \\ . & . & . & . & . \\ o & o & . & . & 1 \end{bmatrix}$

Each one of the above assumptions should be checked and if it is violated it should be treated. For example if cov(eiej) ≠0 then there is autocorrelation problem and to getrid of this problem we have to transform the variables of the model such that they become without autocorrelation as following

$Y_t^* = Y_t - \hat{p}\, Y_{t-1}$ and $X_t = X_t - \hat{p} X_{t-1}$ where:

$\hat{p} = \frac{\sum_{t=2}^{n} etet-1}{\sum_{t=2}^{n} e_t^2 - 1}$ (t = 1, 2, 3, …….n) The estimated value of outocorrelation

Y_t^*, X_t^* : The new variables of Y_t and X_t respectively

Durbin – Watson statistic (D.W) is a good indicator about $\hat{p}$ where:

D.W= 2(1-$\hat{p}$) ⇒ $\hat{p}$ = 1- $\frac{D,W}{2}$ and since $\hat{p}$ = [-1, 1] Then

D.W= 2(1-O) = 2 if $\hat{p}$ = o ⇒ no autocorrelation.

D.W= 2(1-1) = 0 if $\hat{p}$ = 1 ⇒ Posative autocorrelation

D.W= 2[1-(-1)] = 4 if $\hat{p}$ = -1 ⇒ Negative autocorrelation

So (D.W) value approaches to (2) indicates no autocorrelation. However, Durbin & Watson has made tables to test the serial correlation in least square reg.

2-5-4: Make sure of the validation of the final model; the stable and reasonable estimators and the small MSE and MSPE.

All these steps can be summarized by the following figure:

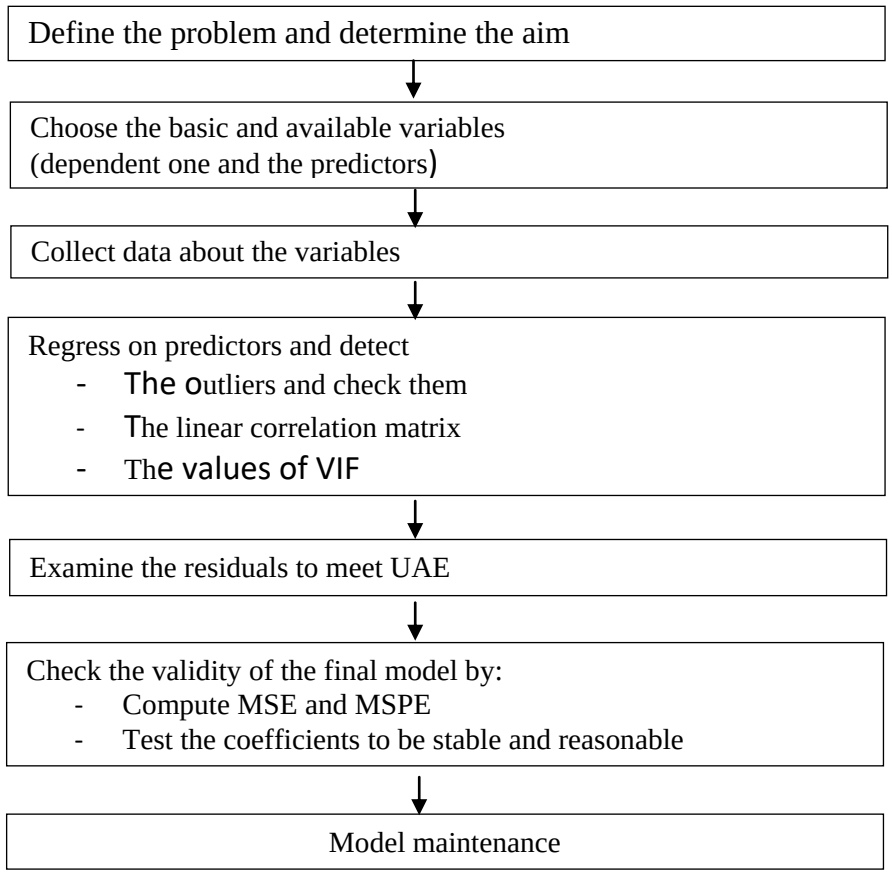


Figure1: Summary of building BLRM.

3- **Applied part:** After we have determined the problem and the aim and chosen variables we have directed to collect data about them

3-1. The Data of random sample: We have taken random sample of size (n=25) families from an area of Baghdad "Alhuria Area" and gotten the following data of their electric energy consumption:

Table(1): data about the local electric energy consumption

No	X1	X2	X3	Y
1	3	750	2	40
2	5	1000	3	50
3	2	1500	3	55
4	7	1650	4	60
5	4	1300	3	45
6	8	1900	4	65
7	10	2300	5	90
8	9	2100	5	80
9	5	2000	3	75
10	7	1800	4	70
11	6	1500	3	60
12	4	1500	2	55
13	5	1900	3	60
14	8	1800	4	60
15	4	1000	2	50
16	3	1100	2	50
17	7	2000	3	75
18	9	2500	4	95
19	5	2000	3	70
20	5	2150	3	70
21	6	1950	3	65
22	3	700	2	45
23	7	800	3	50
24	2	1100	2	55
25	3	1200	2	60

Where:

X_1: family size, X_2= family income (thousand dinars)

X_3: no. of rooms in the family house, y= The cost of electric consumption in thousand dinars.

3-2. The fitted first model by using "SPSS"

3-2-1. The range, the homogenuous variance and the distribution of (ei/s):

ei/s- Range = [-1.655, 1.521] ⇒ There is no outliers.

ei/s ~ N(0,.935) and the variance is homogenuous(*)

(*) See Tucky Box in the Appendix.

3-2-2. The Correlation matrix:

	X_1	X_2	X_3	Y
X_1	1	.671	.887	.726
X_2	.671	1	.706	.902
X_3	.887	.706	1	.702
Y	.726	.902	.702	1

We notice that each (Xi) is correlated highly with (Y) but there is high correlation also among predictors and this is asign of multicollinearity problem that should be checked and treated.

3-2-3. The values of (VIF) and the tests of multicollinearity:

$\hat{Y} = f(X_1, X_2, X_3) = 24.616 + 1.934X_1 + .021X_2 - 2.014X_3$

s.e.	(4.654)	(1.140)	(.003)	(2.985)
sig.	.000	.104	.000	.507
VIF	___	4.782	2.031	5.238

Since VIF >5 Then it is clear that we have sever multicollinearity that affected on some coefficients such that they looked insignificant and unreasonable. To make sure and determine the causing variable. we have applied Farrar- Glauber method and gotten the following results:

- X^2computed = 49.971 whereas X^2 tabled [.01,3]= 11.34(**)

That is; X^2 comp. > 4 X^2tab. ⇒ There is sever multicollinearity.

- $F(x_1) = 41.632$, $F(x_2) = 11.360$, $F(x_3) = 46.601$

$tx_1x_2 = .649$, $t(x_1 x_3) = 5.983$, $tx_2x_3 = 1.601$

So (X_3) is the causing variable of sever multicollinearity.

Since (X_3) gives the same information like (X_1), it is better to omit it. Notice that the signal of X_3–coefficient is incorrect then it is un reasonable.

3-2-4. The Durbin – Watson statistic.

D.W = 1.06 and from the table d_L = 1.12, du= 1.66 That is:

D.W < d_L ⇒ There is autocorrelation problem also should be treated.

3-2-5. The criteria of the best model: From the ANOVA of (Y) in the first fitted model we get the following:

SSE = 708.833, MSE = 33.754 and by using PRESS selection procedure[Allen, 1971] we get the values of prediction sum of squares (PRESS) = 977.87 that is; MSPE = 45.565 (*)

(**) All the arithmetic operations are in the Appendix.

(*) See the Appendix.

3-3. The second model is gotten by dropping (X_3) and the fitted equation is:

$\hat{Y}$ = f(x_1, x_2) = 22.858 + 1.329X_1 + .020X_2

s.e.	(3.809)	(.694)	(.003)
sig.	.000	.069	.000
VIF	___	1.819	1.819

and MSE = 32.918, MSPE = 42.372, D.W = 1.03, dl = 1.21, du = 1.55 It is clear that the second model is better than the first one because the value of VIF < 5 moreover it approaches to one! So the severity of collinearity has gone which affected the precision of estimators (coefficients) and also the fitted equation where MSE and MSPE become less than before. Now suppose the third variable (X_3) is basic, Then Cp = C_3 = 2.455 ⇒ Bias of dropping X_3 = .545. But the autocorrelation problem is still standing because (D.W) is less than (d_L) in Durbin- Watson Table.

3-4. The third model needed to transform the variables (X_1, X_2, Y) into (X_1^*,X_2^*,Y^*) respectively using the value of ($\hat{P}$= 0.426)(*) inorder to getrid of the autocorrelation problem. The new variables are as in the table (2) - see the Appendix - .

The fitted third model gives the following equation:

$\hat{Y}$=f(X_1^*,X_2^*) = 13.933 + 1.160 X_1^* + .020 X_2^*

s.e.	(2.697)	(.557)	(.003)
sig	.000	.050	.000
VIF	___	1.456	1.456

and MSE = 27.113, MSPE = 37.216, D.W= 1.65, d_L= 1.19, du=1.55

It is clear that the fitted third model is the best because the severity of multicollinearity has gone as longas VIF approach more and more to one and the autocorrelation problem also has gone for D.W > d_u and all the coefficients become significant at level (α = 0.05) and stable and reasonable. So this equation is usable (adequate) to predict and control the local consumption of electric energy, specially when we knew that the value of observed (Fobs.=42.728) equals more than four multiple (Ftab), that is; Fobs.> 4 [F(0.05, 2,21) = 3.47] as Dr. G. E. P. Box has said in the thesis written under his direction[Wetz, 1964]. We can also see that ei/s Range = [-1.478, 2.000] and ei/s ~ IIdN (0, .956) and cp=c_3=3.265⇒ Bias = -.265 which is small and less than in the second fitted model. At last we can summarize the improvements in building the regression model in the following table:

(*) See the arithmetic operations and Tuky Box and the ANOVA in the Apendix.

Table (3): The improvements on regression model

The fitted model	UAE	S.e. for estimators	Bias	Significant stable and reasonable	MSE	MSPE
First: $\hat{y}$= f(X_1, X_2, X_3)	Violated: autocorrelation and multicollinearity	big for X_1 and X_3	Zero	Insignificant unstable and unreason – able for X_3	33.754	46.566
Second: $\hat{y}$= f(X_1,X_2)	Violated: autocorrelation problem only	Small	.545	Insignificant for X_1 only	32.918	42.376
Third: $\hat{Y}$= f(X_1^*,X_2^*)	ALL HAS MET (No problems)	Very small	-.265	All significant, stable and reasonable	27.113	37.215

4- conclusions:

1- The examination of residuals is very important to building the best linear regression model (BLRM)
2- The computation of correlation matrix is also important to building the (BLRM) because it discovers the chosen basic variables and refers to severity of multicollinearity problem.
3- The explanatory variables that are chosen to predict and control the local electric energy consumption are really basic variables because, by them, we got the adequate equation.
4- Building the (BLRM) means achieving the following:

a. Meet UAE in the model.
b. High fit to data indicated by small (MSE) and big (Fobserved).
c. Adequacy for prediction indicated by small (MSPE).
d. The significance, stability and the reason of the unbiased or small biased estimators.

5- References

1. بخيت، د. حسين علي وفتح الله، د. سحر "الاقتصاد القياسي" دار اليازوري العلمية للنشر والتوزيع، الاردن، عمان، 2009.
2. عبد المهدي، د.عادل "تصريح وزير النفط العراقي الى القنوات الفضائية العراقية في 2015/7/30.
3. وزارة التخطيط – الجهاز المركزي للاحصاء "المجموعة الاحصائية السنوية للعام 2010-2011".
4. Allen, D.M. "The prediction sum of squares as acriterion for selecting predietor variables". Technieal Report No. 23, Departmeat of statistics, university of kentuneky, 1971.
5. Draper. N. R. and Smith, H. "Applied Regression Analysis" 3rd edition, John wiley & sons, 1999.
6. Haan, C.T. "Statistical methods in Hydrology", 2nd edition Iowa state university press, Ames, Iowa, 2002.
7. Ostrom, C.W., Jr. "Time series Analysis, Regression techniques: Quantitative Applications in the social sciences" 2nd edition, V. 07-009, Newbury park, sage publication, 1990.
8. Wetz, J.M "criteria for judging adequacy of estimator by an approximating response function", Ph.D. thesis, 1964.

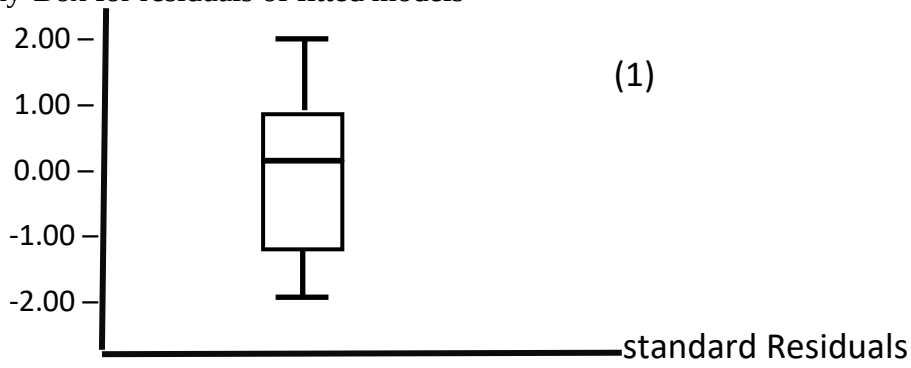
- Appendix -

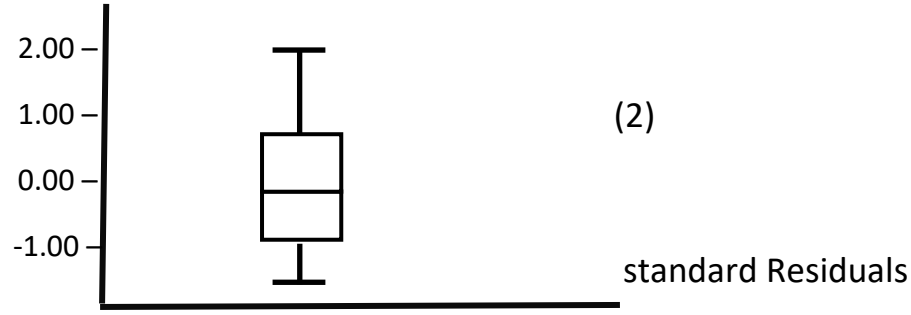
1) Test of Normality for residuals of fitted models:

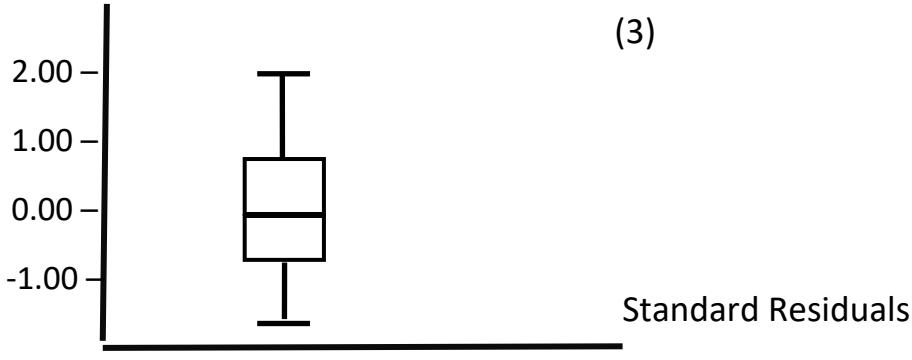
	K.S			Sh.w		
Model	Stabistic	d.f.	Sig.	Stabistic	d.f.	Sig.
1	.107	25	.200*	.960	25	.420
2	.072	25	.200*	.975	25	.760
3	.112	24	.200*	.960	24	.447

Sig. ofk. stest represents the Lower bound of true significance.

2) Tuky-Box for residuals of fitted models







3) Application of Farrar-Glauber Method:

$$|R| = \begin{bmatrix} 1 & .671 & .887 \\ & 1 & .706 \\ & & 1 \end{bmatrix} = \begin{bmatrix} 1 & .706 \\ .706 & 1 \end{bmatrix} - .671 \begin{bmatrix} .671 & .706 \\ .887 & 1 \end{bmatrix} + 887 \begin{bmatrix} .671 & 1 \\ .887 & .706 \end{bmatrix}$$

= .561564- .671 (.044778) + .887 (-.413274) =.104943924

- $X^2 = -(24 - \frac{11}{6})$ (-2.254329) = 49.971 > [X^2(.01,3) = 11.34]

$Fx_1 = \frac{.3955}{.0095}$ = 41.632, $Fx_2 = \frac{.254}{.07236}$ = 11.360, $Fx_3 = \frac{.4045}{.00868}$ = 46.601

Fx_1, Fx_2 and Fx_3 > [F(.05,2,22) = 3.44]

$t_{12} = \frac{0.642586959}{0.990571047}$ = 0.649, $t_{32} = \frac{1.51500429}{0.946398964}$ = 1.601

$t_{31} = \frac{3.691357203}{0.616952996}$ = 5.983

t ($\frac{.05}{2}$, 22) = 2.074

4) Mallow's statistic and the Bias of second and third model:

CP= $C_3 = \frac{724.191}{33.754}$ - (25 – 6) = 2.455

- Bias for second model = 3-2.455 = .545

CP= $C_3 = \frac{569.375}{33.754}$ - (24-6) = 3.265

- Bias for third model = 3- 3.265 = -.265

5- The ANOVA of the third (Last) fitted model

S.O.V	S.S	d.F	M.S	Fobs.	Ftab.	Sig.
Reg.	2316.968	2	1158.484	42.728	3.47	0.00
Residu.	569.375	21	27.113			
Total	2886.343	23				

6- Table(2): The transformed (new) variables:

No.	X_1^*	X_2^*	y^*
1	3.72	680.50	32.96
2	1.40	1074.00	33.70
3	6.15	1011.00	36.57
4	1.02	597.10	19.44
5	6.30	1346.20	45.83
6	6.59	1490.60	62.31
7	4.74	1120.20	41.66
8	1.17	1105.40	40.92
9	4.87	948.00	38.05
10	3.02	733.20	30.18
11	1.49	861.00	29.44
12	3.30	1261.00	36.57
13	5.87	990.60	34.44
14	.59	233.20	24.44
15	1.30	674.00	28.70
16	5.72	1531.40	53.70
17	6.02	1648.00	63.05
18	1.17	935.00	29.53
19	2.87	1298.00	40.18
20	3.87	1034.10	35.18
21	.44	130.70	17.31
22	5.72	501.80	30.83
23	2.00	759.20	33.70
24	2.15	731.40	36.57

Where the estimated value of auto correlation is computed as following: $\hat{P}$= r = $\frac{\sum ei - ei - 1}{\sum e_i^2 - 1}$ = $\frac{308.53}{724.20}$ = 0.426

Table (3): The Residuals of the second fitted model:

No.	X_i^2	$ei.ei-1$
1	3.87	—
2	.11	-.66
3	.58	-.25
4	29.50	4.14
5	88.12	50.99
6	46.26	63.85
7	55.88	-50.84
8	8.05	21.21
9	26.72	14.66
10	2.38	7.98
11	1.16	-1.66
12	11.70	3.69
13	61.07	26.73
14	95.74	76.47
15	2.76	-16.26
16	.95	1.62
17	6.31	2.45
18	95.48	24.54
19	.03	1.65
20	8.16	-.48
21	26.54	14.71
22	16.32	-20.81
23	2.92	6.90
24	53.34	12.48
25	80.25	65.42

Table (4): Allen – PRESS for the three fitted model:

No.	$(DRE_{(1)})^2$	$(DRE_{(2)})^2$	$(DRE_{(3)})^2$
1	5.79	5.34	ـــ
2	1.03	.14	1.21
3	5.19	.90	4.65
4	26.52	33.82	30.82
5	91.30	99.41	70.76
6	54.59	56.14	8.17
7	110.29	88.36	159.51
8	32.33	10.88	.12
9	31.76	33.82	14.91
10	6.61	2.69	.44
11	2.54	1.29	4.55
12	30.60	13.34	14.13
13	75.05	73.08	51.06
14	114.27	117.86	49.57
15	.86	3.26	38.22
16	.51	1.15	.11
17	1.96	7.27	7.13
18	134.77	142.67	118.31
19	.55	.04	25.31
20	14.07	11.55	12.79
21	43.04	30.44	18.10
22	22.71	23.22	69.49
23	3.51	6.91	.05
24	78.19	71.81	49.30
25	90.38	96.80	32.83
PRESS	977.87	932.19	781.54

$^{(1)}(DRE_{(1)})^2$: squared predicted Residuals given from model (1)
$^{(2)}(DRE_{(2)})^2$: squared predicted Residuals given from model (2)
$^{(3)}(DRE_{(3)})^2$: squared predicted Residuals given from model (3)