

comparison between of estimators Robust and Classical in repeated measurement experiments analysis

مقارنة بين المقدرات الحصينة والاعتيادية في تجارب مكررة القياس

saja mehdy jeleel

University of Kufa

saja.mehdy @yahoo.com

Supervisor: Dr. Fadhil Abdul Abbas Abdul Hussein / asis. Prof

Al-Furat Al-Awast Technical University

Abidy_fadhil@yahoo.com

1. Abstract

The main objective of the present study is the use of known statistical methods to overcome the problems in the data and the comparison between them and also to discover the best method and apply it on factual data. This represents the influence of time on children weights, to do that the study considered taking four measurements of weights; they are (the weight in the first month, the weight in the third month, the weight in sixth month, and the weight in the ninth month) from which is one of the governmental health centers in Najaf governorate. The data have been analyzed with different methods including the (Classical) method which represents (Linear Regression Model) and the (Robust Method). Outliers were added to the factual data then analyzed again with the same abovementioned methods. To find out which of the two methods is the best, a comparison has been conducted in which different measurements are used, and they are: Mean Square Error (MSE), Coefficient of Determination (R^2), and (P-Value). The comparison showed that Linear Regression Model is better than Robust Model in case of factual data free of outliers, whereas the Robust Model is found to be better than Linear Regression Model in case of factual data with outliers.

الخلاصة

الهدف الرئيسي من هذه الدراسة هو استخدام اساليب إحصائية تتصف بالحصانة التي لها القدرة على تجاوز المشاكل التي تظهر في البيانات مقارنة مع أساليب (طرائق) تقليدية ومعرفة افضل طريقة تعطي مقدرات لها مقاييس إحصائية جيدة ، حيث تم تطبيق هذه الأساليب على بيانات حقيقية في المجال الطبي وهو يمثل دراسة تأثير الوقت على اوزان الاطفال واعتمدنا على اخذ اربعة قياسات من الاوزان وهي (شهر الاول ،شهر الثالث ،شهر السادس ،شهر التاسع) من مركز صحي خوله بنت زوين وهو من المراكز الصحية الحكومية في النجف.

حيث تم في البدء اختبار هذه البيانات هل تنتمي الى التوزيع الطبيعي ومن ثم قمنا بتحليل البيانات بطرائق مختلفة بما في ذلك الطريقة التقليدية (الكلاسيكية) التي تمثل (نموذج الانحدار الخطي) والطريقة الحصينة في حالة البيانات خالية من مشكلة الشواذ ومن ثم تحليل البيانات بوجود مشكلة الشواذ بإضافة قيم متطرفة على البيانات الحقيقية والمقارنة بينهما اي الطريقتين افضل من خلال الاعتماد على عدة قياسات منها متوسط مربعات الخطأ (MSE) ومعامل التحديد (R^2) و (P-value) وقد وجد ان نموذج الانحدار الخطي هو افضل من الطريقة الحصينة في حالة البيانات الحقيقية الخالية من القيم المتطرفة (الشواذ) ولكن تتفوق الطريقة الحصينة على نموذج الانحدار الخطي في حال وجود مشكلة الشواذ.

2. Introduction

This type of experiment design in medical experiments appears when the measurements of response are repeated more than one time, when the patient is given a treatment in a series period of time.

Also, in agricultural experiments when the re-run of the experiment in the intervals is frequent such as fertilizing trees experiments, the self-correlation between errors is found to be stochastic. The mathematical model can be written according to the following formula:

The Mathematical model can be write according to the following formula:

$$y_{ik} = \mu + \tau_i + \epsilon_{ik}$$

$i=1,2,\dots,n$ $k=1,2, \dots,p$ (*period time*)

The method of data analysis of these experiments depends on the degree of correlation between the random error in the time period (K, K-1).

The present study built a mathematical model using Matlab programming language to compare between Robust and Classical analysis.

3. Objectives

The main objective of the research is to conduct an empirical study to test the frequent measurement in order to discover the influence of time on children weights. The data have been obtained from one of the governmental health centers in Najaf governorate, and it is (Khawla Bint Zwain) health center.

4. Review of Lectures

Standard methods of testing for repeated measures experiments and other mixed linear models are based on the likelihood function or least squares methods (Diggle, P. J. (1988). These methods are known to be sensitive to model misspecification, and can be adversely affected by the presence of outliers. [1]

Catherine Potvin, Martin J. Lechowicz, and Serge Tardif (1990), Studied The Statistical Analysis of Ecophysiological Response Curves Obtained from Experiments Involving Repeated Measures. [8]

Mushfiqur, M. (1995) study a robust analysis of two-way models with repeated measures on both factors is developed using a dispersion function. This robust analysis gives users a complete inference: Estimation, tests for the general linear hypotheses.[5]

Padmanabhan, A.R. , Chinchilli, V. M . , Babu, G. J . (1997), Studied robust analysis of within-unit variances in repeated measurement experiments. The objective of study is to compare the within-unit variances of two or more treatments, products, or techniques. In this situation, a repeated measurement design involving a random effects model, with possibly heterogeneous variances, is appropriate.[9]

Wilcox, R. R. and Keselman, H. J., (2003), Studied Repeated measures one-way a nova based on a modified one – step M-estimator. [13]

Maher, K. (2005), study the case of One-Way Repeated Measures Analysis of Variance under Robust and non-robust with Spss. [6]

Modern Robust Statistical Methods "An Easy Way to Maximize the Accuracy and Power of Your Research", studied by David, M. and Vikki, M. (2008). [2]

5. The Theoretical Side:

5.1 Methods of Repeated Measures Analysis

There are several methods to analyze the repeated measurements, in this section we present some of them and discuss the methods that are appropriate to be used:

- General Linear Modal
- Multivariate Analysis

5.1.1 General Linear Model

This side is a large and complex set of statistical methods tied together by a unifying conceptual framework known as “The General Linear Model” (GLM). This model can be used to answer an amazing variety of research questions within an infinite number of different experimental designs. Basically, the GLM can be used to test almost any hypothesis about a dependent variable (DV) that is measured numerically (e.g., height, income, IQ, age, time needed to run a 100-yard

dash, grade point average, etc.; but not categorical DVs like eye color, sex, etc.). there are special cases of the GLM. So are correlation, regression, and the Analysis of Variance (ANOVA). [7]

Where it will be the analysis of anova table for mathematical model

$$y_{ik} = \mu + \tau_i + \epsilon_{ik} \quad \dots (1)$$

$i=1,2,\dots,n$, $k=1,2, \dots,p$ (*period time*)

Table (1) Analysis of Variance (ANOVA)

source	Ss	df	Ms	F
conditions	SS _{conditions}	(k-1)	MS _{conditions}	$\frac{MS_{conditions}}{MS_{error}}$
subjects	SS _{subjects}	(n-1)	MS _{subjects}	$\frac{MS_{subjects}}{MS_{error}}$
error	SS _{error}	(k-1)(n-1)	MS _{error}	
Total	SS _{Total}	(n-1)		

It is important to mention the significance of using (effect code) for dummy variables to differentiate between the characteristics of the variable instead of the serial numbers which causes that the variable loses a lot of its analytical characteristics and cancels the quantitative importance for the categories of the descriptive or qualitative variable. The variable which has (k) characteristic has been known by the number of (k-1) dummy variables, so as the matrix be (x' x) with full rank. [16]

The function used in the program gives the following formula:

d=dummyvar (time);

5.1.2 Multivariate profile analysis

Profile analysis is an analysis of the profile or a comparison of two or more profiles. Profile analysis is often discussed in the context of administering a battery of p psychological or other tests .In growth curve analysis, where the variables are measured at time intervals, the responses have a natural order. In profile analysis where the variables arise from test scores, there is ordinarily no natural order. A distinction is not always made between repeated measures of the same variable through time and profile analysis of several different commensurate variables on the same individual [11]

In multivariate linear regression (MR) models, one is not interested in predicting only one dependent variable but rather several dependent random variables y_1, y_2, \dots, y_p . Two possible extensions with regard to the set of independent variables for MR models are (1) the design matrix X of independent variables is the same for each dependent variable or (2) each dependent variable is related to a different set of independent variables so that p design matrices are permitted. [12]

General linear model will be as follows :

$$Y=X\beta + \epsilon \quad \dots (2)$$

Such that : $Y_{ij}=[Y_{i1}, Y_{i2}, \dots, Y_{ip}] \sim IN(\mu_i, \Sigma)$
 $i = 1, 2, \dots, I$, $j=1, 2, \dots, p$ (p is period time)

$$\beta^{\wedge} = (X^T X)^{-1} X^T Y \quad \dots (3)$$

The null hypothesis is going to be a test equal to the effect of the response to the same group of people as the:

$$H_0: \mu_1 = \mu_2 = \dots = \mu_p$$

$$H_1: \text{At least two unequal}$$

Using arrays as:

$$\begin{aligned}
 &H_0: c \hat{\beta} A = 0 \\
 &H_1: c \hat{\beta} A \neq 0 \\
 &C_{1 \times 1} = (1) \\
 &\hat{\beta}_{1 \times p} = (\mu_1, \mu_2, \dots, \mu_p) \\
 &A_{p \times p-1} = \begin{bmatrix} 1 & 0 & \dots & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & 0 & \dots & \dots & 0 \\ 0 & 1 & \dots & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 1 & \dots & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \dots & \dots & \dots & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ -1 & \dots & \dots & \dots & -1 \end{bmatrix}
 \end{aligned}$$

As it will become null hypothesis: $H_0:$

$$\begin{bmatrix} \mu_1 & - & \mu_p \\ \mu_2 & - & \mu_p \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \mu_{p-1} & - & \mu_p \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix}$$

To test this hypothesis we use the Wilks test that rely on finding all of the total error square (E) and the total square hypothesis (H) where:

$$E = A'Y'[I - X(X'X)^{-1}]YA \quad \dots (4)$$

$$H = (c \hat{\beta} A)'[C(X'X)^{-1}C']^{-1}(c \hat{\beta} A) \quad \dots (5)$$

As it would reject the null hypothesis under significantly level (α) if:

$$\Lambda = \frac{|E|}{|E+H|} < U(P - 1, 1, N - 1) \quad [17] \quad \dots (6)$$

The multi-variable analysis of variance table is a one-way (MANOVA) as in the following table: [10]

Table (2) the multi-variable analysis of variance (MANOVA)

Source	d.f	ss _p
Groups	I-1	H
Within error	N-I	E
Total	N-1	T

6. Methods of estimation

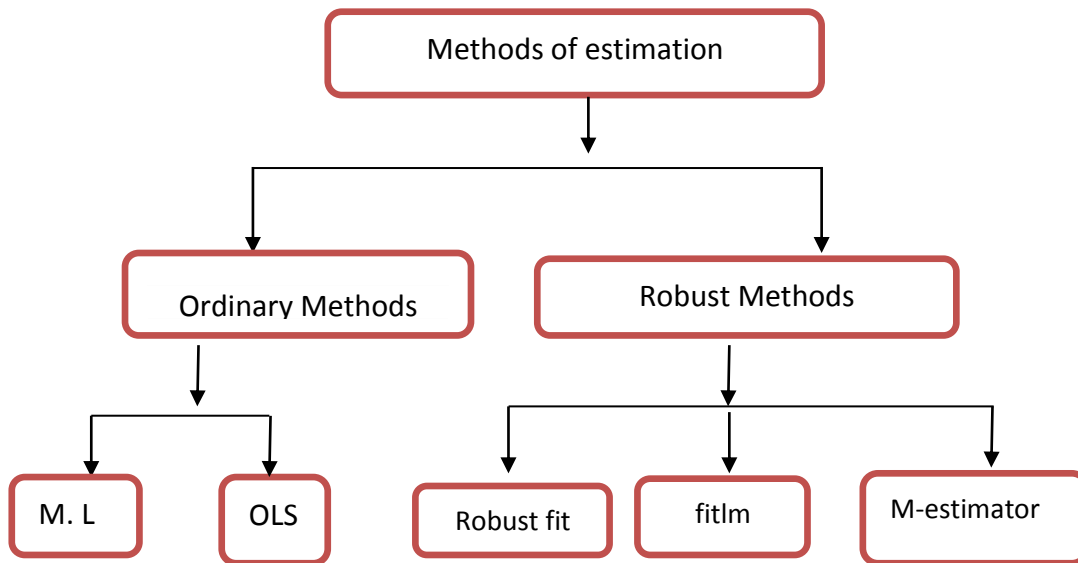


Fig (1) some types of Methods of estimation

6.1 Ordinary Methods [15]

The ordinary ways of estimation depend on the assumption that the parameter to be estimated is fixed and not variable but it is of unknown value. Estimation such as this, the dependence is only on the data of the sample.

Among these are (OLS and M.L) where the mathematical formula for the two samples by comparing between them from proof point of view will be:

$$\beta^{\wedge} = (X^T X)^{-1} X^T Y$$

6.2 Robust Methods

Robust statistics are statistics with good performance for data drawn from a wide range of probability distributions, especially for distributions that are not normal. Robust statistical methods have been developed for many common problems, such as estimating location, scale and regression parameters. One motivation is to produce statistical methods that are not unduly affected by outliers Another motivation is to provide methods with good performance when there are small departures from parametric distributions. From these estimates (M-estimator, fitlm , robust fit). [3]

7. Comparison between Estimation Methods [3]

Three measures compared between the classical and robust estimation methods to get the best way to estimate which are :

1. Mean Square Error (MSE) Which takes the following formula :[4]

$$MSE = \sum_{i=1}^n \frac{(Y_i - \hat{Y}_i)^2}{n} \dots (7)$$

2. P_value

3. Relative efficiency (R²) Which takes the following formula :[14]

$$TSS = \sum (Y - \bar{Y})^2 \dots (8)$$

$$RSS = \sum (Y - \hat{Y})^2 \dots (9)$$

$$R^2 = 1 - \frac{TSS}{RSS} \dots (10)$$

8. The practical side

8.1 Goodness fit

Before starting the statistical analysis of research data, test the hypothesis (Observations belong to the Normal distribution) ,which is the most important conditions for the regression which must be fulfilled in the data analysis. Where it tested through *Easy fit* program, which showed results shown in the following figures:

Normal [#43]					
Kolmogorov-Smirnov					
Sample Size	80				
Statistic	0.08735				
P-Value	0.54556				
Rank	23				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	0.11787	0.13467	0.1496	0.16728	0.17949
Reject?	No	No	No	No	No
Anderson-Darling					
Sample Size	80				
Statistic	0.74867				
Rank	23				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	1.3749	1.9286	2.5018	3.2892	3.9074
Reject?	No	No	No	No	No
Chi-Squared					
Deg. of freedom	6				
Statistic	6.455				
P-Value	0.3742				
Rank	28				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	8.5581	10.645	12.592	15.033	16.812
Reject?	No	No	No	No	No

Fig(2) Output window shows the program easy fit

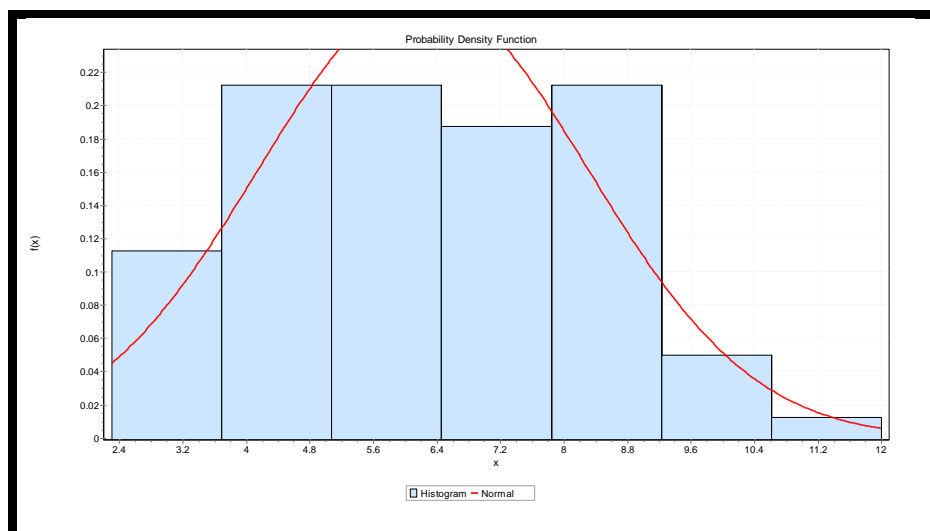


Fig (3) probability Density Function for real data

Form above test results became clear to us not reject the hypothesis above in all test measures, this means that the data belongs normal distribution with parameters ($\mu=6.23$, $\sigma=2.0775$), We will adopt in this aspect on the analysis of data in the two cases, which is the natural state and the status of the presence of outliers

8.2 Natural case:

1- Linear regression model

We used the function in Matlab-the-shelf software for regression analysis that are written in this format:

```
Lm = fitlm (tb, 'weight~time1+time2+time3')
```

```
lm =
```

```
Linear regression model:
```

```
weight ~ 1 + time1 + time2 + time3
```

```
Estimated Coefficients:
```

	Estimate	SE	tStat	pValue
(Intercept)	8.775	0.21972	39.938	9.2758e-53
time1	-4.94	0.31072	-15.898	7.0024e-26
time2	-3.38	0.31072	-10.878	3.7243e-17
time3	-1.86	0.31072	-5.986	6.6299e-08

```
Number of observations: 80, Error degrees of freedom: 76
```

```
Root Mean Squared Error: 0.983
```

```
R-squared: 0.785, Adjusted R-Squared 0.776
```

```
F-statistic vs. constant model: 92.4, p-value = 2.78e-25
```

anova_lm =

	SumSq	DF	MeanSq	F	pValue
time1	244.04	1	244.04	252.76	7.0024e-26
time2	114.24	1	114.24	118.33	3.7243e-17
time3	34.596	1	34.596	35.832	6.6299e-08
Error	73.378	76	0.9655		

2- Linear regression model (robust fit):

We used the function in Matlab-the-shelf software for the analysis of the robust that are written in this regression formula:

`h=fitlm(tb,'weight~time1+time2+time3','robustOpts','on')`

h =

Linear regression model (robust fit):
 weight ~ 1 + time1 + time2 + time3

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	8.6827	0.22191	39.127	4.1058e-52
time1	-4.8435	0.31383	-15.434	3.9428e-25
time2	-3.2703	0.31383	-10.421	2.6699e-16
time3	-1.7665	0.31383	-5.6287	2.9021e-07

Number of observations: 80, Error degrees of freedom: 76

Root Mean Squared Error: 0.992

R-squared: 0.775, Adjusted R-Squared 0.766

F-statistic vs. constant model: 87.1, p-value = 1.57e-24

anova_ro =

	SumSq	DF	MeanSq	F	pValue
time1	234.6	1	234.6	238.2	3.9428e-25
time2	106.95	1	106.95	108.59	2.6699e-16
time3	31.203	1	31.203	31.682	2.9021e-07
Error	74.851	76	0.98489		

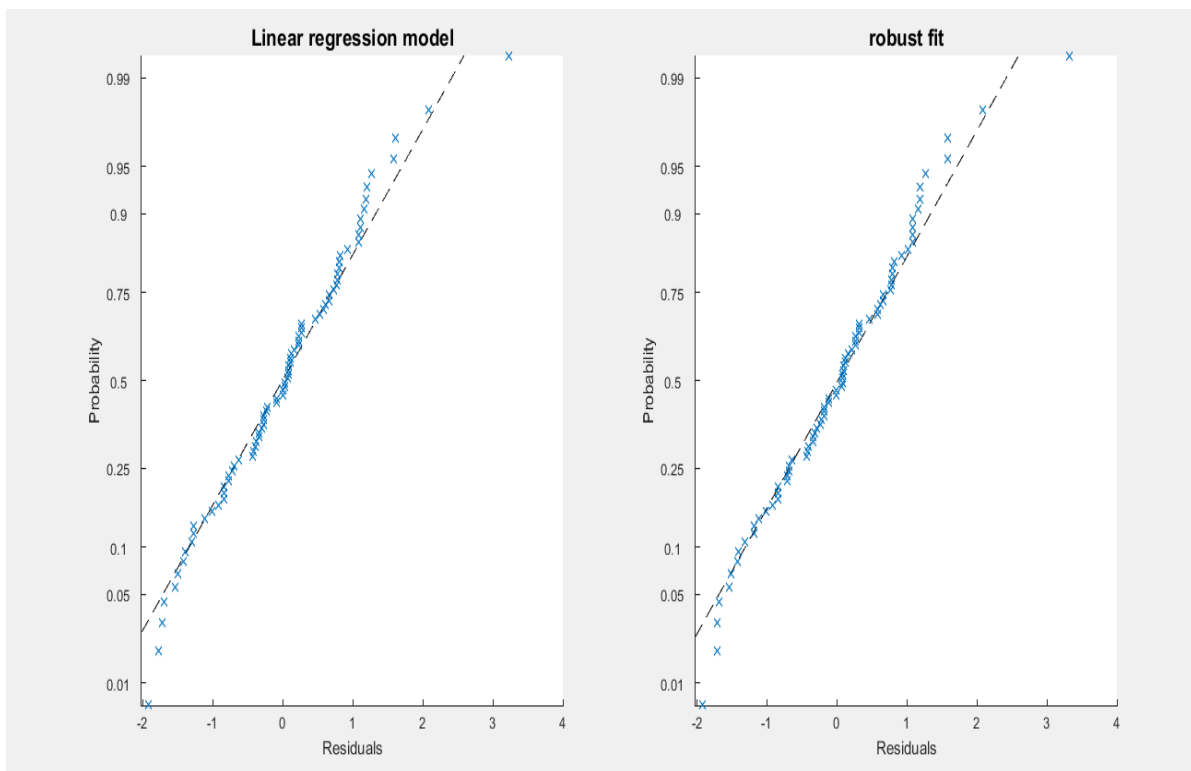


Fig (4) plot Residuals for Natural data

3- Discuss the results :

In light of the previous review the most important results can be summarized as follows:

- In case of Linear Regression Model, the Mean Square Errors (MSE) is smaller than in case of Robust Fit.
- In contrast to the previous point, the coefficient of determination (R^2) is larger in case of Linear Regression Model than in case of Robust Fit.
- It is noted that (p-value) is smaller in case of Linear Regression Model than in case of Robust Fit.

8.3 outliers case:

The detection of outliers has been of concern to statisticians and other scientists for over a century. Those seen that are not consistent with the rest of the group data for any variable of variables for a particular phenomenon or set of phenomena, may be of a great value, or may be

small and located in one of the parties to the total views ranked in ascending or descending, and can be defined statistically as she viewing derived from various community for the community in question, i.e., that the original community polluted by observations of another community, and these are called views persistent, there are several reasons for the emergence of views anomalies including measurement errors, recording or sampling errors.

Where we've added abnormal values on previous statements by the following equation:

$$Weight(i) = Weight(i) + Weight(i) * 0.8$$

Where we have added six of outliers on real data that we used previously in the natural state and randomly selected these values also comes: $i=10, 20, 12, 33, 14, 55$

1- Linear regression model

We used the function in Matlab-the-shelf software for regression analysis that are written in this format:

`Lm = fitlm (tb, 'weight~time1+time2+time3')`

```
lm =
```

```
Linear regression model:
```

```
weight ~ 1 + time1 + time2 + time3
```

```
Estimated Coefficients:
```

	<u>Estimate</u>	<u>SE</u>	<u>tStat</u>	<u>pValue</u>
(Intercept)	9.475	0.41022	23.097	4.273e-36
time1	-5.476	0.58014	-9.439	1.9338e-14
time2	-3.704	0.58014	-6.3846	1.2361e-08
time3	-2.22	0.58014	-3.8266	0.00026478

```
Number of observations: 80, Error degrees of freedom: 76
```

```
Root Mean Squared Error: 1.83
```

```
R-squared: 0.558, Adjusted R-Squared 0.541
```

```
F-statistic vs. constant model: 32, p-value = 1.78e-13
```

anova_lm =

	SumSq	DF	MeanSq	F	pValue
time1	299.87	1	299.87	89.095	1.9338e-14
time2	137.2	1	137.2	40.763	1.2361e-08
time3	49.284	1	49.284	14.643	0.00026478
Error	255.79	76	3.3657		

2- Linear regression model (robust fit):

We used the function in Matlab-the-shelf software for the analysis of the hippocampus that are written in this regression formula:

`h=fitlm(tb,'weight~time1+time2+time3','robustOpts','on')`

h =

Linear regression model (robust fit):
weight ~ 1 + time1 + time2 + time3

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	8.6937	0.27896	31.165	4.844e-45
time1	-4.7941	0.39451	-12.152	1.7106e-19
time2	-3.0567	0.39451	-7.7481	3.3367e-11
time3	-1.867	0.39451	-4.7324	1.0049e-05

Number of observations: 80, Error degrees of freedom: 76

Root Mean Squared Error: 1.25

R-squared: 0.683, Adjusted R-Squared 0.67

F-statistic vs. constant model: 54.5, p-value = 6.65e-19

anova_ro =

	SumSq	DF	MeanSq	F	pValue
time1	229.83	1	229.83	147.67	1.7106e-19
time2	93.435	1	93.435	60.033	3.3367e-11
time3	34.856	1	34.856	22.396	1.0049e-05
Error	118.28	76	1.5564		

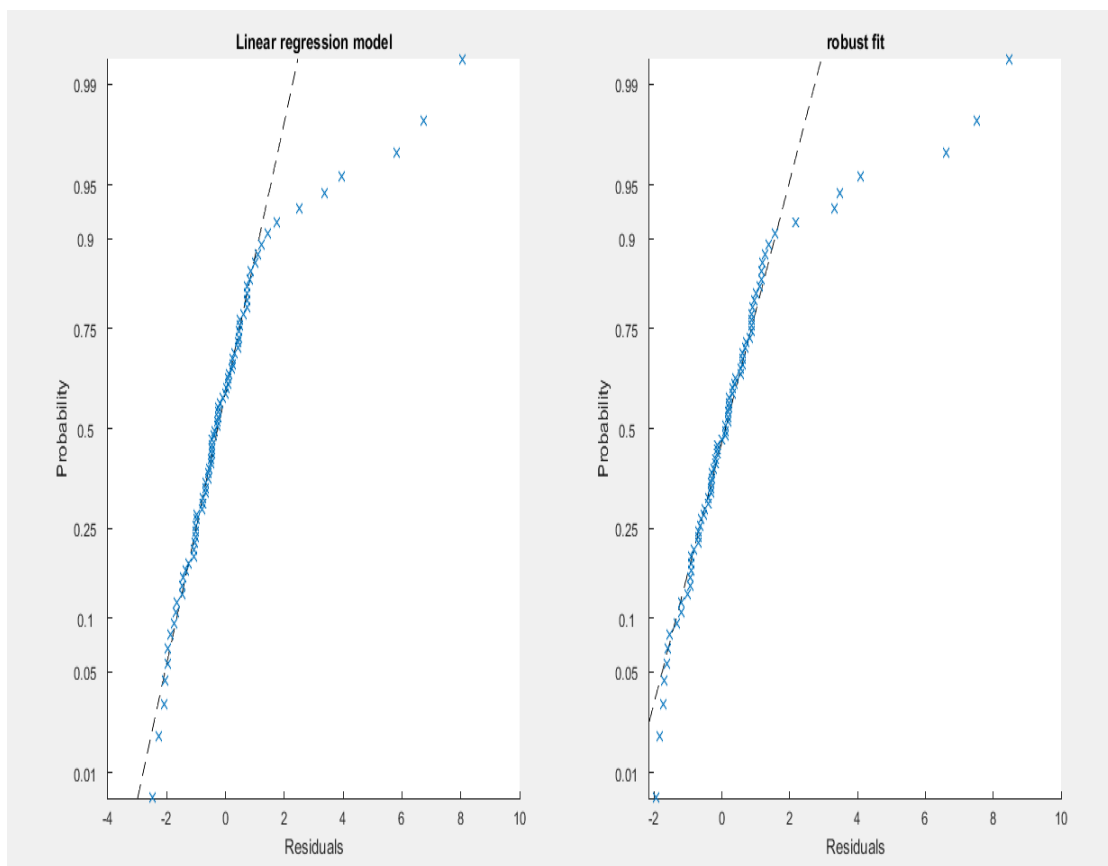


Fig (5) plot Residuals for outliers data

- Note from the above diagram that in the case of a outliers values are more a departure from the regression line on the reverse the natural case.

4- Discuss the results :

In light of previous results, the most important results can be summarized as follows:

- In case of Robust Fit, the Mean Square Errors (MSE) is smaller than in case of Linear Regression Model.
- In contrast to the previous point, the coefficient of determination (R^2) is larger in case of Robust Fit than in case of Linear Regression Model.
- It is noted that (p-value) is smaller in case of Robust Fit than in case of Linear Regression Model.

9. Conclusion

- It is noted that in case of (factual data), the Linear Regression Model is better than Robust Fit because it has smallest Mean Square Errors (MSE) , biggest coefficient of determination (R^2) , and smallest (p-value).
- It is noted that in case of outliers, Robust Fit is better than Linear Regression Model, because it has smallest Mean Square Errors (MSE)) , biggest coefficient of determination (R^2) , and smallest (p-value).

10. Recommendations

In light of the results of the study, it is recommended that:

- In case of data containing outliers, Robust Fit can be used as the most appropriate method of data analysis for its ability to find out best estimation of values in spite of existence of problem in the real data.
- When the data are free from outliers, the Linear Regression Model is proved as the better method for data analysis and the efficiency of this method is proved among others.

11. References

- [1] Diggle, P. J. (1988) An approach to the analysis of repeated measures. *Biometrics* 44, 959-971.
- [2] David, M. and Vikki, M. (2008). Modern Robust Statistical Methods "An Easy Way to Maximize the Accuracy and Power of Your Research" , October 2008 , American Psychologist 591, Vol. 63, No. 7 .
- [3] Huber P . J., and Ronchetti E . M., (2009). "**Robust Statistic**". WILEY, 2th Edithion.
- [4] Kapur. J. N., and Saxena. H.C (2009) "**Mathematical Statistics**", S . CHAND & Company LTD.
- [5] Mushfiqur , M., R. (1995), Robust analysis of two-way models with repeated measures on both factors , Test, vol.4, no.1, 39-62.
- [6] Maher, K., (2005), One-Way Repeated Measures Analysis of Variance, Zayed University Office of Research SPSS for Windows® Workshop Series .
- [7] Miller, J. and Haden², P., Statistical Analysis with The General Linear Model¹, Copyright (c) 1988–1990, 1998–2001, 2006. Version: February 16, 2006
- [8] Potvin , C.; Iechowicz , M.J. ; Tardif, S. ,(1990), The Statistical Analysis of Ecophysiological Response Curves Obtained from Experiments Involving Repeated Measures, Ecology, Vol.71, No.4, 1389-1400.
- [9] Padmanabhan, A.R., Chinchilli VM., Babu G.J.(1997), Robust analysis of within- unit variances in repeated measurement experiments, biometrics 35, 1520-1526.
- [10] Reinsel , G. , (1982) , "Multivariate repeated – Measurement Growth Curve Models With Multivariate random – effect covariance structure " , JASA , 77, 190 – 195.
- [11] Rencher, A . C . , (2002) , "Methods of Multivariate Analysis " , second edition ,
- [12] Timm, N.H.(2002) , "applied multivariate analysis " , Springer – verlag New York , Inc.
- [13] Wilcox , R . R. ; Keselman , H . J. , (2003) , Repeated measures one-way anova based on a modified one – step M-estimator , British journal of mathematical and statistical psychology , vol. 56, No. 10, 15-25 .
- [14] Wendy L . Martine Z, Angel R. Martine Z,(2008). "Computational Statistics Hand book with MATLAB". 2nd Edition, computer science and data analysis series, Chapman & Hall /CRC .
- [15] الصفاوي ، صفاء يونس وطه ، عمار حازم . (2005) . " بعض طرائق المقدرات التقليدية ومقدر بيز لمعاملات نموذج الانحدار الخطي العام (دراسة مقارنة مع تطبيق في مجال طبي) " . مجلة تنمية الرافدين ، المجلد (80) ، العدد (27) ، الصفحات (91 – 103) .
- [16] دبوب ، مروان عبد العزيز ؛ هدية ، وكاع علي، (2008) ، العلاقة بين تحليل التباين وتحليل انحدار المتغيرات الوهمية ، المجلة العراقية الاحصائية ، 132-113
- [17] رضا ، صباح ، منفي ، (1993) ، " طرق تحليل القياسات المكررة وتطبيقها على اللياقة البدنية " ، رسالة ماجستير ، كلية الادارة والاقتصاد ، جامعة بغداد .

Appendix

Table(1) Data of article (Body weight of children in Khawla Bint Zwain Health center/Najaf)

Children	Time			
	Month1	Month2	Month3	Month4
1	4.5	5	7	9.7
2	2.3	3.7	6.2	9.5
3	3	4	7	9
4	4.1	4.5	8	8.8
5	4.6	6.5	8.1	8.5
6	3.5	6.5	9	12
7	4.5	6	8	9.6
8	3.2	4.7	5.8	8
9	4.1	5.5	5.9	8.8
10	4.1	6.2	7	7.5
11	3.6	5.3	5.5	7
12	5	7	7.7	9
13	5.1	5.4	6	9
14	3	6.2	8.5	9.3
15	4	5.1	6.7	8
16	3.9	6.6	7.5	8.9
17	3.4	5.3	7.7	8.4
18	3	4.1	5.2	7.5
19	3.5	3.9	5	8.5
20	4.3	5.5	6.5	8.5

Programs written by Matlab software ver 2015 a

```
%Linear Regression Model for Natural data:
clc
A= xlsread('saja.xlsx');
weight=A(:,1);time=A(:,2);d=dummyvar(time);
time1=d(:,1);time2=d(:,2);time3=d(:,3);
tb=table(weight,time1,time2,time3,'variableNames',{'weight','time1','time2','time3'});
lm = fitlm(tb,'weight~time1+time2+time3')
anova_lm=anova(lm) %Linear regression model (robust fit):
h=fitlm(tb,'weight~time1+time2+time3','robustOpts','on')
anova_ro=anova(h) % plotResiduals
subplot(1,2,1)
plotResiduals(lm,'probability')
title('Linear regression model','FontSize',14)
subplot(1,2,2)
plotResiduals(h,'probability')
title('robust fit','FontSize',14)
%Linear regression model for outliers data:
Clc
```

```
A= xlsread('saja.xlsx');weight=A(:,1);
weight(10)=weight(10)+weight(10)*0.8;weight(20)=weight(20)+weight(20)*0.8;
weight(12)=weight(12)+weight(12)*0.8;weight(33)=weight(33)+weight(33)*0.8;
weight(14)=weight(14)+weight(14)*0.8;weight(55)=weight(55)+weight(55)*0.8;

time=A(:,2); d=dummyvar(time); time1=d(:,1);time2=d(:,2); time3=d(:,3);
tb=table(weight,time1,time2,time3,'variableNames',{'weight','time1','time2','time3'});
lm = fitlm(tb,'weight~time1+time2+time3')
anova_lm=anova(lm)
%Linear regression model (robust fit)for outliers data:
h=fitlm(tb,'weight~time1+time2+time3','robustOpts','on')
anova_ro=anova(h);% plotResidualsfor outliers data
subplot(1,2,1); plotResiduals(lm,'probability'); title('Linear regression model')
subplot(1,2,2); plotResiduals(h,'probability'); title('robust fit')
```