

( 2005/6/13 2005/4/5 )

: (1990-89) :  
100 (2000-99) (1995-94)

## The Effect of Outliers on the Regression Analysis Results with Application on the Premature Births

**Marwan A. Dabdoub**

**Farah A. Younis**

*Department of Statistics  
College of Computers Science and Mathematics  
University of Mosul*

### ABSTRACT

The outliers were founded as results of the regression analysis and also by box plot graph. The outlier values were tackled by trimmed mean.

The data about the life of premature births and other nine factors were collected from three records which represented by pre-blockade years 89-1990, the years 94-1995 which are under blockade and before memorandum of understanding and after memorandum of understanding the years 99-2000. A random sample of size 100 premature births were taken from each period.

It was found that the outliers affected regression analysis results, so it is necessary to search for outliers whenever regression analysis is needed.

The pregnancy period and the weight of premature had the major effect on his life in pre-blockade and after memorandum of understanding periods, before the period of memorandum of understanding two factors were added they were: period of nursery stay and type of disease that the premature births suffered from. The abortion cases increase in last period.

.(2001 )

1755

Boscovich

.(Anderews and Pregibon, 1973)

Huber (1973)

Tukey (1977) ."

"

Box and

whisker plots with five number summaries

Rousseeuw (1987)

(1999)

i

( $e_i$ )

(2002)

premature

37

( )

2.50

(2002)

37

Al-Zubeidi (1994)

%11.00 %32.02

(2000)

(2002)

%17.00

%18.00

(2004)

...

:

:

:

(MSe)

(R<sup>2</sup>)

:

X<sub>j</sub>

y<sub>i</sub>

(2002 )

$$y_i = \beta_0 + \sum_{j=1}^m \beta_j x_{ij} + \varepsilon_i \quad \dots \quad (1)$$

( )

= n i = 1, 2, ..., n:

( )

.i

j = 1, 2, ..., m

m :x<sub>ij</sub>

:β<sub>0</sub>

m :β<sub>j</sub>

.i

:ε<sub>i</sub>

e<sub>i</sub>

Residuals

e<sub>i</sub>s

(-2)-(-2)

%95.45

(Myers, 1986)

:

Outliers

Minitab

:

:

Unusual Observations :

X

(2)-(2)

(e<sub>i</sub>s)

:

R

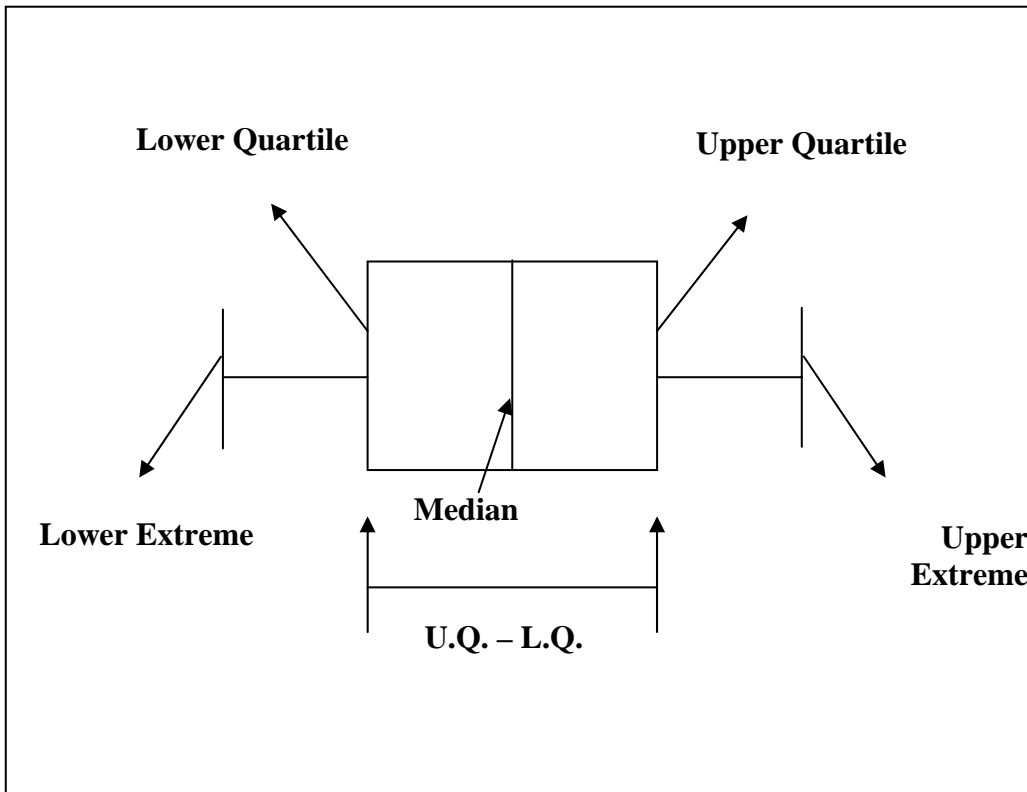
$\hat{y}$

Box

1

plot

.(2001 )



:1

:

(1998 )

n-2 Trimmed mean

...

Stepwise selection procedure :

(1988)

(1987)

Minitab

Forward selection procedure

Backward elimination procedure

F(out)

F(in)

:

R<sup>2</sup>

(1998 )

SST

SSReg.

SSe

MSe

(n-m-1)

SSe

:

m

n

$$SST = \sum(y_i - \bar{y})^2 = (SSReg. = \sum(\hat{y}_i - \bar{y})^2) + (SSe = \sum(y_i - \hat{y}_i)^2) \quad \dots \quad (2)$$

$$R^2 = (SSReg. / SST) = (\sum(y_i - \bar{y})^2 - \sum(y_i - \hat{y}_i)^2) / SST \quad \dots \quad (3)$$

$$MSe = SSe / (n - m - 1) = \sum(y_i - \hat{y}_i)^2 / (n - m - 1) \quad \dots \quad (4)$$

$$MSe \quad R^2 \quad (4 \quad 3)$$

-2)-(2)

$$\sum(y_i - \hat{y}_i)^2$$

(

MSe

R<sup>2</sup>

SSe

R<sup>2</sup>

F

MSe R<sup>2</sup>

MSe

R<sup>2</sup>

F

MSe  
R<sup>2</sup>

$\hat{y}_i$   
MSe R<sup>2</sup>

:

-

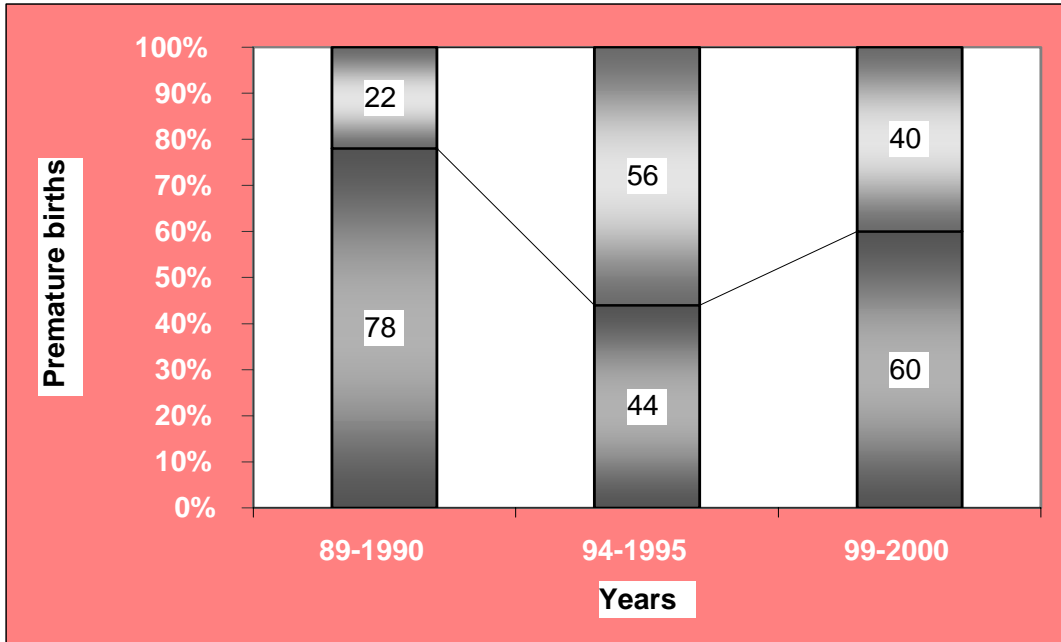
-94) (1990-89) :  
100 (2000-99) (1995  
Response ( )  
(explanatory variables ) predictor variables

:

:X4 ( ) :X3 ( ) :X2 ( ) :X1  
:D1) :X6 ( 1= 0= ) :X5 ( )  
:D4 :D3 :D2  
:X7 ( )  
( 1= 0= ) : :X9 :X8

SPSS V.11.5 Minitab V.13.12 :

.Excel (Office 97)



( ) ( ) :2

%56.00 (1995-94)

2

(1990-89)

%22.00

(2000-99)

%40.00

.%16.00

:



:1990-89

1  
 St Resid 27  
 X (-2)-(2)

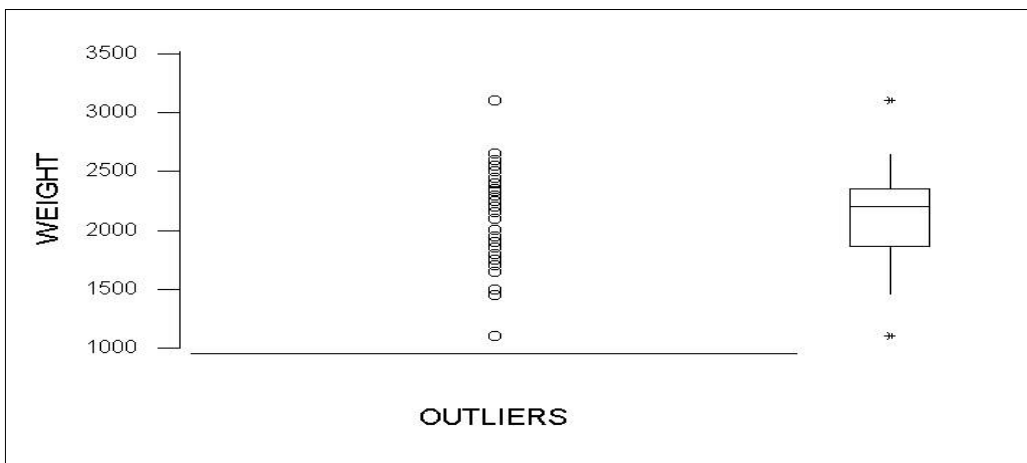
:1

Obs	RESULT	Fit	SE Fit	Residual	St Resid
27	1.0000	1.3140	0.1364	-0.3140	-1.5600 X

(3 )

27

3 1



:3

2

...

:2

Step	1	2	3
Constant	-1.2390	-2.1220	-2.3480
WEIGHT	0.0010	0.0005	0.0005
PR. AGE		0.0550	0.0540
MO. AGE			0.0100
(MSe) <sup>1/2</sup>	0.2740	0.2550	0.2460
R <sup>2</sup>	57.1700	63.2200	66.0700

3100.00 27

. 2241.00

:

3

.27

:3

Step	1	2	3
Constant	-1.3870	-1.6720	-2.2680
WEIGHT	0.0010	0.0010	0.0007
MO. AGE		0.0114	0.0108
PR. AGE			0.0420
(MSe) <sup>1/2</sup>	0.2630	0.2520	0.2430
R <sup>2</sup>	60.4900	64.2100	67.0500

75

3

(4 )

. 1611.00

1100.00

:4

Obs	RESULT	Fit	SE Fit	Residual	St Resid
75	0.0000	-0.1627	0.0863	0.1627	0.7200 X

75 27

:2

3 2  
 .R<sup>2</sup> Mse 2 1  
 2 1 R<sup>2</sup> 3 MSe  
 . %67.04 %60.39  
 MSe R<sup>2</sup>  
 .MSe R<sup>2</sup>

**:1995-94**

:

5

:5

Step	1	2	3
Constant	-1.1029	-0.7337	-0.7955
WEIGHT	0.0009	0.0008	0.0009
DISEASE		-0.126	-0.1470
PERIOD			0.01480
(MSe) <sup>1/2</sup>	0.3200	0.2790	0.2690
R <sup>2</sup>	59.2200	69.3100	71.8800

76 19 8

.(6 )

:6

Obs	RESULT	Fit	SE Fit	Residual	St Resid
8	1.0000	0.9819	0.0993	0.0181	0.0700 X
19	0.0000	0.2747	0.0953	-0.2747	-1.0900 X
76	0.0000	-0.1837	0.0936	0.1837	0.7300 X

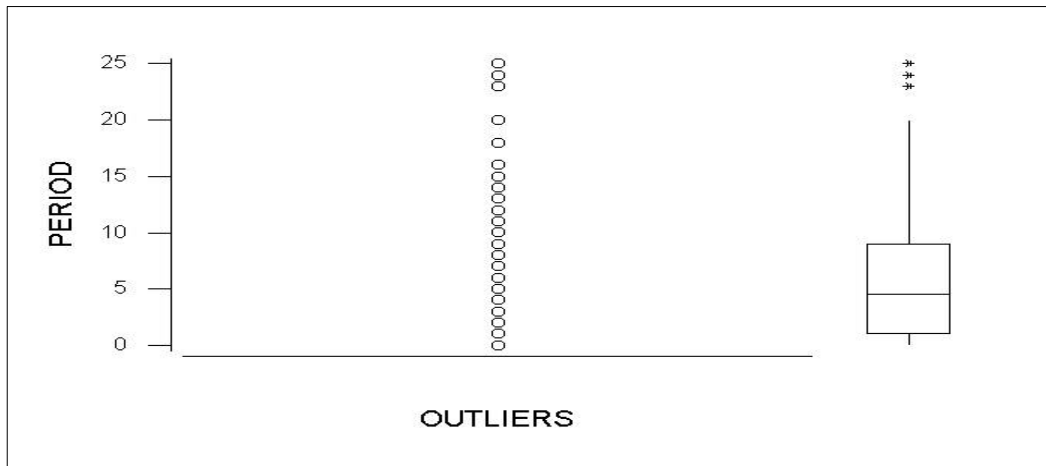
...

6

4

23 25 24

5



:4

20

56

MSe R<sup>2</sup>

6

3

5

%.71.88

%71.52

R<sup>2</sup>

:2000-99

7

St Resid

R

6

RX

19

(-2)-(-2)

( )

:7

Obs	RESULT	Fit	SEFit	Residual	StResid
6	1.0000	0.4894	0.0721	0.5106	2.0500 R
19	1.0000	0.5079	0.1542	0.4921	2.3700RX

:

.(8 )

:8

Step	1	2	3
Constant	-1.3060	-2.3400	-2.3170
WEIGHT	0.0010	0.0007	0.0007
PR. AGE		0.0500	0.0530
BIRTHS B. PR.			-0.0400
(MSe) <sup>1/2</sup>	0.2790	0.2640	0.2560
R <sup>2</sup>	68.3100	71.7300	73.8300

X

19

6

:

(9 )

:9

Step	1	2	3
Constant	-1.3060	-2.3400	-2.3160
WEIGHT	0.0010	0.0007	0.0007
PR. AGE		0.0500	0.0540
ABORTION			-0.0770
(MSe) <sup>1/2</sup>	0.2790	0.2640	0.2510
R <sup>2</sup>	68.3100	71.7300	74.7800

...

9 8

7

19

(-2)-(2)

6

22 20 27

70 52 5

42.00

80

27.00

MSe R<sup>2</sup>

:

F

:

-1: 9 5 2:

.

:

-2 .

:

:

:

H<sub>0</sub>: μ<sub>1</sub> = μ<sub>2</sub> = μ<sub>3</sub>

:

:

H<sub>1</sub>: at least two of μ<sub>j</sub>'s are not equal.

Duncan's multiple range test

.(2002 )

:

F

		1912.50		2111.86
				1714.42
%4.50	2.50		(1999)	
		1994	%20.00	1990
				:
			(2002)	
	F			
31.84				30.84
		32.42		
				:
28.19	28.77 :			
(3 )			26.92	
40.00	20.00		(2002)	
				:
19	18	20		

...

(5 )

%56.00

:

(1989)

:

(Chi-sq.=12.0308, d.f.=2,  $\alpha=0.0050$ )

%79.00

%59.00

%66.00

:

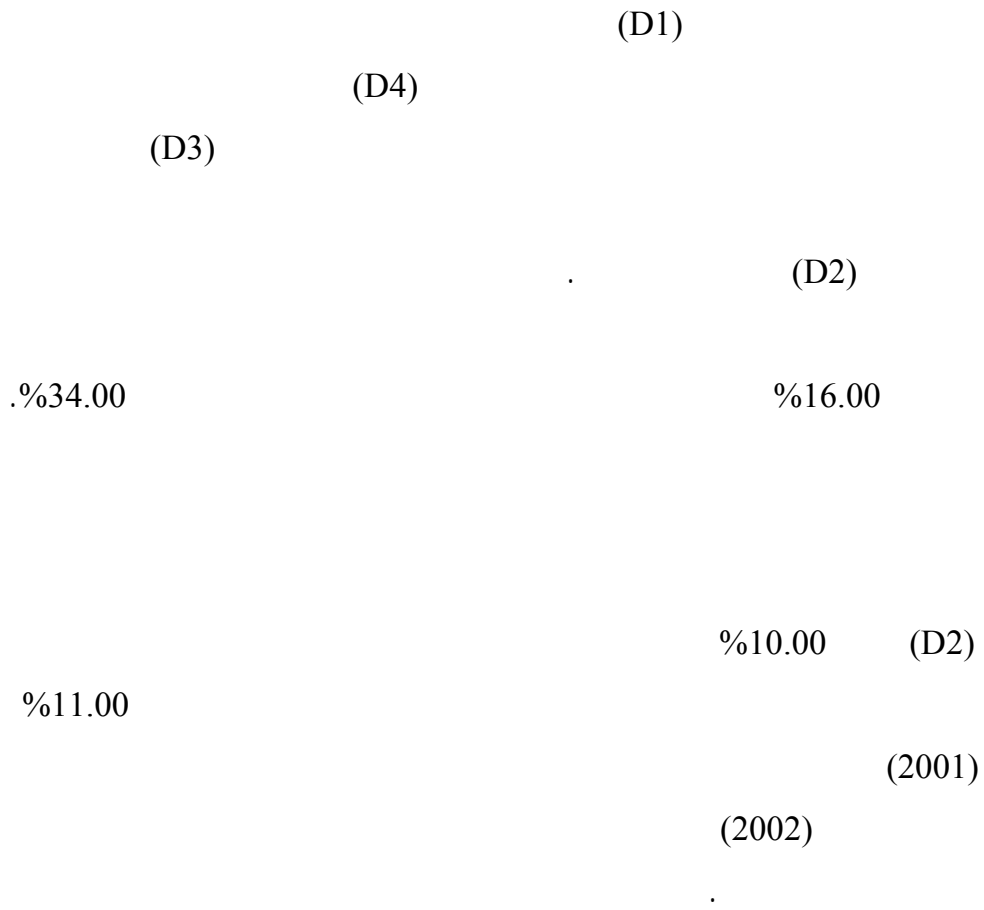
( )

:

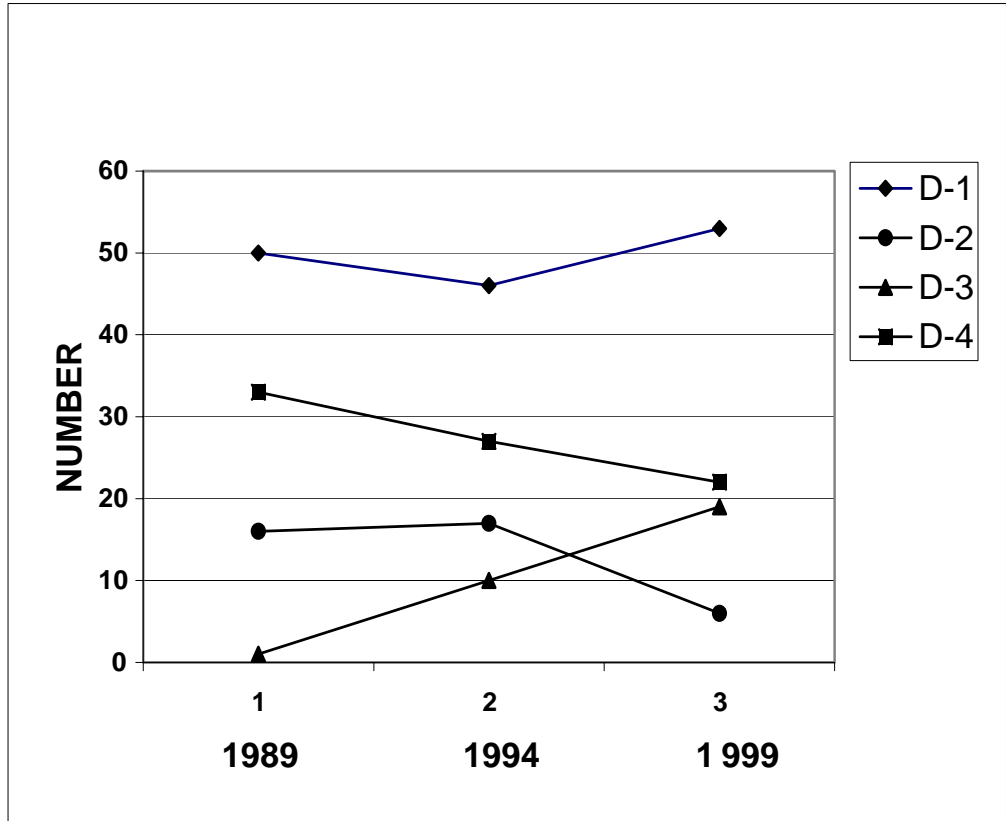
(2004 )

( Chi-sq.=24.6080, d.f.=6, P-value=0.0000)





...



:5

-1

-2

-3

MSe

$R^2$

-4

-5

%.16.00

.2004  
.2002  
.42-31  
.1998  
.1987  
.2001  
.1999  
.2004  
.2001  
.2000  
.1989  
.2002  
.1998  
- -  
.360-353 53 -20  
.1999

...

.2001

.2002

.9-1

13

.1988

.2002

- Al-Zubeidi, I.T., 1994. Perinatal Mortality and Birth Weight in the Valuation Maternal Health Services, Ph.D. Thesis, University of Mosul-Iraq.
- Anderews, D.E. and Pregibon, D., 1973. Finding the Outliers That Matter, J.R. Statist. Soc., (40), pp. 87-93.
- Huber, P.J., 1973. Robust Regression a Symptotics Conjectures and Monte Carlo, Ann. Statist., 19 (2),pp.136-140.
- Myers, R.H., 1986. Classical and Modern Regression With Applications, Duxbury Press, Massachusetts, USA.
- Rousseeuw, P., 1987. Robust Regression and Outliers Detection, John Wil-ey and Sons, N.Y., USA.
- Tukey, J.W., 1977. Exploratory Data Analysis, Wesley Redding, Addison, USA.