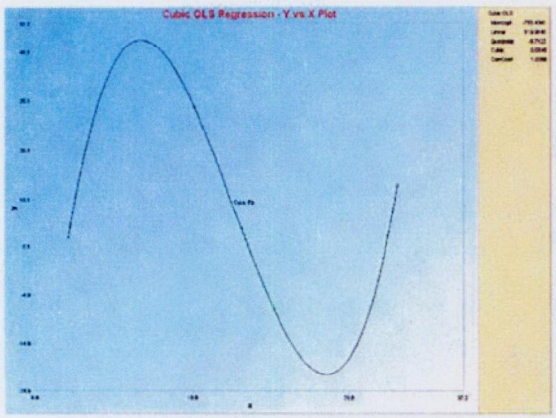
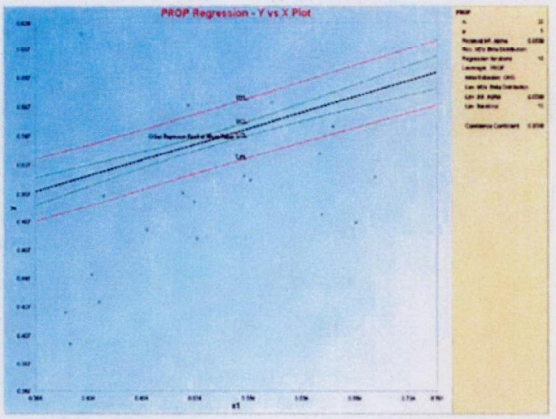
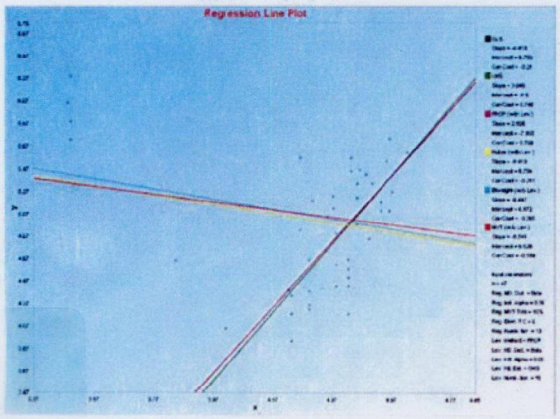
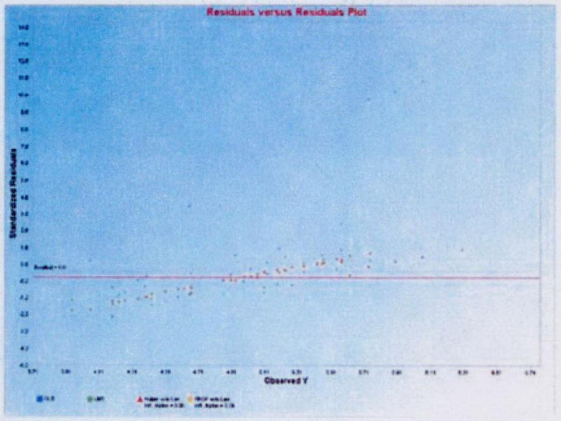
EPA United States Environmental Protection Agency

Scout 2008 Version 1.0 User Guide Part III









RESEARCH AND DEVELOPMENT

EPA/600/R-08/038
February 2009
www epa gov

Scout 2008 Version 1.0 User Guide

(Second Edition, December 2008)

John Nocerino

U.S. Environmental Protection Agency
Office of Research and Development
National Exposure Research Laboratory
Environmental Sciences Division
Technology Support Center
Characterization and Monitoring Branch
944 E. Harmon Ave.
Las Vegas, NV 89119

Anita Singh, Ph.D.[1]
Robert Maichle[1]
Narain Armbya[1]
Ashok K. Singh, Ph.D.[2]

[1]Lockheed Martin Environmental Services
1050 E. Flamingo Road, Suite N240
Las Vegas, NV 89119

[2]Department of Hotel Management
University of Nevada, Las Vegas
Las Vegas, NV 89154

U.S. Environmental Protection Agency
Office of Research and Development
Washington, DC 20460

Notice

The United States Environmental Protection Agency (EPA) through its Office of Research and Development (ORD) funded and managed the research described here. It has been peer reviewed by the EPA and approved for publication. Mention of trade names and commercial products does not constitute endorsement or recommendation by the EPA for use.

The Scout 2008 software was developed by Lockheed-Martin under a contract with the USEPA. Use of any portion of Scout 2008 that does not comply with the Scout 2008 User Guide is not recommended.

Scout 2008 contains embedded licensed software. Any modification of the Scout 2008 source code may violate the embedded licensed software agreements and is expressly forbidden.

The Scout 2008 software provided by the USEPA was scanned with McAfee VirusScan and is certified free of viruses.

With respect to the Scout 2008 distributed software and documentation, neither the USEPA, nor any of their employees, assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed. Furthermore, the Scout 2008 software and documentation are supplied "as-is" without guarantee or warranty, expressed or implied, including without limitation, any warranty of merchantability or fitness for a specific purpose.

Acronyms and Abbreviations

% NDs	Percentage of Non-detect observations
ACL	alternative concentration limit
A-D, AD	Anderson-Darling test
AM	arithmetic mean
ANOVA	Analysis of Variance
AOC	area(s) of concern
B*	Between groups matrix
BC	Box-Cox-type transformation
BCA	bias-corrected accelerated bootstrap method
BD	break down point
BDL	below detection limit
BTV	background threshold value
BW	Black and White (for printing)
CERCLA	Comprehensive Environmental Response, Compensation, and Liability Act
CL	compliance limit, confidence limits, control limits
CLT	central limit theorem
CMLE	Cohen's maximum likelihood estimate
COPC	contaminant(s) of potential concern
CV	Coefficient of Variation, cross validation
D-D	distance-distance
DA	discriminant analysis
DL	detection limit
DL/2 (t)	UCL based upon DL/2 method using Student's t-distribution cutoff value
DL/2 Estimates	estimates based upon data set with non-detects replaced by half of the respective detection limits
DQO	data quality objective
DS	discriminant scores
EA	exposure area
EDF	empirical distribution function
EM	expectation maximization
EPA	Environmental Protection Agency
EPC	exposure point concentration
FP-ROS (Land)	UCL based upon fully parametric ROS method using Land's H-statistic

Gamma ROS (Approx.)	UCL based upon Gamma ROS method using the bias-corrected accelerated bootstrap method
Gamma ROS (BCA)	UCL based upon Gamma ROS method using the gamma approximate-UCL method
GOF, G.O.F.	goodness-of-fit
H-UCL	UCL based upon Land's H-statistic
HBK	Hawkins Bradu Kaas
HUBER	Huber estimation method
ID	identification code
IQR	interquartile range
K	Next K, Other K, Future K
KG	Kettenring Gnanadesikan
KM (%)	UCL based upon Kaplan-Meier estimates using the percentile bootstrap method
KM (Chebyshev)	UCL based upon Kaplan-Meier estimates using the Chebyshev inequality
KM (t)	UCL based upon Kaplan-Meier estimates using the Student's t-distribution cutoff value
KM (z)	UCL based upon Kaplan-Meier estimates using standard normal distribution cutoff value
K-M, KM	Kaplan-Meier
K-S, KS	Kolmogorov-Smirnov
LMS	least median squares
LN	lognormal distribution
Log-ROS Estimates	estimates based upon data set with extrapolated non-detect values obtained using robust ROS method
LPS	least percentile squares
MAD	Median Absolute Deviation
Maximum	Maximum value
MC	minimization criterion
MCD	minimum covariance determinant
MCL	maximum concentration limit
MD	Mahalanobis distance
Mean	classical average value
Median	Median value
Minimum	Minimum value
MLE	maximum likelihood estimate
MLE (t)	UCL based upon maximum likelihood estimates using Student's t-distribution cutoff value

MLE (Tiku)	UCL based upon maximum likelihood estimates using the Tiku's method
Multi Q-Q	multiple quantile-quantile plot
MVT	multivariate trimming
MVUE	minimum variance unbiased estimate
ND	non-detect or non-detects
NERL	National Exposure Research Laboratory
NumNDs	Number of Non-detects
NumObs	Number of Observations
OKG	Orthogonalized Kettenring Gnanadesikan
OLS	ordinary least squares
ORD	Office of Research and Development
PCA	principal component analysis
PCs	principal components
PCS	principal component scores
PLs	prediction limits
PRG	preliminary remediation goals
PROP	proposed estimation method
Q-Q	quantile-quantile
RBC	risk-based cleanup
RCRA	Resource Conservation and Recovery Act
ROS	regression on order statistics
RU	remediation unit
S	substantial difference
SD, *Sd, sd*	standard deviation
SLs	simultaneous limits
SSL	soil screening levels
S-W, SW	Shapiro-Wilk
TLs	tolerance limits
UCL	upper confidence limit
UCL95, 95% UCL	95% upper confidence limit
UPL	upper prediction limit
UPL95, 95% UPL	95% upper prediction limit
USEPA	United States Environmental Protection Agency
UTL	upper tolerance limit
Variance	classical variance
W*	Within groups matrix

WiB matrix	Inverse of W* cross-product B* matrix
WMW	Wilcoxon-Mann-Whitney
WRS	Wilcoxon Rank Sum
WSR	Wilcoxon Signed Rank
Wsum	Sum of weights
Wsum2	Sum of squared weights

Table of Contents

Chapter 9

Regression

The Regression module in Scout also offers most of the classical and robust multiple linear regression (including regression diagnostic methods) methods available in the current literature, similar to the Outlier/Estimates module. The multiple linear regression model with p explanatory (x-variables, leverage variables) variables is given by:

$$y_i = \left(x_{i1}b + x_{i2}b_2 + ...+ x_{ip}b_p\right)+ e_i .$$

The residuals, e_i, are assumed to be normally distributed as $N\left(0,s^2\right)$; i = 1, 2, ..., n.

The classical ordinary least square (OLS) method has a "0" break down point and can get distorted by the presence of even a single outlier, as in the classical mean vector and the covariance matrix.

Let $$x'_i = \left(x_{i1},x_{i2},...,x_{ip}\right),\ b' = \left(b_1,b_2,...,b_p\right).$$

The objective here is to obtain a robust and resistant estimate, $\hat{b}$, of b using the data set, $\left(y_i,x'_i\right)$; i = 1, 2, ..., n. The ordinary least squares (OLS) estimate, $\hat{b}_{OLS}$, of $\hat{b}$ is obtained by minimizing the residual sum of squares; namely, $\sum_{i=1}^{n} r_i^2$, where $r_i = y_i - x_i'\hat{b}_{OLS}$. Like the classical mean, the estimate, $\hat{b}_{OLS}$, of b has a "zero" break down point. This means that the estimate, $\hat{b}_{OLS}$, can take an arbitrarily aberrant value even by the presence of a single regression outlier (y-outlier) or leverage point (x-outlier), leading to a distorted regression model. The use of robust procedures that eliminate or dampen the influence of discordant observations on the estimates of regression parameters is desirable.

In regression applications, anomalies arising out of p-dimension space of the predictor variables, (e.g., due to unexpected experimental conditions), are called leverage points. Outliers in the response variable (e.g., due to unexpected outcomes, such as unusual reactions to a drug), are called regression or vertical outliers. The leverage outliers are divided into two categories: significant leverages ("bad" or inconsistent) and insignificant ("good" or consistent) points.

The identification of outliers in a data set and the identification of outliers in a regression model are two different problems. It is very desirable that a procedure distinguishes between good and bad outliers. In practice, in order to achieve high break down point, some methods (e.g., LMS method) fail to distinguish between good and bad leverage points.

In robust regression, the objective is twofold: 1) the identification of vertical (y-outliers, regression outliers) outliers and distinguishing between significant and insignificant leverage points, and 2) the estimation of regression parameters that are not influenced by the presence of the anomalies. The robust estimates should be in close agreement with classical OLS estimates when no outlying observations are present. Scout also offers several formal graphical displays of the regression and leverage results.

Scout provides several methods to obtain multiple linear regression models. Those available options include:

- **Ordinary Least Squares Regression** (OLS)
  Minimizes the least squared residuals.

- **Least Median/Percentile Squares Regression** (LMS/LPS)
  Minimizing the "**h**th" ordered squared residuals (Rousseeuw, 1984).

- **Biweight Regression**
  Conducted using Tukey's Biweight criterion (Beaton and Tukey, 1974).

- **Huber Regression**
  Conducted using Huber influence function (Huber, 1981).

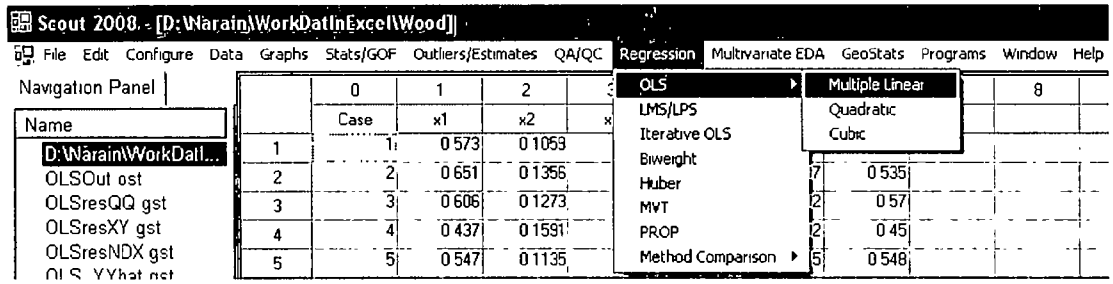
- **MVT Regression**
  Conducted using Multivariate Trimming Methods (Devlin et al., 1981).

- **PROP Regression**
  Conducted using PROP influence function (Singh and Nocerino, 1995).

Scout also provides the user with the option of identifying leverage outliers. If the leverage option is selected, then the outliers arising in the p-dimensional space of the predictor variables (X-space) are identified first. Those leverage points can be identified using various options available in Scout. The leverage points are identified using the same outlier methods as incorporated in the outlier module of Scout. The MDs for the leverage option are computed using the selected x-variables only. The weights obtained used in the leverage option are used at the initial regression option. The regression option is iterated some number of times to identify all of the regression outliers and bad leverage points. This process also distinguishes between good and bad leverage points.

9.1 Ordinary Least Squares (OLS) Linear Regression Method

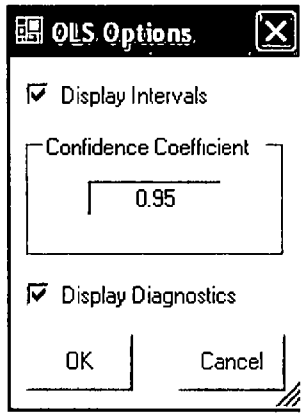
1. Click **Regression ▶ OLS ▶ Multiple Linear.**



2. The "**Select Variables**" screen (Section 3.3) will appear.

   - Select the dependent variable and one or more independent variables from the "**Select Variables**" screen.

   - Click on the "Options" button.



     - The "**Display Intervals**" check box will display the "**Summary Table for Prediction and Confidence Limits**" in the output sheet.

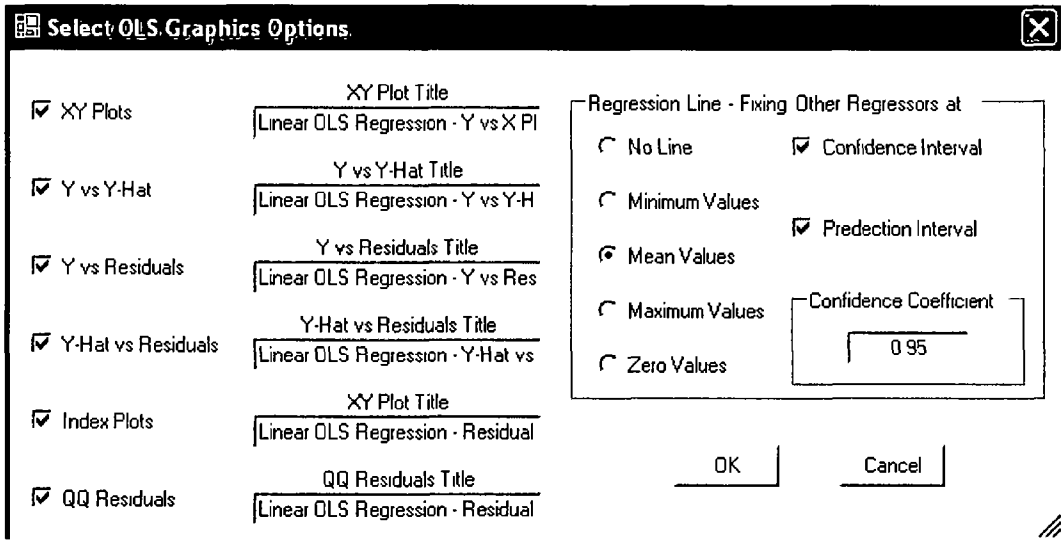
     - The "**Display Diagnostics**" check box will display the "**Regression Diagnostics Table**" and the "**Lack of Fit ANOVA Table**" (only if there are replicates in the independent variables).

     - Click "**OK**" to continue or "**Cancel**" to cancel the options.

   - If the results have to be produced by using a Group variable, then select a group variable by clicking the arrow below the "**Group by Variable**" button. This will result in a drop-down list of available variables. The

user should select and click on an appropriate variable representing a group variable.

- Click on the "**Graphics**" button and check all boxes.



  - A regression line can be drawn in the multivariate setting by choosing a single independent (regressor) variable and fixing other variables at the provided options using "**Regression Line – Fixing Other Regressors at**" option.

  - Specify the confidence or/and prediction band for the regression line using the "**Confidence Intervals**" and the "**Prediction Intervals**" check boxes.

  - Specify the "**Confidence Level**" for the bands.

  - Click "**OK**" to continue or "**Cancel**" to cancel the options.

- Click "**OK**" to continue or "**Cancel**" to cancel the OLS procedure.

Output for OLS Regression.
Data Set used: Wood (predictor variables p = 5).

	Ordinary Least Squares Linear Regression Analysis Output
Date/Time of Computation	10/30/2008 11.05:40 AM
User Selected Options	
From File	D:\Narain\WorkDatInExcel\Wood
Full Precision	OFF
Confidence Level for Intervals	0.95
lay Confidence and Prediction Limits	True
Display Regresion Diagnostics	True
Title for Residual QQ Plot	Linear OLS Regression - Residuals QQ Plot
Title Residual Index Plot	Linear OLS Regression - Residuals Index Plot
Title For Y vs X Plots	Linear OLS Regression - Y vs X Plot
Confidence Level for Regression Line	0 95
Display Confidence Band	True
Display Prediction Band	True
Title for Y-Hat vs Residuals Plot	Linear OLS Regression - Y-Hat vs Residuals Plot
Title for Y vs Residuals Plot	Linear OLS Regression - Y vs Residuals Plot
Title for Y vs Y-Hat Plot	Linear OLS Regression - Y vs Y-Hat Plot

Number of Observations	20
Dependent Variable	y
Number of Selected Regression Variables	5
Independant Variable	x1
Independant Variable	x2
Independant Variable	x3
Independant Variable	x4
Independant Variable	x5

Correlation Matrix

	y	x1	x2	x3	x4	x5
y	1	-0 145	0 611	0 47	-0 6	0 629
x1	-0.145	1	-0 246	-0 604	0 528	-0.641
x2	0 611	-0 246	1	0 388	-0.498	0 248
x3	0 47	-0.604	0 388	1	-0 24	0 659
x4	-0 6	0 528	-0 498	-0 24	1	-0 512
x5	0 629	-0 641	0 248	0 659	-0 512	1

Eigenvalues of Correlation Matrix							
Eval 1	Eval 2	Eval 3	Eval 4	Eval 5	Eval 6		
3 357	1.114	0.713	0.588	0 173	0.054		
		Sum of Eigenvalues	6				
Regression Estimates and Inference Table							
Paramater	DOF	Estimates	Std. Error	T-values	p-values	Tol Values	VIF
intercept	1	0.422	0.169	2.494	0.0253	N/A	N/A
x1	1	0.441	0.117	3.77	0.00222	0.27	3.701
x2	1	-1.475	0.487	-3.029	0.00931	0.264	3.786
x3	1	-0.261	0.112	-2.332	0.0339	0 583	1.715
x4	1	0.0208	0.161	0.129	0.388	0.299	3.346
x5	1	0.171	0.203	0.84	0.27	0.268	3.725
OLS ANOVA Table							
Source of Variation			**SS**	**DOF**	**MS**	**F-Value**	**P-Value**
Regression			0.0344	5	0.00687	11.81	0.0001
Error			0.00814	14	5.8158E-4		
Total			0.0425	19			
R Square				0 808			
Adjusted R Square				0.74			
Sqrt(MSE) = Scale				0.0241			
Regression Table							
Obs	Y Vector	Yhat	Residuals	Hat[i,i]	Res/Scale	Stude~ Res	
1	0.534	0.551	-0.0175	0.278	-0.725	-0 853	
2	0.535	0 534	0.00114	0.132	0.0472	0.0507	
3	0.57	0.54	0 03	0.22	1.243	1.407	
4	0.45	0.441	0.00855	0.258	0.355	0.412	
5	0.548	0.524	0.0242	0.222	1.002	1.137	
6	0.431	0 442	-0.0109	0.259	-0.452	-0.525	
7	0.481	0.459	0.0219	0.53	0.907	1.323	
8	0.423	0.424	-8 415E-4	0.289	-0.0349	-0.0414	
9	0.475	0 485	-0.00955	0.348	-0.396	-0.49	
10	0.486	0.496	-0.01	0 449	-0.415	-0.559	
11	0.554	0.506	0.0479	0.317	1.986	2.403	

Output for OLS Regression (continued).

Summary Table for Prediction and Confidence Limits

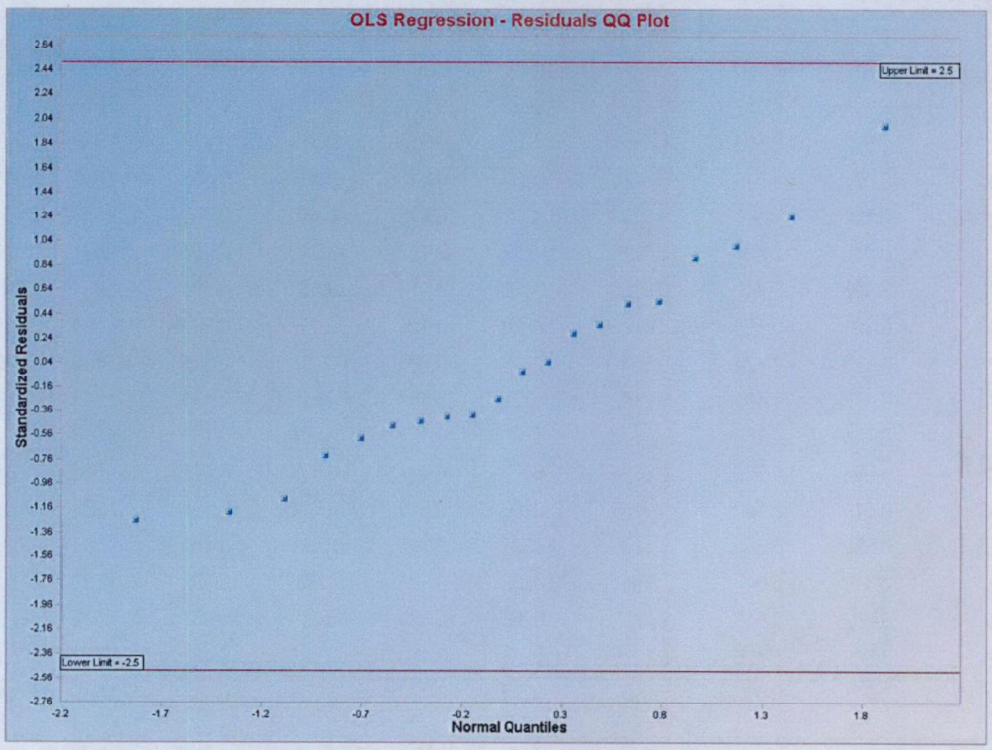
Obs	Y Vector	Yhat	s(Yhat)	s(pred)	LCL	UCL	LPL	UPL	Residuals
1	0 534	0.551	0 0127	0 0273	0 524	0.579	0.493	0.61	-0.0175
2	0.535	0 534	0 00875	0 0257	0 515	0.553	0.479	0 589	0.00114
3	0.57	0 54	0 0113	0 0266	0.516	0.564	0.483	0.597	0.03
4	0 45	0.441	0 0123	0.0271	0 415	0 468	0.383	0.499	0.00855
5	0 548	0 524	0.0114	0.0267	0.499	0 548	0.467	0.581	0.0242
6	0 431	0 442	0.0123	0.0271	0.416	0.468	0.384	0.5	-0 0109
7	0.481	0 459	0.0176	0 0298	0.421	0.497	0.395	0 523	0.0219
8	0.423	0.424	0.013	0.0274	0.396	0.452	0 365	0.483	-8.415E-4
9	0 475	0 485	0 0142	0 028	0 454	0.515	0.424	0.545	-0 00955
10	0.486	0 496	0 0162	0.029	0 461	0.531	0.434	0.558	-0 01
11	0 554	0 506	0 0136	0.0277	0 477	0 535	0 447	0.565	0 0479
12	0 519	0 548	0 0154	0.0286	0 515	0 581	0 486	0.609	-0.0289
13	0 492	0 504	0.0129	0.0274	0.476	0 531	0.445	0.562	-0 0117
14	0 517	0 547	0 00866	0 0256	0.529	0 566	0 492	0.602	-0.0304
15	0.502	0 516	0 00941	0 0259	0.496	0 536	0.461	0 572	-0.0141
16	0.508	0 495	0.0175	0 0298	0 458	0.533	0.431	0.559	0.0126
17	0.52	0 526	0 013	0.0274	0.498	0.554	0.467	0.585	-0.00615
18	0.506	0.499	0 0131	0.0274	0 471	0.527	0.44	0.558	0.00685
19	0 401	0 427	0 013	0.0274	0.399	0.455	0.368	0.486	-0 0261
20	0 568	0 555	0 0136	0.0277	0.526	0.584	0.495	0.614	0.0131

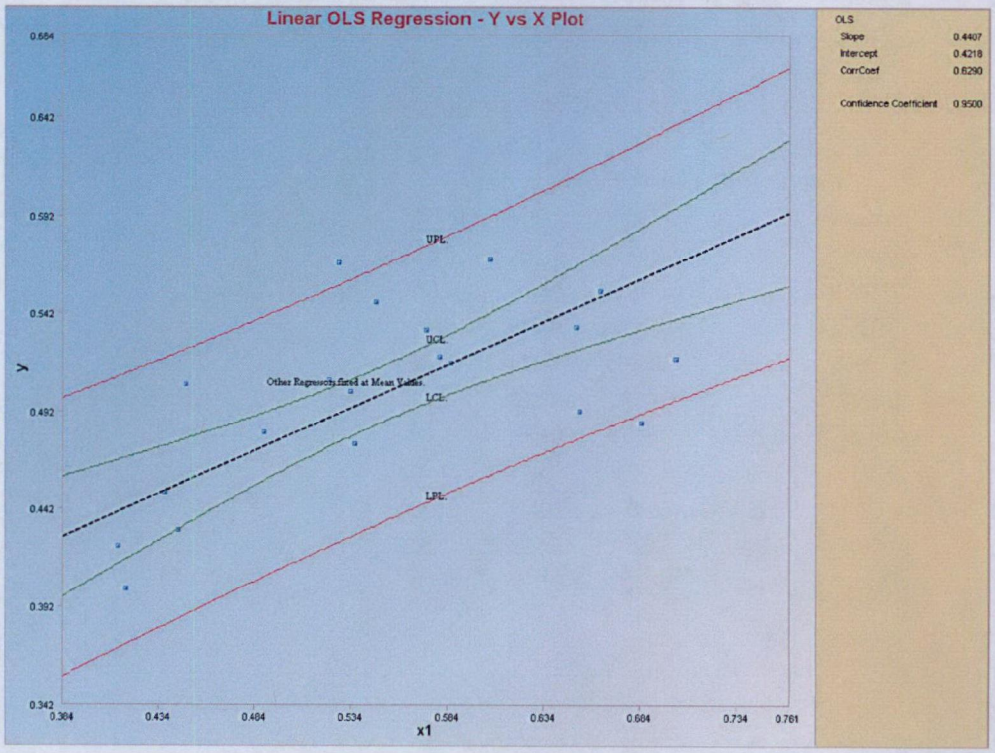
No replicates in the data - Lack of Fit ANOVA Table not displayed

Regression Diagnostics Table

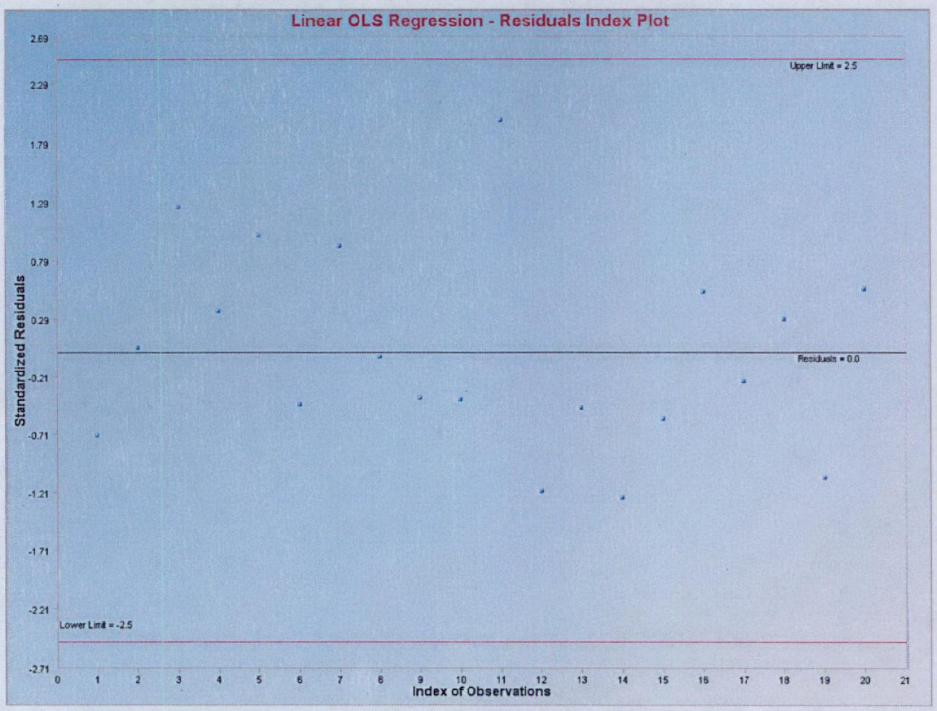
Obs. #	Residuals	H(i,i)	CD(i)	t(i)	DFFITS
1	-0.0175	0 278	0.0466	-0 876	-0.543
2	0 00114	0 132	6 4894E-5	0 0507	0 0197
3	0 03	0 22	0.0929	1.518	0 806
4	0.00855	0 258	0 00985	0.414	0.245
5	0 0242	0.222	0.0616	1.193	0 638
6	-0.0109	0 259	0 0161	-0.53	-0.314
7	0.0219	0.53	0.329	1.414	1.502
8	-8.415E-4	0 289	1.1582E-4	-0.0414	-0 0264
9	-0 00955	0 348	0.0214	-0 495	-0 362
10	-0.01	0.449	0 0425	-0.566	-0 511

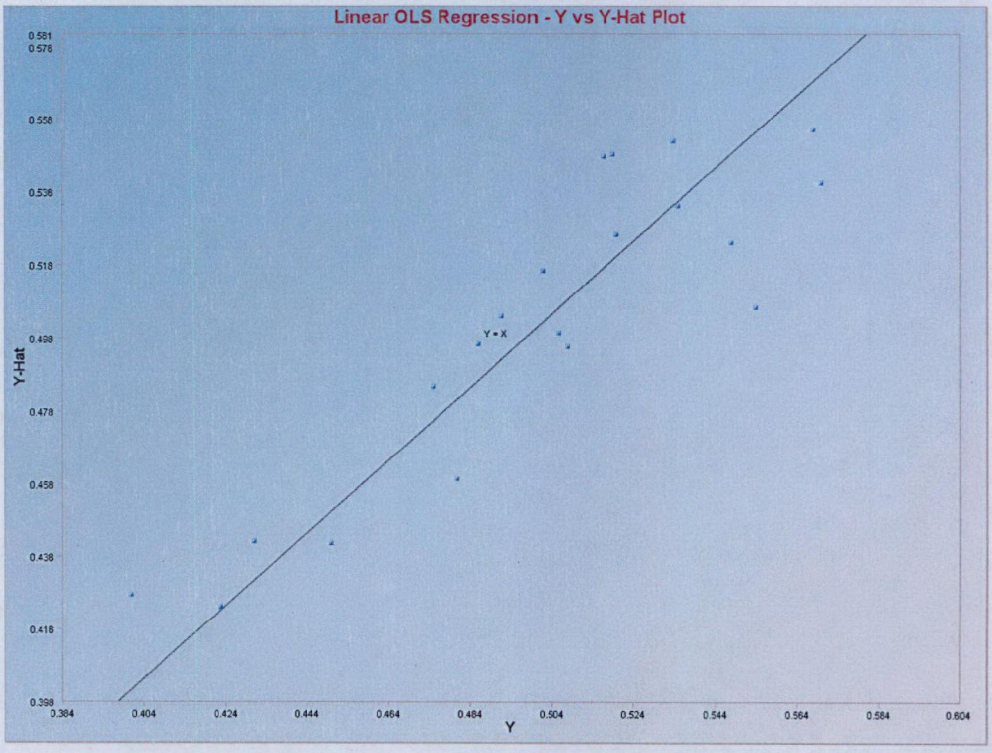
Output for OLS Regression (continued).



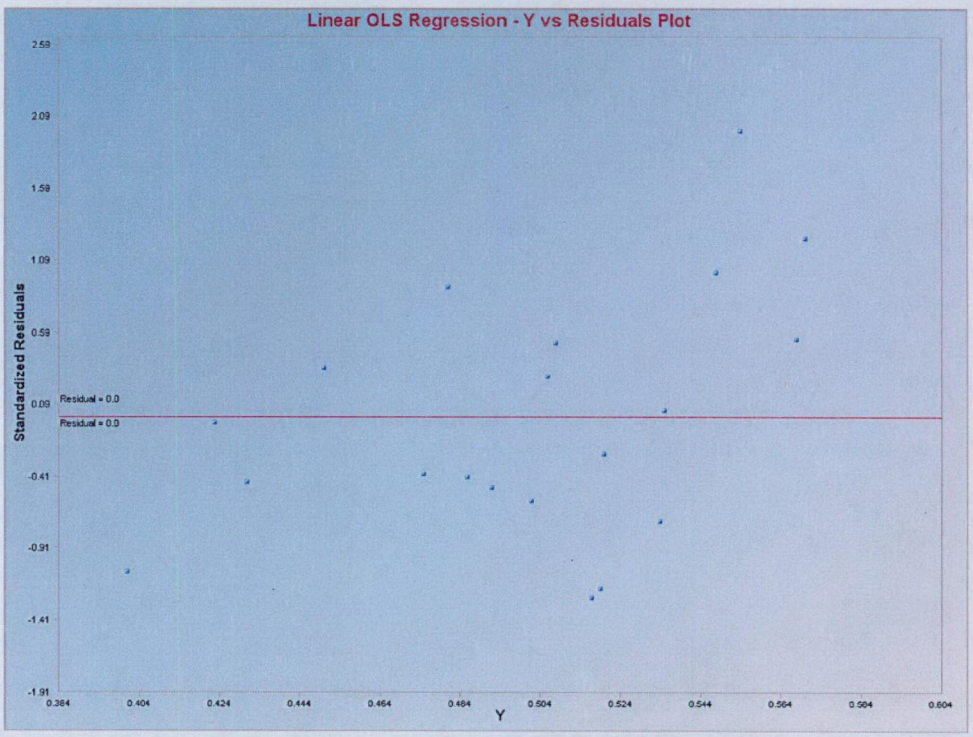


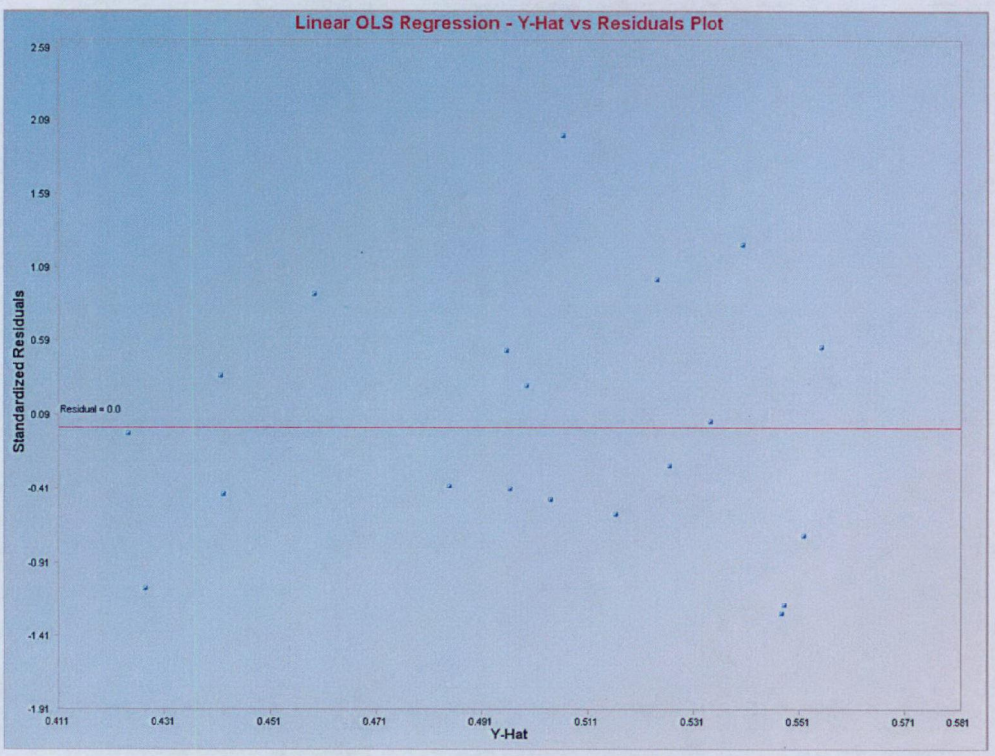
Output for OLS Regression (continued).





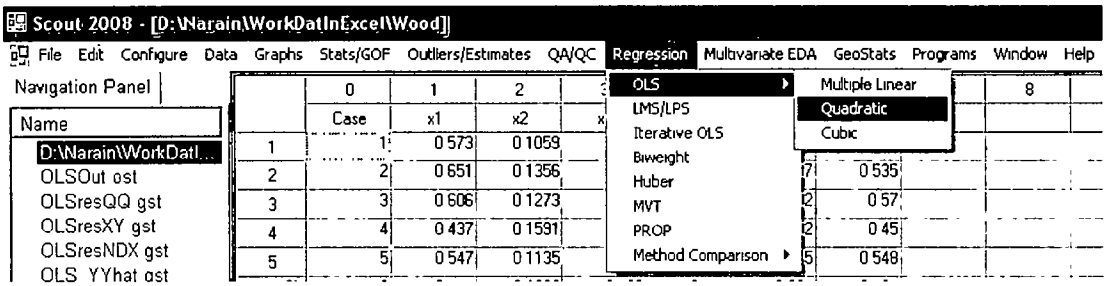
Output for OLS Regression (continued).





9.2 OLS Quadratic/Cubic Regression Method

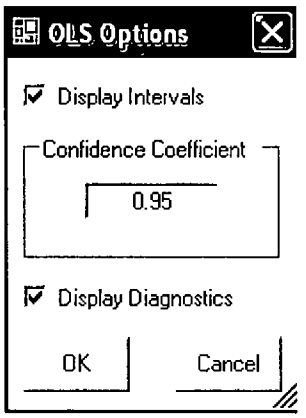
1. Click **Regression ▶ OLS ▶ Quadratic or Cubic.**



2. The "**Select Variables**" screen (Section 3.3) will appear.

   - Select the dependent variable and one or more independent variables from the "**Select Variables**" screen.

   - Click on the "Options" button.



     - The "**Display Intervals**" check box will display the "**Summary Table for Prediction and Confidence Limits**" in the output sheet.

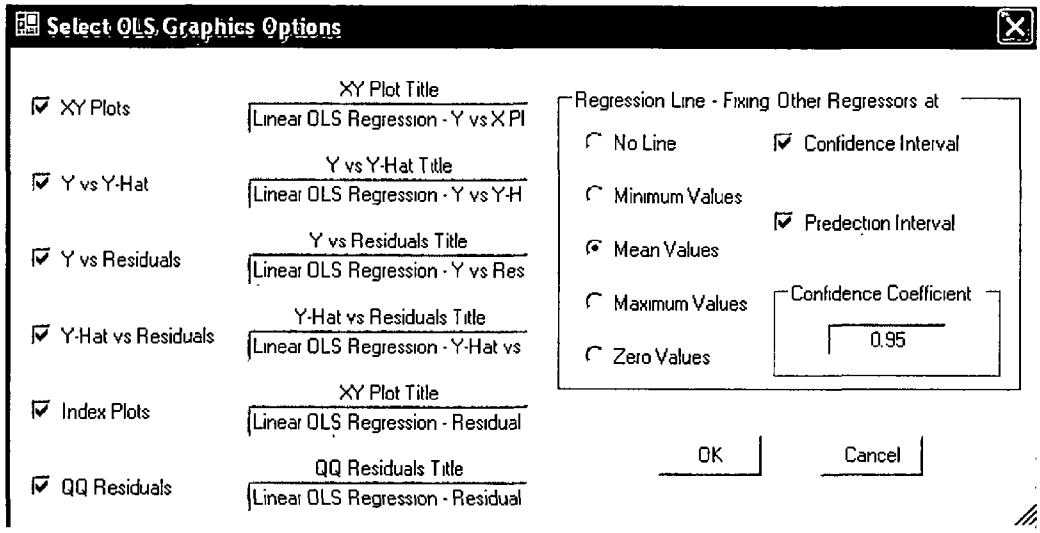
     - The "**Display Diagnostics**" check box will display the "**Regression Diagnostics Table**" and the "**Lack of Fit ANOVA Table**" (only if there are replicates in the independent variables).

     - Click "**OK**" to continue or "**Cancel**" to cancel the options.

   - If the results have to be produced by using a Group variable, then select a group variable by clicking the arrow below the "**Group by Variable**" button. This will result in a drop-down list of available variables. The

user should select and click on an appropriate variable representing a group variable.

- Click on the "**Graphics**" button and check all boxes.



  - "**Regression Line – Fixing Other Regressors at**" option is not used in this quadratic regression module.

  - Specify the confidence or/and prediction band for the regression line using the "**Confidence Intervals**" and the "**Prediction Intervals**" check boxes.

  - Specify the "**Confidence Level**" for the bands.

  - Click "**OK**" to continue or "**Cancel**" to cancel the options.

- Click "**OK**" to continue or "**Cancel**" to cancel the OLS procedure.

Output for OLS Regression.
Data Set used: Wood (predictor variables p = 5).

	Ordinary Least Squares Quadratic Regression Analysis Output
Date/Time of Computation	10/30/2008 1.14.57 PM
User Selected Options	
From File	D.\Narain\WorkDatInExcel\Wood
Full Precision	OFF
Confidence Level for Intervals	0.95
ay Confidence and Prediction Limits	True
Display Regresion Diagnostics	True
Residual QQ Plot	Not Selected
Residual Index Plot	Not Selected
Title For Y vs X Plots	Quadratic OLS Regression - Y vs X Plot
onfidebce Level for Regression Line	0.95
Display Confidence Band	True
Display Prediction Band	True
Y vs Residuals Plot	Not Selected
Y vs Residuals Plot	Not Selected
Y vs Y-Hat Plot	Not Selected

Number of Observations	20
Dependent Variable	y
Number of Selected Regression Variables	1
Independant Variable	x1

Correlation Matrix

	y	x1	Squared
y	1	0.997	0 629
x1	0 997	1	0.588
Squared	0.629	0.588	1

Eigenvalues of Correlation Matrix

Eval 1	Eval 2	Eval 3	
2.493	0.505	0.0015	
		Sum of Eigenvalues	3

Output for OLS Regression (continued).

Regression Estimates and Inference Table							
Paramater	DOF	Estimates	Std. Error	T-values	p-values	Tol Values	VIF
intercept	1	-0.643	0 26	-2.47	0.0252	N/A	N/A
x1	1	3 906	0 957	4.08	9.1497E-4	0.00573	174 6
Squared	1	-3.237	0 863	-3.75	0 00185	0.00573	174.6

OLS ANOVA Table					
Source of Variation	**SS**	**DOF**	**MS**	**F-Value**	**P-Value**
Regression	0.0284	2	0 0142	17.2	0.0001
Error	0.0141	17	8.2683E-4		
Total	0 0425	19			
R Square		0.669			
Adjusted R Square		0.63			
Sqrt(MSE) = Scale		0.0288			

Regression Table						
Obs	Y Vector	Yhat	Residuals	Hat[i,i]	Res/Scale	Stude~ Res
1	0 534	0.532	0 00157	0 103	0.0545	0.0576
2	0 535	0.528	0.00697	0 116	0.242	-0.258
3	0 57	0.535	0.0346	0.0923	1 204	1.264
4	0 45	0.446	0.00411	0.16	0 143	0.156
5	0 548	0.525	0 0229	0.106	0 795	0 84
6	0.431	0.453	-0 0223	0.137	-0.774	-0 833
7	0 481	0.493	-0.0121	0.0873	-0 421	-0.441
8	0 423	0.418	0.00481	0.293	0 167	0.199
9	0.475	0.521	-0 0457	0.103	-1.591	-1.68
10	0 486	0 514	-0 0278	0.247	-0 966	-1.114
11	0.554	0 523	0.0305	0.149	1.062	1.151
12	0.519	0.503	0.0158	0.391	0.549	0.703
13	0.492	0.527	-0 0354	0.12	-1 231	-1 313
14	0.517	0 534	-0 0174	0 0993	-0 606	-0.639
15	0.502	0 52	-0.0179	0.103	-0 621	-0.656
16	0.508	0.515	-0.00653	0 099	-0 227	-0 239
17	0.52	0 534	-0.0136	0.101	-0.475	-0.501
18	0.506	0 457	0 0487	0.126	1.692	1.81
19	0.401	0.423	-0.0221	0.264	-0 767	-0 895
20	0.568	0.517	0.0509	0.101	1.772	1.869

Output for OLS Regression (continued).

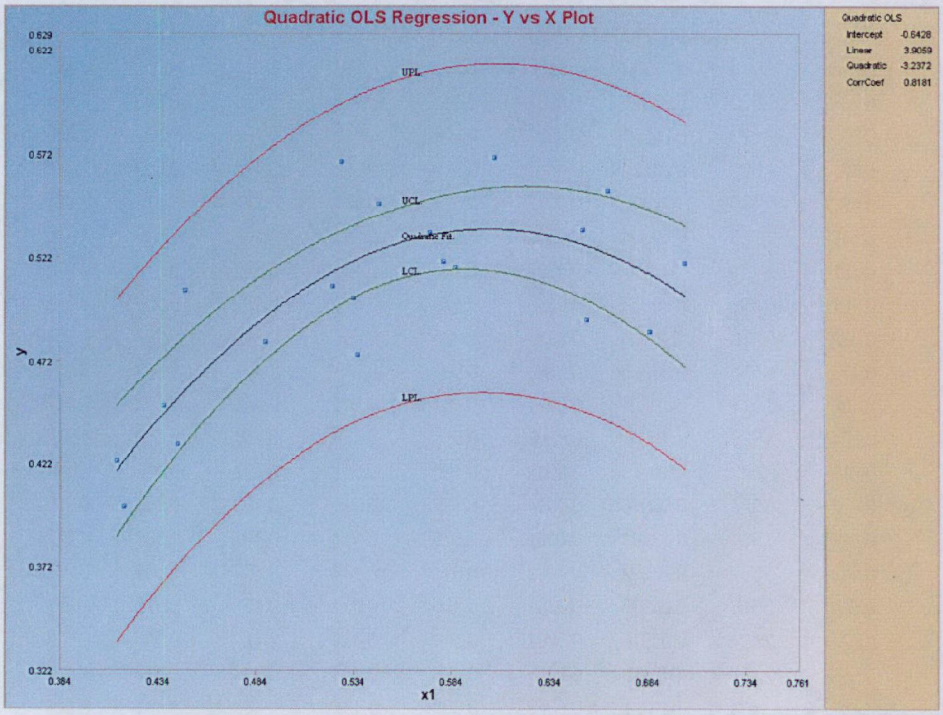
Summary Table for Prediction and Confidence Limits									
Obs	Y Vector	Yhat	s(Yhat)	s(pred)	LCL	UCL	LPL	UPL	Residuals
1	0 534	0 532	0.00925	0 0302	0 513	0 552	0.469	0 596	0 00157
2	0 535	0 528	0 00981	0 0304	0 507	0 549	0 464	0 592	0 00697
3	0.57	0 535	0 00874	0 0301	0 517	0.554	0 472	0.599	0 0346
4	0.45	0 446	0 0115	0 031	0.422	0.47	0.381	0.511	0 00411
5	0.548	0 525	0 00934	0 0302	0 505	0 545	0.461	0 589	0 0229
6	0 431	0 453	0 0106	0 0307	0 431	0 476	0 389	0 518	-0.0223
7	0 481	0 493	0 0085	0 03	0 475	0 511	0 43	0 556	-0 0121
8	0.423	0 418	0 0156	0 0327	0 385	0.451	0.349	0.487	0 00481
9	0 475	0 521	0 00925	0 0302	0.501	0.54	0.457	0.584	-0 0457
10	0.486	0 514	0 0143	0 0321	0.484	0.544	0.446	0.582	-0 0278
11	0.554	0 523	0 0111	0 0308	0 5	0 547	0 458	0 589	0 0305
12	0 519	0 503	0 018	0.0339	0.465	0.541	0.432	0.575	0 0158
13	0 492	0 527	0.00998	0 0304	0 506	0 548	0 463	0.592	-0 0354
14	0.517	0.534	0.00906	0 0301	0.515	0.554	0.471	0.598	-0 0174
15	0.502	0.52	0.00922	0.0302	0 5	0 539	0.456	0 584	-0.0179
16	0 508	0 515	0 00905	0 0301	0 495	0 534	0 451	0 578	-0.00653
17	0 52	0 534	0 00916	0 0302	0 514	0 553	0 47	0 597	-0 0136
18	0 506	0 457	0.0102	0 0305	0 436	0 479	0 393	0.522	0 0487
19	0.401	0 423	0 0148	0 0323	0 392	0 454	0 355	0.491	-0 0221
20	0.568	0 517	0.00913	0 0302	0 498	0 536	0 453	0 581	0 0509

No replicates in the data - Lack of Fit ANOVA Table not displayed

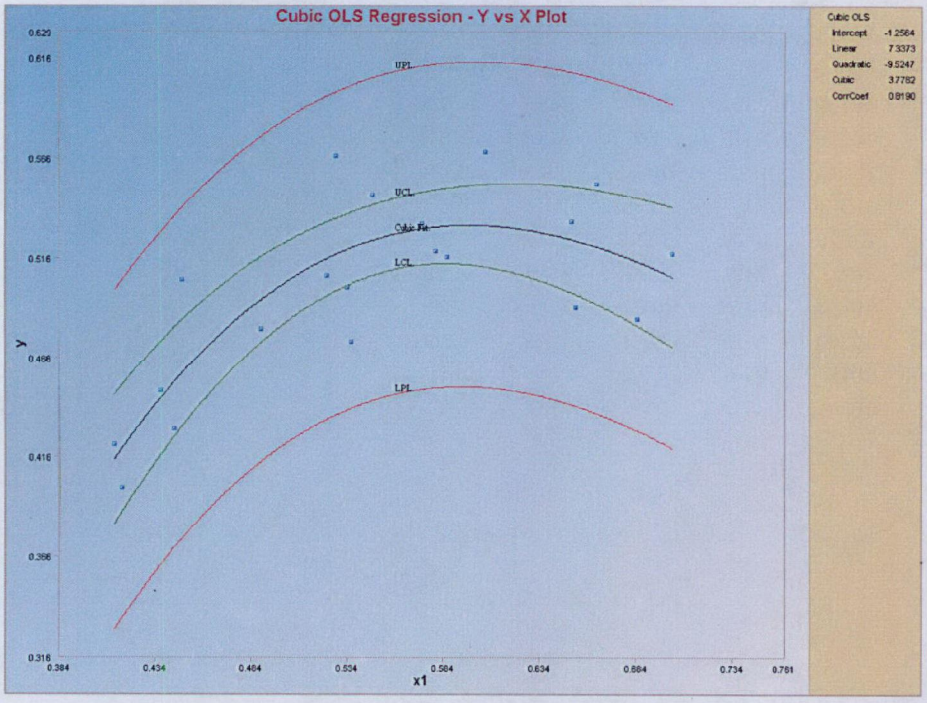
Regression Diagnostics Table

Obs. #	Residuals	H[i,i]	CD[i]	t[i]	DFFITS
1	0 00157	0 103	1 2758E-4	0 0576	0 0196
2	0 00697	0.116	0.00292	0 258	0 0938
3	0 0346	0 0923	0 0542	1 328	0 423
4	0 00411	0 16	0 00154	0.156	0 0681
5	0.0229	0 106	0 0278	0 858	0 295
6	-0 0223	0 137	0 0366	-0 851	-0 339
7	-0 0121	0 0873	0.00621	-0 444	-0 137
8	0 00481	0 293	0 00549	0 199	0.128
9	-0 0457	0.103	0 109	-1 84	-0 625
10	-0 0278	0 247	0 136	-1 157	-0 663

Output for OLS Regression (continued) – Quadratic Fit.



Output for OLS Regression (continued) – Cubic Fit.



9.3 Least Median/Percentile Squares (LMS/LPS) Regression Method

Break Down Point of LMS Regression Estimates

The break down (BD) points for LMS (k~0.5) and least percentile of squared residuals (LPS, k>0.5) regression methods as incorporated in Scout are summarized in the following table. Note that, LMS is labeled as LPS when k>0.5. In the following the fraction, k is given by $0.5 \leq k < 1$. For example, for median, the fraction, k =0.5, for 75th percentile, fraction, k = 0.75, and so forth.

Approximate Break Down Point for LMS or LPS Regression Estimates

No. of Explanatory Vars., p = 1

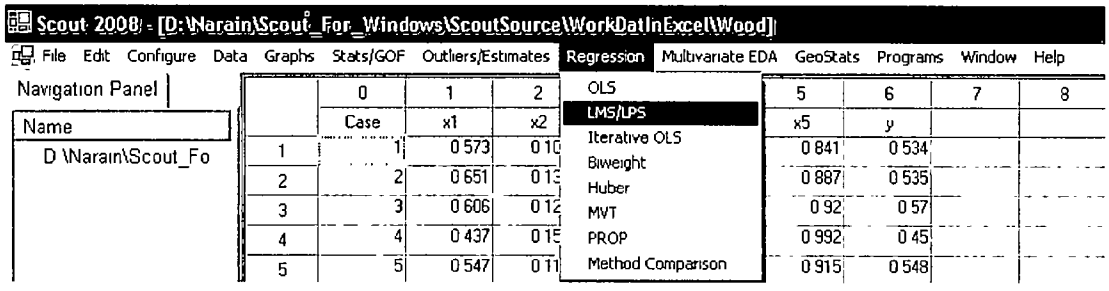
Minimizing Squared Residual	**BD**
Pos = [n/2], k = 0.5	(n-Pos)/n
Pos = [(n+1)/2]	(n-Pos)/n
Pos = [(n+p+1)/2]	(n-Pos)/n
LPS ~ Pos = [n*k], k > 0.5	(n-Pos)/n

No. of Explanatory Vars., p > 1

Minimizing Squared Residual	**BD**
Pos = [n/2], k = 0.5	(n-Pos-p+2)/n
Pos = [(n+1)/2]	(n-Pos-p+2)/n
Pos = [(n+p+1)/2]	(n-Pos-p+2)/n
LPS ~ Pos = [n*k], k > 0.5	(n-Pos-p+2)/n

Here [x] = greatest integer contained in x, and k represents a fraction: $0.5 \leq k < 1$. Pos stands for position/index of an entry in ordered array (of size n) of squared residuals. The squared residual at position, Pos is being minimized. For example, when Pos = [n/2], the median of squared residuals is being minimized.

1. Click **Regression ▷ LMS.**

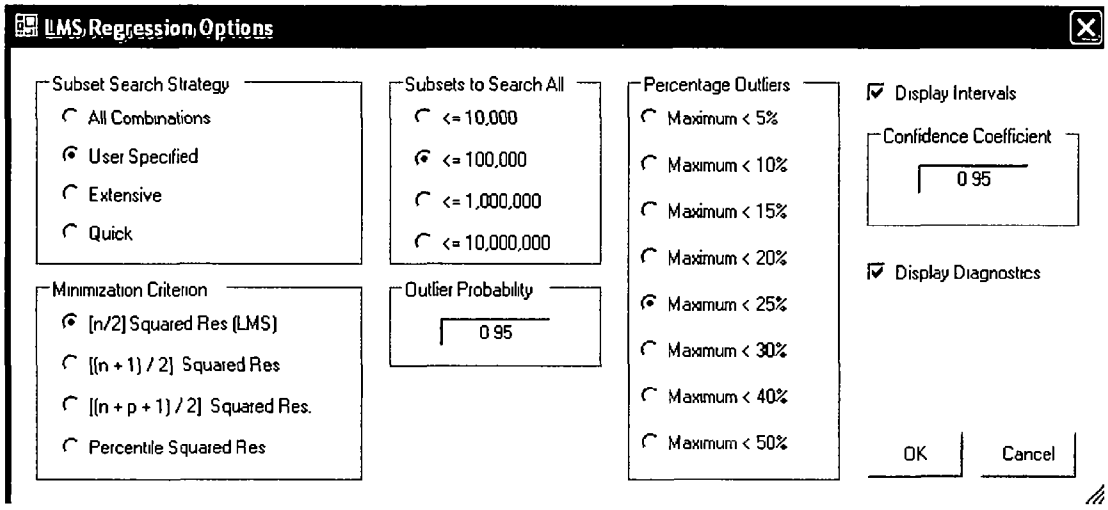


2. The "**Select Variables**" screen (Section 3.3) will appear.

- Select the dependent variable and one or more independent variables from the "**Select Variables**" screen.

- If the results have to be produced by using a Group variable, then select a group variable by clicking the arrow below the "**Group by Variable**" button. This will result in a drop-down list of available variables. The user should select and click on an appropriate variable representing a group variable.

- Click on the **"Options"** button to get the options window and then click on "**User Specified**" in "**Subset Search Strategy**" box.



*Note: The **Subset Search Strategy** allows the user to specify the number of initial subsets of size p+1 to be used to obtain the residuals (regression models) from a total of* $\binom{n}{p}$ *subsets. The user can specify the **Percentage of Outliers**, **Outlier Probability** (usually closer to 1), and the **Minimization Criterion** (order of the squared residual to minimize) (Leroy and Rousseeuw, 1987).*

- Specify "**Subsets to Search.**" The default is "**<=100,000.**"

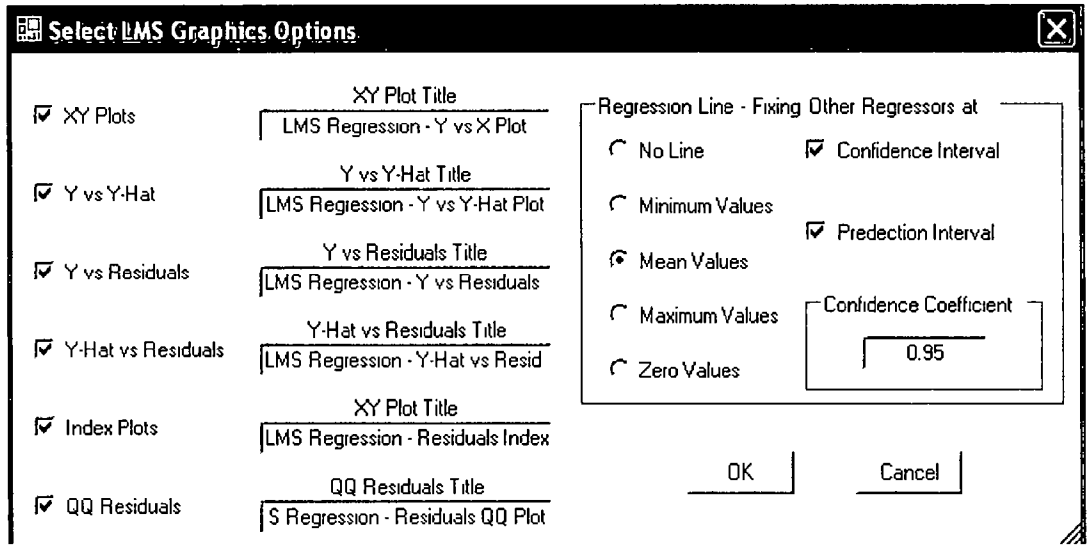
- Specify "**Percentage Outliers.**" The default is "**< 25%.**"

- Specify "**Outlier Probability.**" The default is "**0.95.**"

- Specify "**Minimization Criterion.**" The default is "**Median Squared Residual.**"

- Click on "**OK**" to continue or "**Cancel**" to cancel the options.

- Click on "**Graphics**" for the graphics options and specify the preferred graphs.



  - Specify the required graphs and the input parameters.
  - Click on "**OK**" to continue or "**Cancel**" to cancel the options.
- Click on "**OK**" to continue or "**Cancel**" to cancel the computations.

Output example: The data set "**WOOD.xls**" was used for LMS regression. It has 5 predictor variables (p) and 20 observations. A total of 38760 subsets of size p+1 (6) observations were used find the best subset meeting the minimization criterion of least median of squared residuals.

Output for LMS Regression.

Data Set used: Bradu (predictor variables p = 5, Minimization Criterion = Median Squared Residuals).

Least Median Squared (LMS) Regression Analysis Output	
Date/Time of Computation	3/4/2008 9.35 11 AM
User Selected Options	
From File	D \Narain\Scout_For_Windows\ScoutSource\WorkDatInExcel\Wood
Full Precision	OFF
Subset Search Strategy	User Specified Criteria
Percentage Outliers	Maximum Outliers <= 0 25
Percentage Outliers	Outlier Probability <= 0.95
Search All Cutoff	Do all combinations if <= 100000
Minimization Criterion	Median of Squared Residuals
Title for Residual QQ Plot	LMS Regression - Residuals QQ Plot
Residual Index Plot	Not Selected
Y vs X Plots	Not Selected
Title for Y-Hat vs Residuals Plot	LMS Regression - Y-Hat vs Residuals Plot
Y vs Residuals Plot	Not Selected
Y vs Y-Hat Plot	Not Selected

Number of Selected Regression Variables	5
Number of Observations	20
Dependent Variable	y

Correlation Matrix

	y	x1	x2	x3	x4	x5
y	1	-0 145	0 611	0 47	-0 6	0 629
x1	-0.145	1	-0 246	-0 604	0 528	-0 641
x2	0 611	-0 246	1	0 388	-0 498	0 248
x3	0 47	-0 604	0 388	1	-0 24	0 659
x4	-0 6	0 528	-0 498	-0 24	1	-0 512
x5	0 629	-0 641	0 248	0 659	-0 512	1

Eigenvalues of Correlation Matrix

Eval 1	Eval 2	Eval 3	Eval 4	Eval 5	Eval 6
0 054	0.173	0 588	0.713	1.114	3 357

OLS Estimates of Regression Parameters

Intercept	x1	x2	x3	x4	x5
0 422	0.441	-1 475	-0.261	0 0208	0.171

Output for LMS Regression (continued).

Stdv of Estimated Regression Parameters								
Intercept	x1	x2	x3	x4	x5			
0.169	0.117	0.487	0.112	0.161	0 203			

OLS ANOVA Table					
Source of Variation	**SS**	**DOF**	**MS**	**F-Value**	**P-Value**
Regression	0 0344	5	0.00687	11.81	0 0001
Error	0.00814	14	5 8158E-4		
Total	0.0425	19			
OLS Scale Estimate				0.0241	
R Square				0.808	

Least Median of Squared Residual Regression

Total Number of Elemental Subsets of size (6)	38760
Total Number of Elemental Subsets of size (6) Searched	38760
Number of Non-Singular Elemental Subsets of size (6)	38760

Best Elemental Subset of size 6 Found						
	y	x1	x2	x3	x4	x5
Obs # 7	0.481	0.489	0.123	0.562	0.455	0.824
Obs # 10	0.486	0.685	0.156	0.631	0.564	0.914
Obs # 11	0.554	0 664	0 159	0.506	0.481	0.867
Obs # 12	0.519	0 703	0 134	0.519	0.484	0.812
Obs # 15	0.502	0.534	0.114	0.521	0 57	0.889
Obs # 16	0 508	0.523	0.132	0.505	0.612	0.919

Best Subset satisfies minimization criterion.

LMS Estimates of Regression Parameters (Using Best Subset)					
Intercept	x1	x2	x3	x4	x5
0.37	0.172	-0 073	-0 524	-0.441	0.644

Stdv of Estimated Regression Parameters (Using Best Subset)					
Intercept	x1	x2	x3	x4	x5
0.874	0 604	2 516	0.579	0.832	1.051

Output for LMS Regression (continued).

Minimizing 10th Ordered Squared Residual	
Value of Minimum Criterion	2.0999E-6
Approximate Breakdown Value	0.35
Unweighted Sigma Estimate based upon LMS Residuals	0.125
Initial Robust LMS Scale Estimate (Adjusted for dimensionality)	0 00691

LMS Regression Table Based Upon Best Subset

Obs #	Y	Yhat	Residuals	Hat[i,i]	Res/Sigma	Student	Res/Scale	Weights	C Res~cale
1	0 534	0 522	0 0122	0.278	0.0976	0 115	1 761	1	-0.725
2	0.535	0 527	0.00787	0.132	0.0632	0 0678	1.139	1	0.0472
3	0.57	0 569	8.8370E-4	0.22	0.00709	0.00803	0.128	1	1 243
4	0.45	0 652	-0 202	0.258	-1 625	-1 887	-29.31	0	0 355
5	0.548	0.538	0 00972	0.222	0.078	0 0885	1.407	1	1.002
6	0.431	0.662	-0 231	0 259	-1.857	-2.158	-33 5	0	-0.452
7	0.481	0 481	-3.43E-14	0.53	-2.75E-13	-4.02E-13	-4.97E-12	1	0.907
8	0.423	0 65	-0.227	0 289	-1 824	-2 163	-32.91	0	-0.0349
9	0.475	0 489	-0 0141	0.348	-0 113	-0.14	-2.037	1	-0 396
10	0.486	0.486	-4 35E-14	0.449	-3.49E-13	-4 71E-13	-6.30E-12	1	-0.415
11	0.554	0 554	-1.14E-14	0 317	-9 18E-14	-1 11E-13	-1.66E-12	1	1 986
12	0.519	0.519	-3.52E-14	0.41	-2 82E-13	-3 68E-13	-5.09E-12	1	-1.198
13	0.492	0.491	0.00145	0.287	0 0116	0.0138	0.21	1	-0.485
14	0.517	0.522	-0.00472	0.129	-0.0379	-0 0406	-0.684	1	-1.261
15	0.502	0.502	-6 55E-14	0 152	-5 26E-13	-5 71E-13	-9.48E-12	1	-0 587
16	0.508	0.508	-7.33E-14	0.526	-5 88E-13	-8.54E-13	-1.06E-11	1	0.524
17	0.52	0.521	-0 00114	0 289	-0.00913	-0 0108	-0.165	1	-0.255
18	0.506	0 522	-0 0161	0 294	-0 129	-0.154	-2.329	1	0 284
19	0.401	0.666	-0.265	0.292	-2 129	-2.53	-38.4	0	-1.084
20	0.568	0.567	9.6663E-4	0.318	0.00776	0 00939	0.14	1	0.545

Reweighted LMS Estimates of Regression Parameters

Intercept	x1	x2	x3	x4	x5
0 377	0.217	-0 085	-0.564	-0.4	0.607

Reweighted LMS Stdv of Estimated Regression Parameters

Intercept	x1	x2	x3	x4	x5
0 054	0 0421	0.198	0.0435	0.0654	0.0786

Output for LMS Regression (continued).

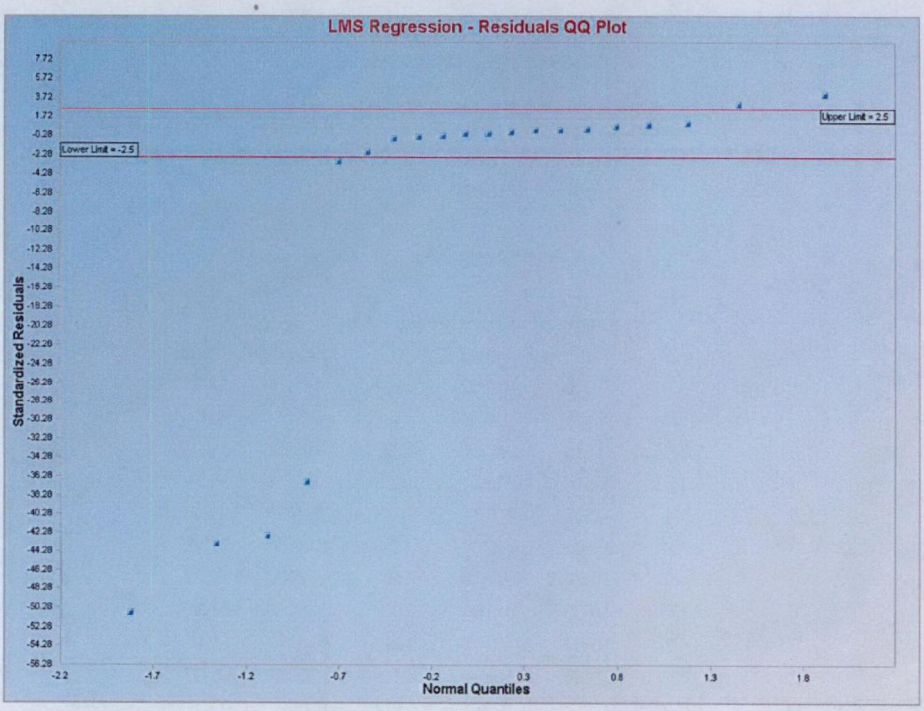
Reweighted LMS ANOVA Table					
Source of Variation	SS	DOF	MS	F-Value	P-Value
Regression	0 0128	5	0.00255	46	0 0000
Error	5.5517E-4	10	5 5517E-5		
Total	0 0133	15			
			R Square	0 958	
			Final Reweighted LMS Scale Estimate	0 00745	

Reweighted LMS Regression Table						
Obs #	Y	Yhat	Residuals	Hat(i,i)	Student	Res/Scale
1	0 534	0 526	0 00802	0 278	1 267	1 076
2	0 535	0 531	0 00444	0 132	0.639	0 595
3	0 57	0 57	2 3614E-4	0 22	0 0359	0 0317
4	0.45	0 64	-0 19	0 258	-29 67	-25.55
5	0 548	0 535	0 013	0 222	1 979	1 745
6	0.431	0 651	-0 22	0 259	-34 32	-29 54
7	0.481	0.474	0 00658	0 53	1 288	0 883
8	0 423	0 639	-0 216	0 289	-34.37	-28.98
9	0 475	0 493	-0 00775	0 348	-1.288	-1 04
10	0 486	0 486	-2 958E-4	0 449	-0 0535	-0.0397
11	0 554	0 557	-0 00274	0 317	-0 445	-0 368
12	0 519	0 525	-0 00642	0 41	-1 122	-0.862
13	0 492	0 489	0.00319	0 287	0 507	0 428
14	0 517	0 524	-0.00712	0.129	-1.023	-0 955
15	0 502	0 502	4 6552E-4	0 152	0 0679	0 0625
16	0 508	0 508	-7 691E-5	0 526	-0.015	-0 0103
17	0 52	0 521	-7 971E-4	0 289	-0.127	-0.107
18	0 506	0.515	-0 00928	0 294	-1 482	-1 246
19	0 401	0 655	-0 254	0 292	-40 5	-34 07
20	0.568	0 569	-0 00145	0 318	-0 235	-0 194

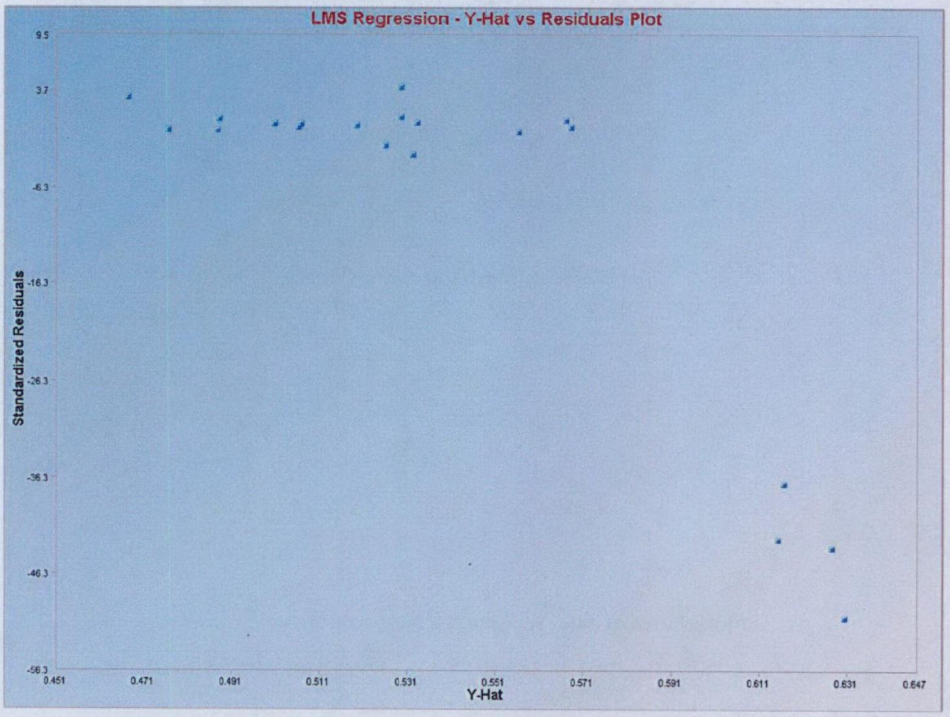
Final Weighted Correlation Matrix						
	y	x1	x2	x3	x4	x5
y	1	0 75	0 271	-0 0958	-0.173	0 147
x1	0.75	1	0 497	-0.132	-0.0367	-0 097
x2	0 271	0 497	1	-0.226	-0 0031	-0 755
x3	-0 0958	-0 132	-0.226	1	0 733	0 138
x4	-0 173	-0.0367	-0.0031	0 733	1	0 245
x5	0 147	-0 097	-0 755	0 138	0 245	1

Eigenvalues of Final Weighted Correlation Matrix					
Eval 1	Eval 2	Eval 3	Eval 4	Eval 5	Eval 6
0 012	0.196	0.39	1.431	1.604	2 368

Output for LMS Regression (continued).



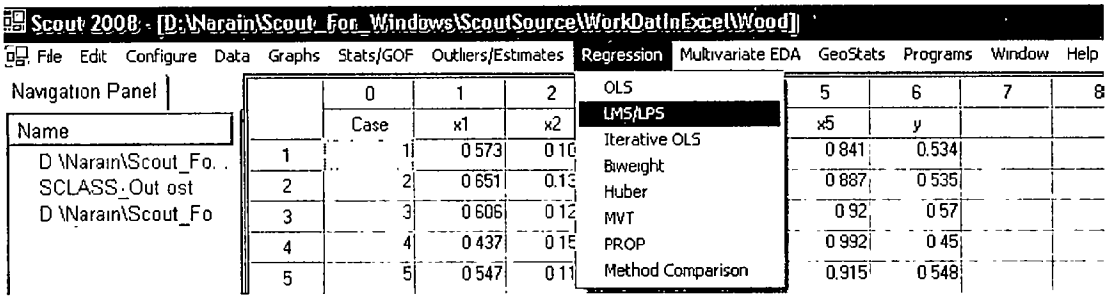
Output for LMS Regression (continued).



Interpretation of Graphs: Observations which are outside of the horizontal lines in the graph are considered to be regression outliers.

9.3.1 Least Percentile of Squared Residuals (LPS) Regression

1. Click **Regression ▶ LMS/LPS.**

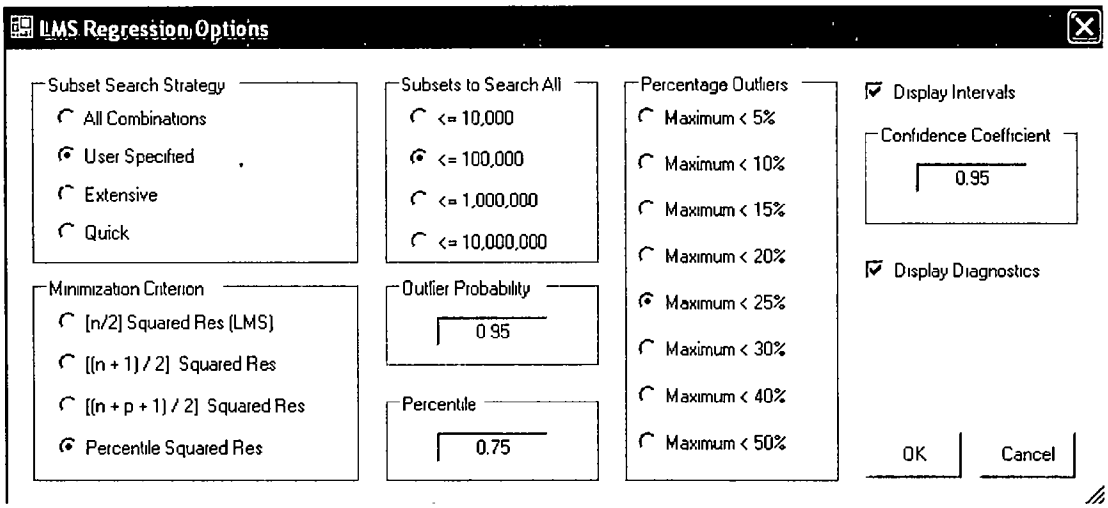


2. The "**Select Variables**" screen (Section 3.3) will appear.

   - Select the dependent variable and one or more independent variables from the "**Select Variables**" screen.

   - If the results have to be produced by using a Group variable, then select a group variable by clicking the arrow below the "**Group by Variable**" button. This will result in a drop-down list of available variables. The user should select and click on an appropriate variable representing a group variable.

   - Click on the "**Options**" button to get the options window and then click on "**User Specified**" in "**Subset Search Strategy**" box.



   - Specify "**Subsets to Search All.**" The default is "**<=100,000.**"

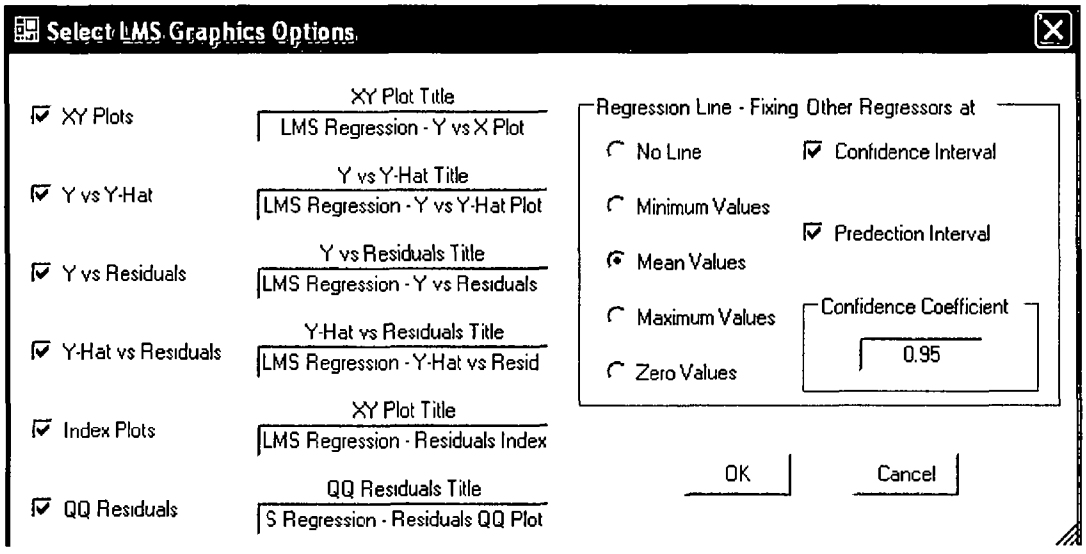
   - Specify "**Percentage Outliers.**" The default is "**<25%.**"

   - Specify "**Outlier Probability.**" The default is "**0.95.**"

- Specify "**Minimization Criterion**" as "**Percentile Squared Res.**" The default is "**0.75**."

- Click on "**OK**" to continue or "**Cancel**" to cancel the options.

- Click on "**Graphics**" for the graphics options and specify the preferred graphs.



- Click on "**OK**" to continue or "**Cancel**" to cancel the options.

- Click on "**OK**" to continue or "**Cancel**" to cancel the computations.

Output for LPS Regression.
Data Set used: Bradu (predictor variables p = 3, Minimization Criterion = 0.75 percentile).

	Least Percentile Squared (LPS) Regression Analysis Output
Date/Time of Computation	2/25/2008 11:08:20 AM
User Selected Options	
From File	D:\Narain\Scout_For_Windows\ScoutSource\WorkDatInExcel\BRADU
Full Precision	OFF
Subset Search Strategy	User Specified Criteria
Percentage Outliers	Maximum Outliers <= 0.25
Percentage Outliers	Outlier Probability <= 0 95
Search All Cutoff	Do all combinations if <= 100000
Minimization Criterion	The 0.75 Percentile of Squared Residuals
Residual QQ Plot	Not Selected
Residual Index Plot	Not Selected
Y vs X Plots	Not Selected
Title for Y-Hat vs Residuals Plot	LMS Regression - Y-Hat vs Residuals Plot
Y vs Residuals Plot	Not Selected
Y vs Y-Hat Plot	Not Selected

Number of Selected Regression Variables	3
Number of Observations	75
Dependent Variable	y

Correlation Matrix

	y	x1	x2	x3
y	1	0 946	0.962	0.743
x1	0 946	1	0.979	0 708
x2	0 962	0 979	1	0 757
x3	0.743	0.708	0.757	1

Eigenvalues for Correlation Matrix

Eval 1	Eval 2	Eval 3	Eval 4
0 0172	0.0556	0 368	3.559

OLS Estimates of Regression Parameters

Intercept	x1	x2	x3
-0.388	0 239	-0.335	0.383

Output for LPS Regression (continued).

Stdv of Estimated Regression Parameters							
Intercept	x1	x2	x3				
0.416	0.262	0.155	0.129				

OLS ANOVA Table					
Source of Variation	SS	DOF	MS	F-Value	P-Value
Regression	543.3	3	181.1	35.77	0.0000
Error	359.5	71	5.063		
Total	902.8	74			
			OLS Scale Estimate	2.25	
			R Square	0.602	

Least 0.75 Percentile of Squared Residual Regression		
Total Number of Elemental Subsets Searched	10000	
Number of Non-Singular Elemental Subsets	10000	

Best Elemental Subset of size 4 Found							
	y	x1	x2	x3			
Obs # 19	0.1	0.8	2.9	1.6			
Obs # 1	9.7	10.1	19.6	28.3			
Obs # 3	10.3	10.7	20.2	31			
Obs # 72	-0.2	0.6	2	1.5			

Best Subset satisfies minimization criterion.							

Percentile Squared Estimates of Regression Parameters (Using Best							
Intercept	x1	x2	x3				
-1.045	0.219	0.272	0.113				

Stdv of Estimated Regression Parameters (Using Best Subset)							
Intercept	x1	x2	x3				
0.572	0.36	0.213	0.177				

Minimizing 56th Ordered Squared Residual							
					Value of Minimum Criterion	0.606	
					Approximate Breakdown Value	0.24	
					Unweighted Sigma Estimate based upon LPS Residuals	3.088	

Output for LPS Regression (continued).

Initial Robust LPS Scale Estimate (Adjusted for dimensionality)						0 901			
LPS (0.75th) Regression Table Based Upon Best Subset									
Obs #	Y	Yhat	Residuals	Hat[i,i]	Res/Sigma	Student	Res/Scale	Weights	C Res~cale
1	9 7	9.7	1 356E-11	0.063	4.393E-12	4 538E-12	1 506E-11	1	1 502
2	10.1	9 882	0 218	0 0599	0.0707	0.073	0.242	1	1.775
3	10.3	10 3	1 530E-11	0.0857	4.954E-12	5.181E-12	1.698E-11	1	1 334
4	9 5	10 56	-1 058	0 0805	-0 343	-0 357	-1.174	1	1 138
5	10	10.47	-0.469	0 0729	-0.152	-0.158	-0.52	1	1.36
6	10	10 17	-0.173	0.0756	-0.0559	-0.0582	-0.192	1	1.527
7	10 8	10.23	0 568	0.068	0.184	0.191	0 631	1	2 006
8	10.3	9 713	0 587	0.0631	0.19	0.196	0.652	1	1.705
9	9 6	10 22	-0 617	0.08	-0.2	-0 208	-0.685	1	1.204
10	9.9	9.778	0 122	0 0869	0.0394	0.0412	0.135	1	1.35
11	-0.2	11.85	-12.05	0.0942	-3 903	-4.101	-13 38	0	-3.48
12	-0.4	12.03	-12.43	0.144	-4.024	-4.349	-13.79	0	-4.165
13	0.7	12 5	-11 8	0.109	-3 822	-4 049	-13.1	0	-2.719
14	0 1	14 46	-14 36	0 564	-4.65	-7.04	-15.94	0	-1.69
15	-0.4	0 725	-1 125	0 0579	-0.364	-0 375	-1.249	1	-0 294
16	0 6	0 266	0 334	0.0759	0 108	0 113	0 371	1	0.385
17	-0 2	-0.587	0 387	0.0393	0 125	0 128	0.43	1	0 287
18	0	0 12	-0.12	0 0231	-0.0387	-0.0392	-0.133	1	-0 175
19	0 1	0 1	-2.54E-12	0 0312	-8.24E-13	-8 37E-13	-2 82E-12	1	0 29
20	0.4	0.807	-0.407	0.0476	-0.132	-0.135	-0.452	1	0.151
21	0.9	0.337	0 563	0.0294	0 182	0.185	0.625	1	0.299
22	0 3	0.128	0.172	0.0457	0.0557	0.057	0.191	1	0.415
23	-0.8	0.109	-0.909	0.0293	-0.294	-0.299	-1.009	1	-0 19
24	0 7	-0.0783	0.778	0.0261	0 252	0.255	0.864	1	0 602
25	-0.3	-0.781	0.481	0 022	0 156	0.158	0.534	1	-0 136
26	-0.8	0 333	-1 133	0 0318	-0.367	-0 373	-1 257	1	-0 214
27	-0 7	0 685	-1 385	0 0417	-0.449	-0 458	-1.538	1	-0 612
28	0 3	-0 207	0 507	0.0235	0 164	0 166	0 563	1	-0.108
29	0 3	-0 447	0.747	0.0178	0 242	0 244	0 829	1	0.176
30	-0 3	-0 208	-0 0924	0.0466	-0.0299	-0.0307	-0.103	1	-0 564
31	0	0 127	-0 127	0 059	-0.0412	-0 0424	-0 141	1	-0 12
32	-0.4	-0 249	-0 151	0 0364	-0.049	-0 0499	-0 168	1	0 247
33	-0 6	0 296	-0 896	0.0264	-0 29	-0 294	-0 995	1	-0.0485
34	-0 7	-0 879	0 179	0 032	0 0578	0.0588	0.198	1	-0.301
35	0 3	0.626	-0.326	0.0342	-0.105	-0 107	-0 361	1	-0 178

(The complete regression table is not shown.)

Output for LPS Regression (continued).

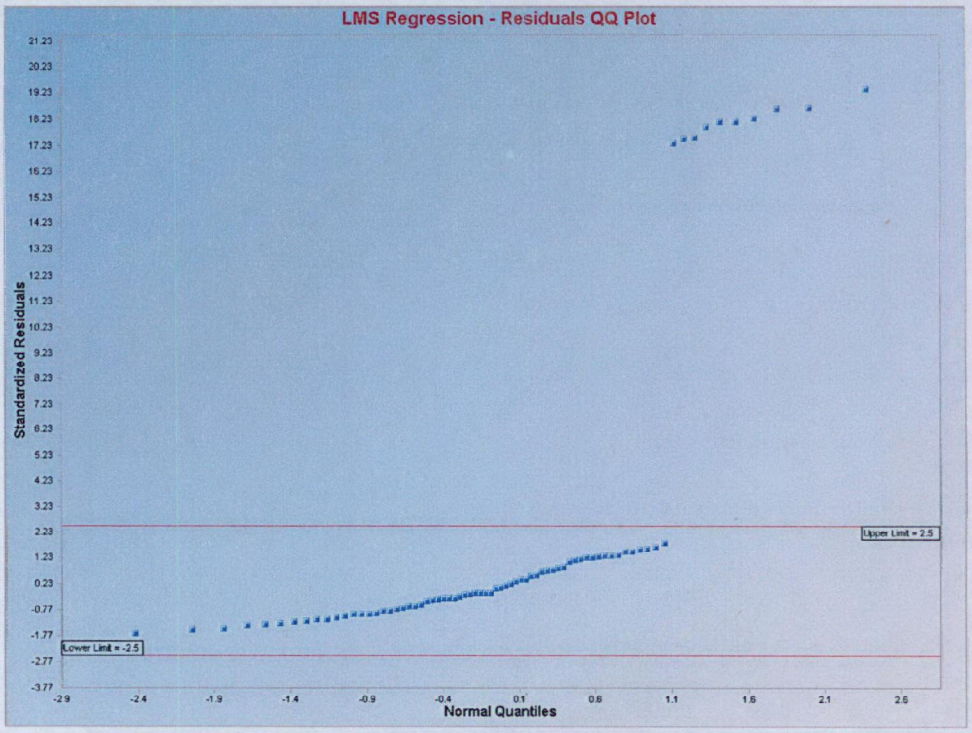
Reweighted LPS Estimates of Regression Parameters							
Intercept	x1	x2	x3				
-0 93	0 143	0 191	0 184				
Reweighted LPS Stdv of Estimated Regression Parameters							
Intercept	x1	x2	x3				
0.13	0.0795	0.0718	0.0505				
Reweighted LPS ANOVA Table							
Source of Variation			**SS**	**DOF**	**MS**	**F-Value**	**P-Value**
Regression			865 3	3	288 4	635	0.0000
Error			30 43	67	0.454		
Total			895.7	70			
					R Square	0 966	
					Final Reweighted LPS Scale Estimate	0 674	

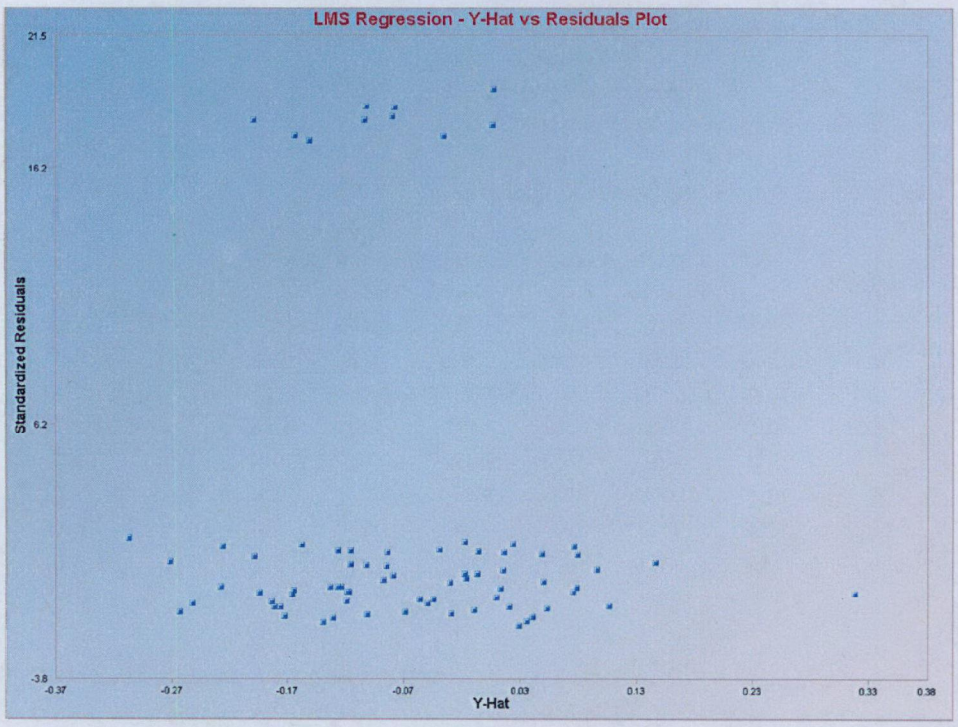
Reweighted LPS Regression Table							
Obs #	Y	Yhat	Residuals	Hat[i,i]	Student	Res/Scale	
1	9 7	9 475	0.225	0.063	0 346	0.335	
2	10.1	9 671	0 429	0.0599	0.657	0 637	
3	10 3	10.17	0 127	0.0857	0.197	0.189	
4	9.5	10.44	-0.936	0.0805	-1.448	-1.388	
5	10	10 31	-0.306	0.0729	-0.471	-0.454	
6	10	9 893	0 107	0.0756	0.165	0 158	

(The complete regression table is not shown.)

73	0.4	-0.157	0 557	0 0426	0.844	0 826	
74	-0.9	-0 215	-0.685	0.05	-1 043	-1.016	
75	0 2	-0 331	0 531	0.0621	0 813	0.788	
Final Weighted Correlation Matrix							
	y	x1	x2	x3			
y	1	0 939	0 946	0.943			
x1	0 939	1	0 985	0 977			
x2	0 946	0 985	1	0 98			
x3	0 943	0 977	0 98	1			
Eigenvalues for Final Weighted Correlation Matrix							
Eval 1	Eval 2	Eval 3	Eval 4				
0 0142	0 0244	0.0762	3 885				

Output for LPS Regression (continued).





Output for LPS Regression.
Data Set used: Bradu (predictor variables p = 3, Minimization Criterion = 0.9 percentile).

Least 0.9 Percentile of Squared Residual Regression

Total Number of Elemental Subsets Searched	10000
Number of Non-Singular Elemental Subsets	10000

Best Elemental Subset of size 4 Found

	y	x1	x2	x3
Obs # 31	0	3.1	1 4	1
Obs # 32	-0 4	0 5	2 4	0 3
Obs # 3	10.3	10.7	20 2	31
Obs # 45	-0 5	1 9	0.1	0 6

Best Subset satisfies minimization criterion.

ercentile Squared Estimates of Regression Parameters (Using Best

Intercept	x1	x2	x3
-0 951	0 167	0.171	0.194

Stdv of Estimated Regression Parameters (Using Best Subset)

Intercept	x1	x2	x3
0 554	0.349	0.206	0.171

Minimizing 67th Ordered Squared Residual	
Value of Minimum Criterion	1 664
Approximate Breakdown Value	0 0933
Unweighted Sigma Estimate based upon LMS Residuals	2 991
Initial Robust LPS Scale Estimate (Adjusted for dimensionality)	0 909

LPS (0.9th) Regression Table Based Upon Best Subset

Obs #	Y	Yhat	Residuals	Hat(ι,ι)	Res/Sigma	Student	Res/Scale	Weights	C Res~cale
1	9 7	9 573	0 127	0.063	0 0424	0 0438	0.139	1	1 502
2	10.1	9 743	0 357	0 0599	0.119	0 123	0.393	1	1.775
3	10.3	10 3	1 723E-13	0 0857	5.760E-14	6 024E-14	1 895E-13	1	1 334
4	9 5	10 52	-1.024	0 0805	-0.342	-0 357	-1 126	1	1.138
5	10	10.41	-0.406	0 0729	-0.136	-0.141	-0 447	1	1 36
6	10	10	-0.00147	0 0756	-4.906E-4	-5.102E-4	-0 00161	1	1 527
7	10.8	10 02	0 783	0.068	0 262	0.271	0.861	1	2 006
8	10.3	9 637	0 663	0.0631	0 222	0 229	0.729	1	1 705

Output for LPS Regression (continued).

8	10.3	9 637	0 663	0 0631	0 222	0 229	0 729	1	1 705
9	9 6	10 22	-0.618	0 08	-0 207	-0 215	-0 68	1	1 204
10	9.9	9 845	0.0552	0 0869	0.0185	0 0193	0 0607	1	1.35
11	-0 2	11 77	-11 97	0.0942	-4 003	-4 206	-13 17	0	-3 48
12	-0.4	12 16	-12 56	0 144	-4 199	-4 538	-13.82	0	-4.165
13	0 7	12.09	-11 39	0 109	-3 807	-4.034	-12.53	0	-2 719
14	0.1	13 29	-13 19	0 564	-4.408	-6 673	-14 5	0	-1 69
15	-0.4	0 52	-0.92	0 0579	-0.307	-0.317	-1.011	1	-0.294
16	0 6	5 8698E-4	0 599	0 0759	0 2	0 208	0.659	1	0 385
17	-0.2	-0 639	0.439	0 0393	0.147	0.15	0.483	1	0 287
18	0	0 0945	-0 0945	0.0231	-0 0316	-0 0319	-0 104	1	-0.175
19	0 1	-0 0122	0 112	0 0312	0 0375	0 0381	0 123	1	0.29
20	0 4	0.574	-0.174	0 0476	-0 0582	-0.0596	-0 191	1	0 151
21	0 9	0 228	0 672	0 0294	0 225	0 228	0 74	1	0 299
22	0.3	0 0303	0 27	0.0457	0 0902	0 0923	0 297	1	0.415
23	-0 8	-0 0692	-0.731	0 0293	-0 244	-0 248	-0 804	1	-0 19
24	0 7	-0 244	0 944	0 0261	0 316	0 32	1 039	1	0 602
25	-0.3	-0 706	0.406	0.022	0 136	0.137	0 447	1	-0.136
26	-0 8	0 247	-1 047	0 0318	-0.35	-0 356	-1.152	1	-0 214
27	-0 7	0 59	-1.29	0 0417	-0 431	-0 44	-1 419	1	-0 612
28	0 3	-0 126	0.426	0.0235	0 142	0 144	0 468	1	-0 108
29	0 3	-0 442	0.742	0 0178	0 248	0.25	0 816	1	0 176
30	-0.3	0 0288	-0 329	0.0466	-0.11	-0.113	-0 362	1	-0.564
31	0	2 146E-14	-2 15E-14	0 059	-7.17E-15	-7 39E-15	-2 36E-14	1	-0 12
32	-0.4	-0 4	2 520E-14	0 0364	8 425E-15	8 583E-15	2 772E-14	1	0.247
33	-0 6	0 119	-0 719	0 0264	-0 241	-0 244	-0 791	1	-0 0485
34	-0 7	-0 748	0.0484	0.032	0 0162	0 0165	0 0533	1	-0 301
35	0 3	0 559	-0.259	0 0342	-0 0865	-0 088	-0 285	1	-0.178
36	-1	0 132	-1.132	0 0231	-0 378	-0 383	-1 245	1	-0.522
37	-0 6	0 0819	-0.682	0 0587	-0 228	-0 235	-0 75	1	-0 102
38	0.9	-0.457	1.357	0 021	0.454	0.458	1.493	1	0 557
39	-0 7	-0 367	-0.333	0 035	-0.111	-0.113	-0 366	1	-0 567
40	-0.5	-0 294	-0 206	0.03	-0 069	-0 0701	-0 227	1	-0 0102
41	-0.1	0.453	-0.553	0 0524	-0 185	-0 19	-0 608	1	-0 49
42	-0.7	-0.206	-0.494	0 0554	-0 165	-0 17	-0 543	1	-0.482
43	0.6	-0.197	0.797	0 0606	0 266	0 275	0.877	1	0 766
44	-0 7	0 0561	-0.756	0 0406	-0 253	-0 258	-0.832	1	-0 801
45	-0 5	-0 5	-4.16E-14	0 029	-1 39E-14	-1 41E-14	-4 58E-14	1	-0 339
46	-0.4	0 0173	-0.417	0 0377	-0.14	-0 142	-0 459	1	-0 634

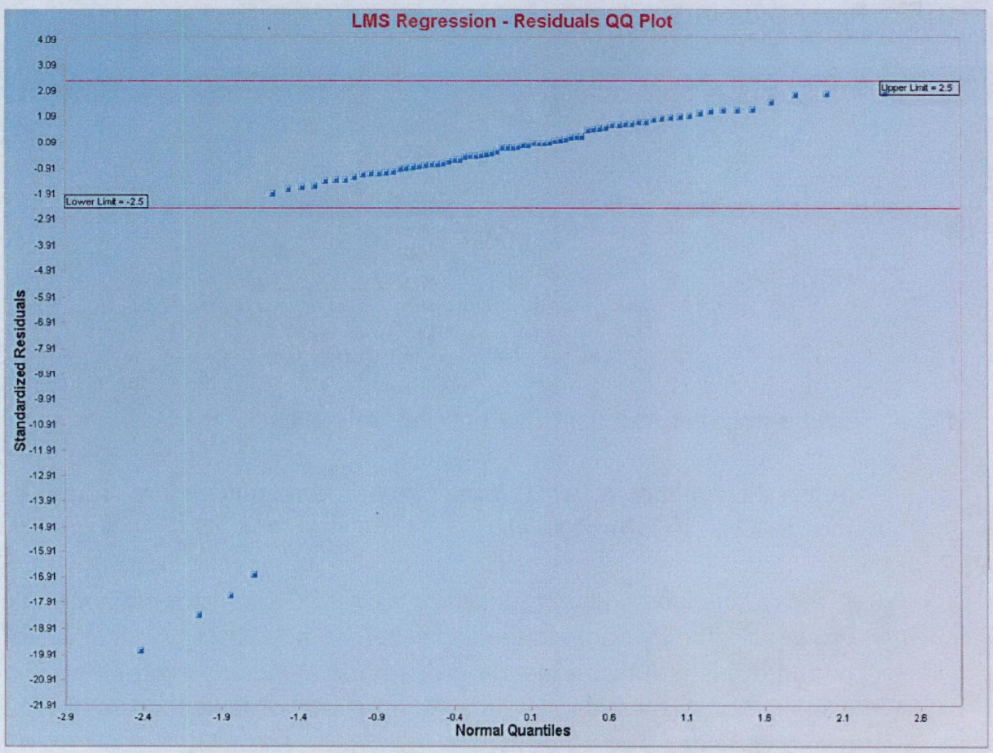
(The complete regression table is not shown.)

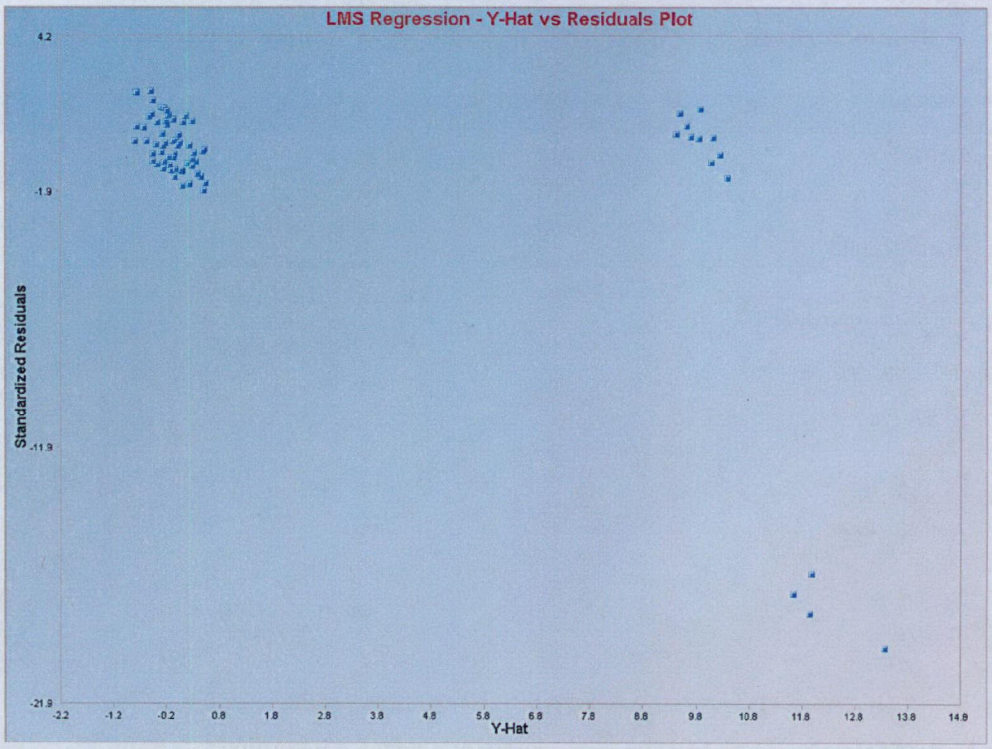
Output for LPS Regression (continued).

Reweighted LPS Estimates of Regression Parameters									
Intercept	x1	x2	x3						
-0.93	0.143	0.191	0.184						
Reweighted LMS Stdv of Estimated Regression Parameters									
Intercept	x1	x2	x3						
0.13	0.0795	0.0718	0.0505						
Reweighted LPS ANOVA Table									
Source of Variation			**SS**	**DOF**	**MS**	**F-Value**	**P-Value**		
Regression			865.3	3	288.4	635	0.0000		
Error			30.43	67	0.454				
Total			895.7	70					
					R Square	0 966			
					Final Reweighted LMS Scale Estimate	0 674			
Reweighted LPS Regression Table									
Obs #	Y	Yhat	Residuals	Hat[i,i]	Student	Res/Scale			
1	9 7	9.475	0 225	0.063	0 346	0.335			
2	10.1	9.671	0.429	0 0599	0.657	0 637			
3	10 3	10 17	0 127	0 0857	0 197	0 189			
4	9.5	10.44	-0.936	0.0805	-1 448	-1 388			
5	10	10 31	-0 306	0.0729	-0 471	-0.454			
6	10	9.893	0.107	0.0756	0 165	0 158			
7	10 8	9.927	0.873	0.068	1 341	1 295			
8	10.3	9 538	0.762	0.0631	1.168	1 13			
9	9.6	10.13	-0.525	0.08	-0 812	-0.779			
10	9.9	9.748	0.152	0.0869	0.236	0.225			
11	-0.2	11.68	-11.88	0.0942	-18.52	-17.63			
12	-0.4	12	-12.4	0.144	-19.89	-18.4			
13	0.7	12.02	-11.32	0.109	-17.79	-16 79			
14	0.1	13.4	-13.3	0.564	-29.88	-19.74			
15	-0.4	0.497	-0.897	0.0579	-1 372	-1.332			
16	0.6	-0 0111	0.611	0.0759	0 943	0 907			
17	-0.2	-0.588	0.388	0.0393	0.587	0 575			
18	0	0.0736	-0 0736	0.0231	-0.11	-0.109			
19	0 1	0.033	0 067	0 0312	0 101	0 0994			
20	0 4	0 568	-0.168	0.0476	-0.256	-0 25			

(The complete regression table is not shown.)

Output for LPS Regression (continued).

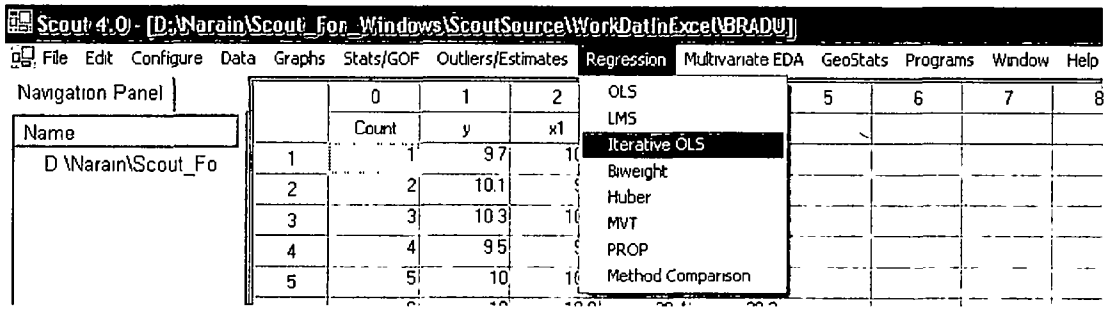




The 75th percentile minimization criterion finds 14 observations (1, 2, 3, 4, 5, 6, 7, 8, 9 and 10) as outliers and 90th percentile minimization criterion finds four observations (11, 12, 13 and 14) as outliers.

9.4 Iterative OLS Regression Method

1. Click **Regression ▶ Iterative OLS.**

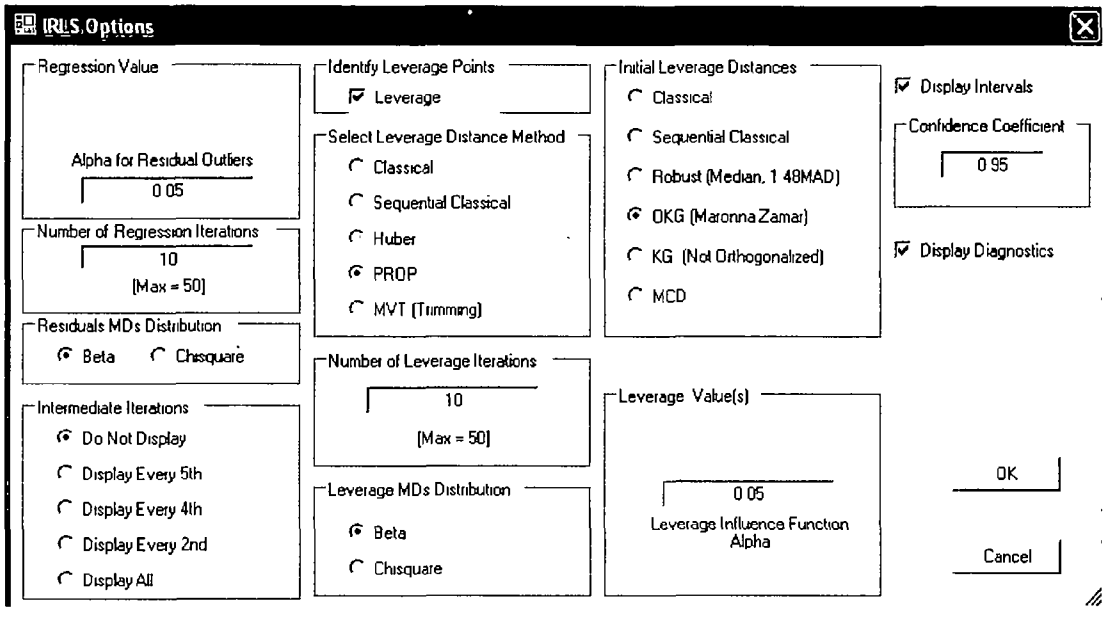


2. The "**Select Variables**" screen (Section 3.3) will appear.

   - Select the dependent variable and one or more independent variables from the "**Select Variables**" screen.

   - If the results have to be produced by using a Group variable, then select a group variable by clicking the arrow below the "**Group by Variable**" button. This will result in a drop-down list of available variables. The user should select and click on an appropriate variable representing a group variable.

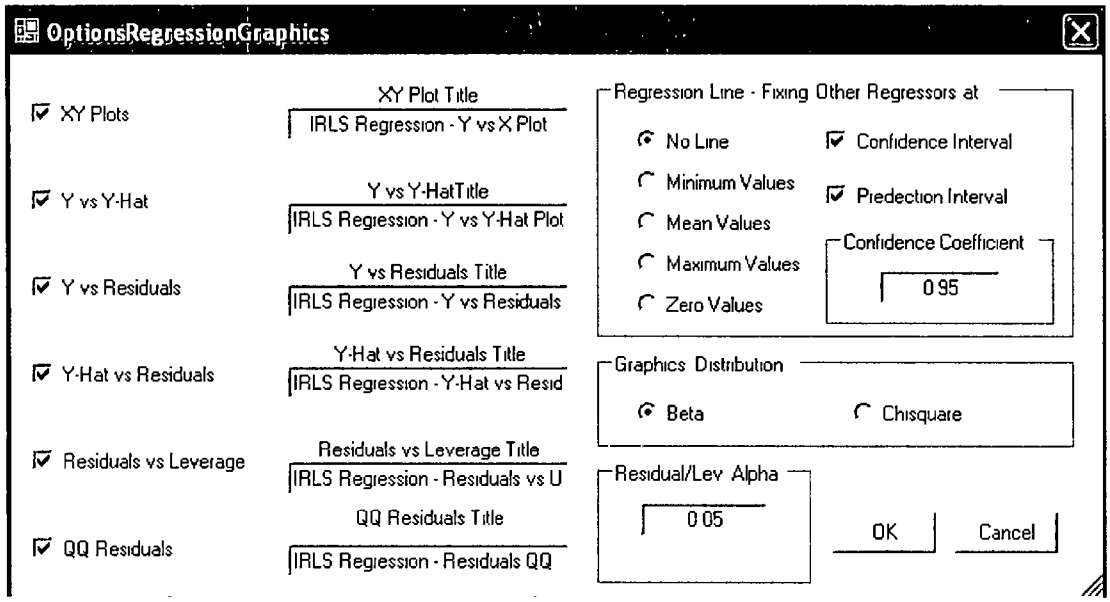
   - Click on the "**Options**" button to get the options window.



   - Specify the "**Regression Value.**" The default is "**0.05.**"

- Specify the "**Number of Regression Iterations.**" The default is"**10.**"
- Specify the "**Regression MDs Distribution.**" The default is "**Beta.**"
- Specify the "**Identify Leverage Points.**" The default is "**On.**"
- Specify the "**Select Leverage Distance Method.**" The default is "**PROP.**"
- Specify the "**Number of Leverage Iterations.**" The default is"**10.**"
- Specify the "**Leverage Initial Distances**" The default is "**OKG (Maronna Zamar).**"
- Specify the "**Leverage Value.**" The default is "**0.05.**"
- Click "**OK**" to continue or "**Cancel**" to cancel the options.

- Click on the **"Graphics"** button to get the options window.



- Specify the preferred plots and the input parameters.
- Click "**OK**" to continue or "**Cancel**" to cancel the options.

- Click "**OK**" to continue or "**Cancel**" to cancel the computations.

Output example: The data set "**BRADU.xls**" was used for iterative OLS regression. It has 3 predictor variables (p) and 75 observations. When the "**Leverage**" option is on, the leverage distances are calculated and outlying observations are obtained iteratively using initial estimates as median and OKG matrix and the leverage option as PROP (i.e., using PROP influence function). Then the weights are assigned to observations and those weights are used in finding the regression outliers iteratively. When the leverage option is off, all observations are assigned one (1) as weights and then the regression outliers are found iteratively. Finally, the estimated regression parameters are calculated.

Output for Iterative OLS (Leverage ON with PROP function and OKG initial start).
Data Set Used: Bradu (predictor variables p = 3).

	Regression Analysis Output
Date/Time of Computation	3/4/2008 9 50 32 AM
User Selected Options	
From File	D \Naram\Scout_For_Windows\ScoutSource\WorkDatInExcel\BRADU
Full Precision	OFF
Selected Regression Method	Iterative Reweighted Least Squares (IRLS)
Alpha for Residual Outliers	0 05 (Used to Identify Vertical Regression Outliers)
Number of Regression Iterations	10 (Maximum Number if doesn't Converge)
Leverage	Identify Leverage Points (Outliers in X Space)
Selected Leverage Method	PROP
Initial Leverage Distance Method	OKG (Maronna Zamar) Matrix
Squared MDs	Beta Distribution used for Leverage Distances based upon Selected Regression (Leverage) Variables
Leverage Distance Alpha	0 05 (Used to Identify Leverage Points)
Number of Leverage Iterations	10 (Maximum Number if doesn't Converge)
Y vs Y-hat Plot	Not Selected
Y vs Residual Plot	Not Selected
Y-hat vs Residual Plot	Not Selected
Y vs X Plots	Not Selected
Title for Residual QQ Plot	IRLS Regression - Residuals QQ Plot
Residual Band Alpha	0 05 (Used in Graphics Residual Bands)
Title Residual vs Distance Plot	IRLS Regression - Residuals vs Unsquared Leverage Distance Plot
Show Intermediate Results	Do Not Display Intermediate Results
	Intermediate Results Shown on Another Output Sheet

Leverage Points are Outliers in X-Space of Selected Regression Variables

Number of Selected Regression Variables	3
Number of Observations	75
Dependent Variable	y

Residual Values used with Graphics Display

Upper Residual Indvidual (0 05) MD	1 94
Lower Residual Indvidual (0 05) MD	-1 94

Correlation Matrix

	y	x1	x2	x3
y	1	0 946	0 962	0 743
x1	0 946	1	0 979	0 708
x2	0 962	0 979	1	0 757
x3	0 743	0 708	0 757	1

Eigenvalues of Correlation Matrix

Eval 1	Eval 2	Eval 3	Eval 4
0 0172	0 0556	0 368	3 559

Ordinary Least Squares (OLS) Regression Results

Estimates of Regression Parameters

Intercept	x1	x2	x3
-0 388	0 239	-0 335	0 383

Stdv of Estimated Regression Parameters

Intercept	x1	x2	x3
0 416	0 262	0 155	0 129

Output for Iterative OLS (Leverage ON) (continued).

ANOVA Table

Source of Variation	SS	DOF	MS	F-Value	P-Value
Regression	543 3	3	181 1	35.77	0 0000
Error	359.5	71	5.063		
Total	902 8	74			

R Square Estimates	0 602
MAD Based Scale Estimates	1 067
Weighted Scale Estimates	2 25
IQR Estimates	1.468
Det of COV[Regression Coefficients] Matrix	5.5107E-8

Regression Parameters Vector Estimates

Intercept	x1	x2	x3
-0 0105	0 0624	0 0119	-0.107

Stdv of Regression Estimates Vector

Intercept	x1	x2	x3
0 197	0 0689	0 0684	0 0713

ANOVA Table

Source of Variation	SS	DOF	MS	F-Value	P-Value
Regression	0 898	3	0.299	0.94	0 4272
Error	18 14	57	0 318		
Total	19 04	60			

R Square Estimates	0 0472
MAD Based Scale Estimates	0 902
Weighted Scale Estimates	0 564
Individual MD[0]	7 346
IQR Estimates	1 236
Determinant of Leverage S Matrix	1.357

Output for Iterative OLS (Leverage ON) (continued).

Leverage Option Regression Table

Obs	Y Vector	Yhat	Residuals	Hat[i,i]	Res/Scale	Stude~ Res	Wts[i,i]	Res Dist	Lev Dist	OLS R~ist
1	9 7	-2.174	11 87	0 063	21 05	21 74	2 364E-13	21 05	29.44	1.502
2	10 1	-2 265	12 36	0 0599	21 92	22 61	1 074E-13	21.92	30 21	1 775
3	10 3	-2 418	12 72	0 0857	22 54	23 58	1 886E-14	22 54	31 89	1.334
4	9.5	-2.528	12 03	0 0805	21.32	22 23	6.931E-15	21 32	32.86	1 138
5	10	-2 443	12 44	0.0729	22 06	22 91	1 266E-14	22 06	32 28	1 36
6	10	-2.217	12.22	0 0756	21 66	22.52	7 228E-14	21 66	30.59	1 527
7	10 8	-2.219	13 02	0 068	23.08	23 9	6 576E-14	23.08	30.68	2.006
8	10 3	-2 24	12 54	0.0631	22.23	22.97	1.634E-13	22.23	29.8	1 705
9	9 6	-2 475	12 07	0 08	21 4	22 32	1 768E-14	21 4	31 95	1 204
10	9.9	-2 437	12 34	0 0869	21 87	22 89	5 017E-14	21.87	30 94	1 35
11	-0 2	-2 782	2.582	0 0942	4 577	4 809	1.424E-16	4 577	36 64	3 48
12	-0.4	-2.946	2.546	0.144	4 513	4 877	3.684E-17	4 513	37.96	4.165
13	0.7	-2.589	3 289	0 109	5.83	6.177	1 069E-16	5.83	36.92	2 719
14	0.1	-2 556	2.656	0.564	4 708	7 127	1.478E-18	4.708	41 09	1.69
15	-0.4	0 0115	-0 412	0 0579	-0.73	-0 752	1	0 73	2.002	0.294
16	0.6	0.177	0 423	0.0759	0.75	0 78	1	0 75	2.165	0 385
17	-0 2	-0 0128	-0 187	0 0393	-0 332	-0 339	1	0.332	1.938	0 287
18	0	-0.0619	0 0619	0.0231	0.11	0 111	1	0.11	0.786	0 175
19	0.1	-0 0971	0.197	0 0312	0.349	0 355	1	0 349	1 287	0.29
20	0.4	-0 0119	0.412	0.0476	0.73	0.748	1	0 73	2 067	0.151
21	0 9	-0.0253	0 925	0 0294	1 64	1.665	1	1 64	1.059	0.299
22	0.3	-0 151	0.451	0.0457	0.799	0 818	1	0 799	1.746	0.415
23	-0 8	0 0561	-0.856	0 0293	-1.518	-1 54	1	1.518	1.163	0 19
24	0.7	0.0446	0 655	0.0261	1 162	1.177	1	1.162	1.317	0 602
25	-0 3	0 00912	-0 309	0 022	-0.548	-0 554	1	0 548	1 986	0.136
26	-0 8	-0 182	-0 618	0.0318	-1 095	-1 113	1	1 095	1 706	0 214
27	-0 7	-0 085	-0 615	0 0417	-1 09	-1 114	1	1 09	1 994	0 612
28	0 3	-0 103	0 403	0 0235	0.714	0.722	1	0 714	1.036	0.108
29	0.3	-0 0105	0.31	0.0178	0.55	0 555	1	0 55	1 136	0 176
30	-0 3	-0 291	-0 00901	0.0466	-0 016	-0 0164	1	0 016	2 111	0 564
31	0	0.0926	-0 0926	0 059	-0.164	-0.169	1	0 164	1 715	0.12
32	-0 4	0 0173	-0 417	0 0364	-0.74	-0 754	1	0 74	1.763	0.247
33	-0.6	-0.0404	-0 56	0.0264	-0 992	-1.005	1	0.992	1 277	0 0485
34	-0.7	-0 0604	-0 64	0.032	-1 134	-1 152	1	1.134	2.042	0 301
35	0 3	-0 109	0.409	0 0342	0.726	0.738	1	0 726	1 885	0.178
36	-1	-0 204	-0 796	0 0231	-1.41	-1 427	1	1 41	1 144	0.522
37	-0 6	-0 247	-0.353	0 0587	-0 626	-0.646	1	0 626	2.014	0.102

(The complete regression table is not shown.)

Output for Iterative OLS (Leverage ON) (continued).

THE BREAK BETWEEN LEVERAGE AND REGRESSION IS HERE!

Results From the Regression Operation

Regression Parameters Vector Estimates

Intercept	x1	x2	x3
-0 18	0.0814	0.0399	-0 0517

Stdv of Regression Estimates Vector

Intercept	x1	x2	x3
0 104	0.0667	0 0405	0 0354

ANOVA Table

Source of Variation	SS	DOF	MS	F-Value	P-Value
Regression	0.847	3	0.282	0.909	0.4421
Error	18 94	61	0 31		
Total	19 79	64			

R Square Estimates	0 0428
MAD Based Scale Estimates	0 845
Weighted Scale Estimates	0 557
Individual MD(0)	7 346
IQR Estimates	1.132
Det. of COV[Regression Coefficients] Matrix	2.531E-12

Regression Table

Obs	Y Vector	Yhat	Residuals	Hat[i,i]	Res/Scale	Stude~ Res	Wts[i,i]	Res Dist
1	9 7	-0.0386	9 739	0 063	17 48	18 06	0	17 48
2	10 1	-0 0825	10.18	0 0599	18 27	18.85	0	18 27
3	10 3	-0 105	10.41	0 0857	18 67	19 53	0	18 67
4	9 5	-0.155	9.655	0.0805	17 33	18.07	0	17 33
5	10	-0.107	10 11	0.0729	18.14	18 84	0	18 14
6	10	0.00379	9.996	0.0756	17.94	18.66	0	17 94
7	10.8	0 00449	10.8	0.068	19 37	20.07	0	19.37
8	10.3	-0 0807	10.38	0 0631	18.63	19.25	0	18.63
9	9.6	-0.167	9 767	0.08	17.53	18 27	0	17.53
10	9 9	-0.203	10 1	0 0869	18.13	18 98	0	18 13
11	-0 2	-0 136	-0.0641	0 0942	-0 115	-0.121	1	0 115

(The complete regression table is not shown.)

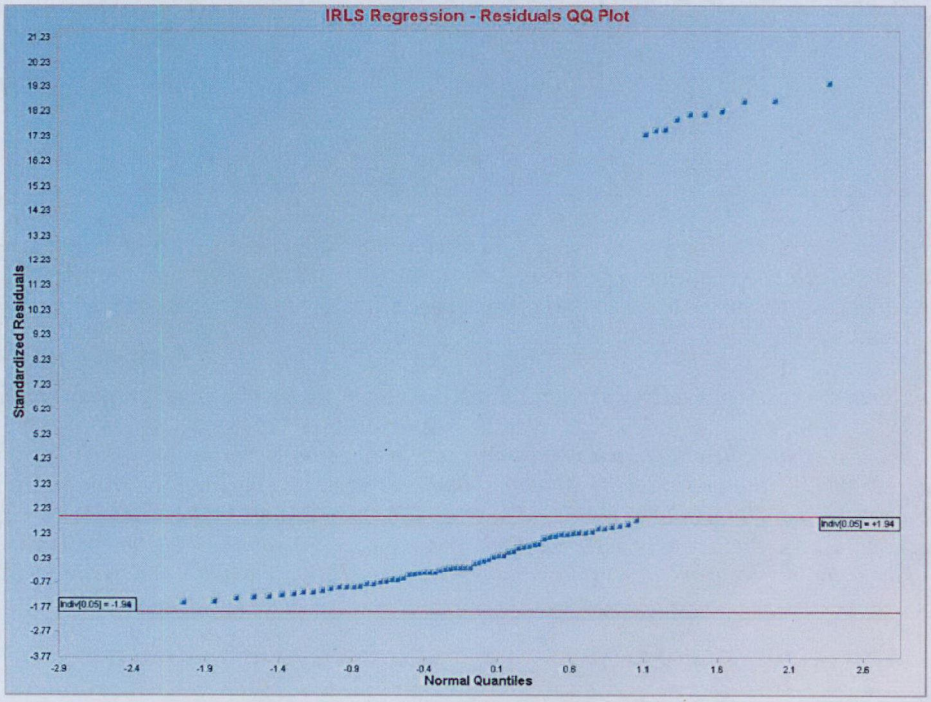
Final Weighted Correlation Matrix

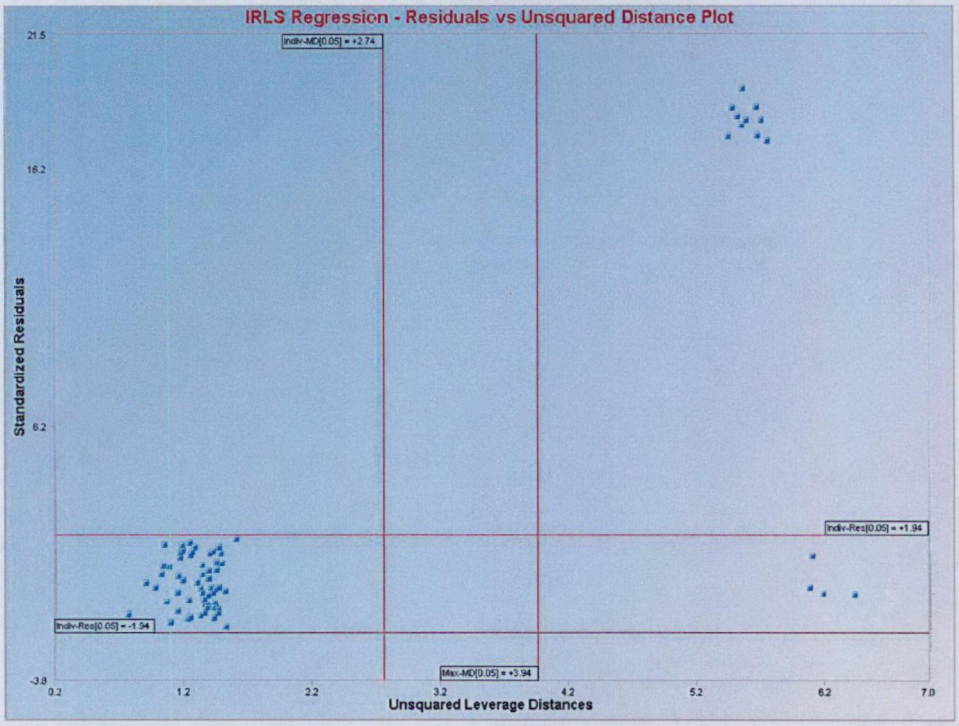
	y	x1	x2	x3
y	1	0.89	0.917	0.0893
x1	0.89	1	0.961	0.063
x2	0.917	0.961	1	0.0261
x3	0.0893	0.063	0.0261	1

Eigenvalues of Final Weighted Correlation Matrix

Eval 1	Eval 2	Eval 3	Eval 4
0.035	0.117	0.997	2.851

Output for Iterative OLS (Leverage ON) (continued).





Interpretation of Graphs: Observations which are outside of the horizontal lines in both of the graphs are considered to be regression outliers. The observations to the right of the vertical lines are considered to be leverage outliers. Observations between the horizontal and to the right of the vertical lines represent good leverage points.

Question: *What are really bad leverage points for this data set in the context of a regression model?*

Answer: *There are contradictory opinions in this respect. So far as outliers are considered, several methods (e.g., MCD, PROP) can identify all of the 14 outliers present in this data set. However, observations 1 through 10 should be considered to be good leverage points as they enhance the regression model and increase the coefficient of determination. Without those 10 points, fitting a regression model to the rest of the 65 points is meaningless. Observations 11 through 14 are outliers and bad leverage points.*

Output for Iterative OLS (Leverage OFF).

Regression Analysis Output	
Date/Time of Computation	3/4/2008 9 54 08 AM
User Selected Options	
From File	D:\Narain\Scout_For_Windows\ScoutSource\WorkDatInExcel\BRADU
Full Precision	OFF
Selected Regression Method	Iterative Reweighted Least Squares (IRLS)
Alpha for Residual Outliers	0 05 (Used to Identify Vertical Regression Outliers)
Number of Regression Iterations	10 (Maximum Number if doesn't Converge)
Leverage	Off
Y vs Y-hat Plot	Not Selected
Y vs Residual Plot	Not Selected
Y-hat vs Residual Plot	Not Selected
Y vs X Plots	Not Selected
Title for Residual QQ Plot	IRLS Regression - Residuals QQ Plot
Residual Band Alpha	0 05 (Used in Graphics Residual Bands)
Title Residual vs Distance Plot	IRLS Regression - Residuals vs Unsquared Leverage Distance Plot
Show Intermediate Results	Do Not Display Intermediate Results
	Intermediate Results Shown on Another Output Sheet

Number of Selected Regression Variables	3
Number of Observations	75
Dependent Variable	y

Residual Values used with Graphics Display

Upper Residual Indvidual (0 05) MD	1 94
Lower Residual Indvidual (0 05) MD	-1.94

Correlation Matrix

	y	x1	x2	x3
y	1	0 946	0 962	0 743
x1	0 946	1	0 979	0 708
x2	0 962	0 979	1	0 757
x3	0 743	0 708	0 757	1

Eigenvalues of Correlation Matrix

Eval 1	Eval 2	Eval 3	Eval 4
0.0172	0 0556	0 368	3 559

Output for Iterative OLS (Leverage OFF) (continued).

Ordinary Least Squares (OLS) Regression Results

Estimates of Regression Parameters

Intercept	x1	x2	x3
-0 388	0.239	-0 335	0 383

Stdv of Estimated Regression Parameters

Intercept	x1	x2	x3
0 416	0 262	0 155	0 129

ANOVA Table

Source of Variation	SS	DOF	MS	F-Value	P-Value
Regression	543 3	3	181 1	35 77	0 0000
Error	359 5	71	5.063		
Total	902 8	74			

R Square Estimate	0 602
MAD Based Scale Estimate	1 067
Weighted Scale Estimate	2 25
IQR Estimate of Residuals	1 468
Det. of COV[Regression Coefficients] Matrix	5 5107E-8

Final Reweighted Regression Results

Estimates of Regression Parameters

Intercept	x1	x2	x3
-0 388	0 239	-0 335	0.383

Stdv of Estimated Regression Parameters

Intercept	x1	x2	x3
0 416	0 262	0 155	0 129

ANOVA Table

Source of Variation	SS	DOF	MS	F-Value	P-Value
Regression	543.3	3	181 1	35 77	0.0000
Error	359 5	71	5 063		
Total	902.8	74			

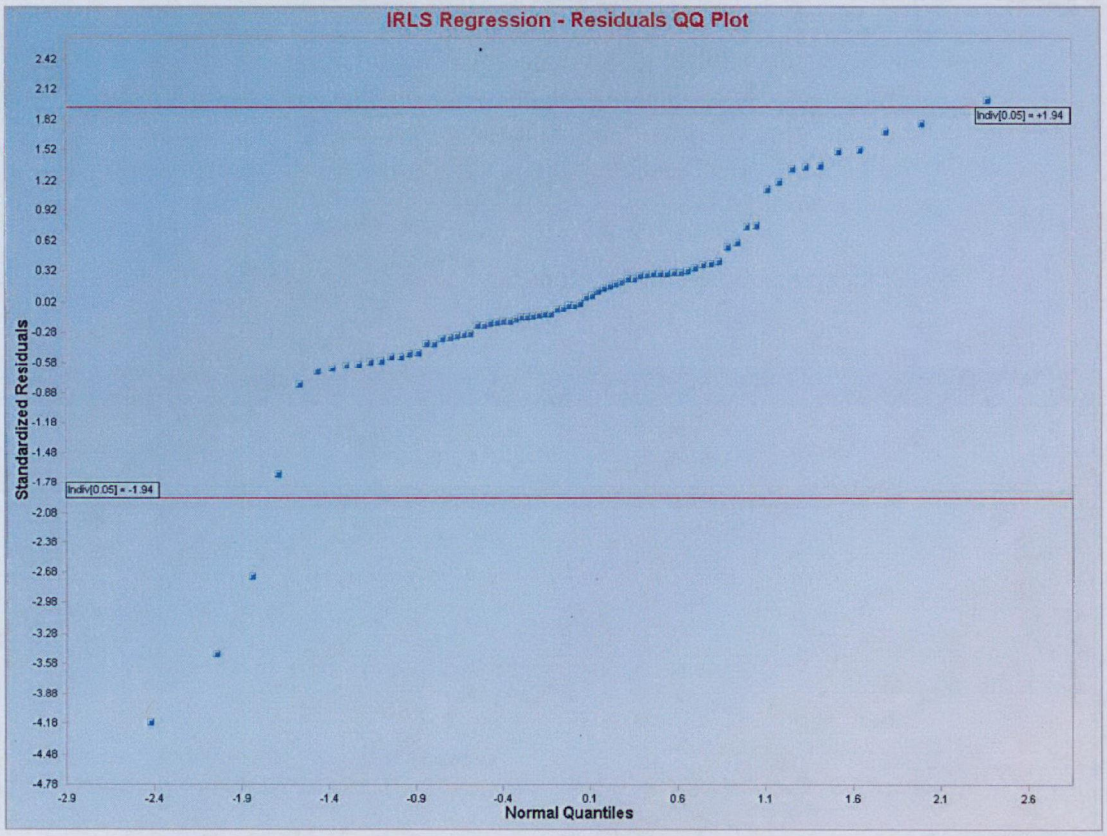
R Square Estimate	0 602
MAD Based Scale Estimate	1 067
Weighted Scale Estimate	2 25
IQR Estimate of Residuals	1 468
Det. of COV[Regression Coefficients] Matrix	5 5107E-8

Regression Table

Obs	Y Vector	Yhat	Residuals	Hat[i,i]	Res/Scale	Stude~ Res	Wts[i,i]	Res Dist
1	9 7	6 32	3 38	0 063	1 502	1 552	1	1 502
2	10 1	6 105	3 995	0 0599	1 775	1 831	1	1 775
3	10 3	7 297	3 003	0 0857	1 334	1 396	1	1 334
4	9 5	6 939	2 561	0 0805	1 138	1 187	1	1 138
5	10	6 939	3 061	0 0729	1 36	1 413	1	1 36

(The complete regression table is not shown.)

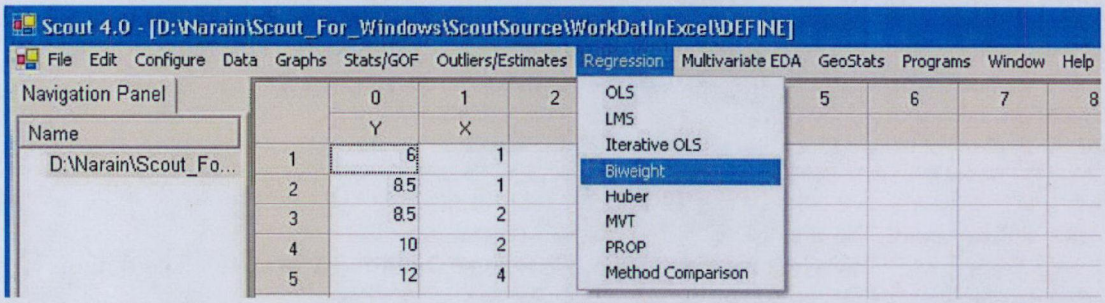
Output for Iterative OLS (Leverage OFF) (continued).



Interpretation of Graphs: Observations which are outside of the horizontal lines in the graph are considered to be regression outliers. The Leverage Distances vs. Standardized residuals plot is not produced. The sequential classical method failed to identify all of the regression outliers.

9.5 Biweight Regression Method

1. Click **Regression ▶ Biweight.**

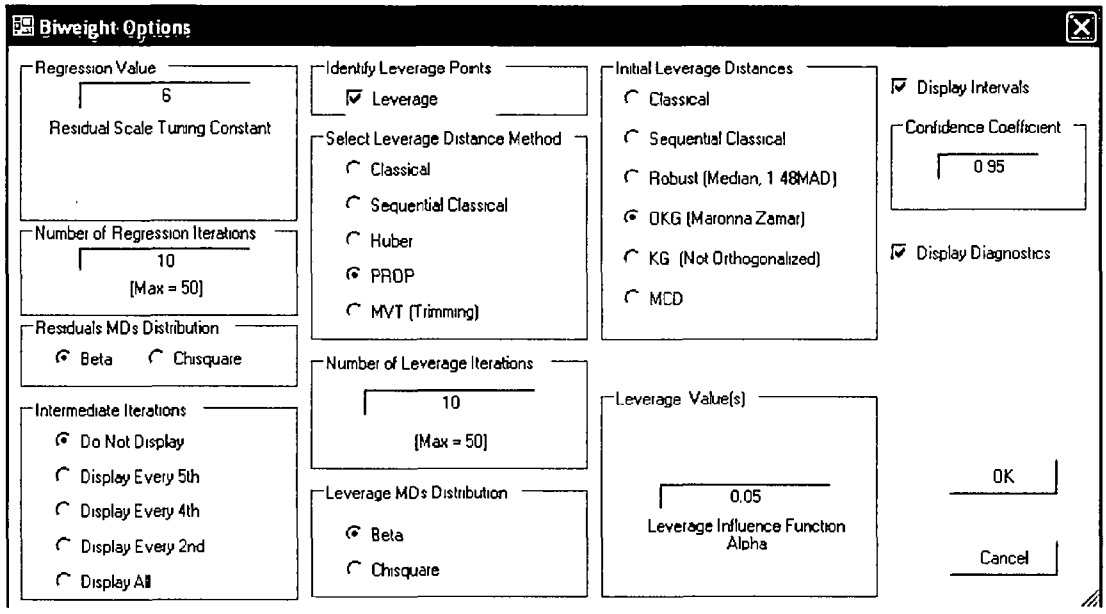


2. The "**Select Variables**" screen (Section 3.3) will appear.

- Select the dependent variable and one or more independent variables from the "**Select Variables**" screen.

- If the results have to be produced by using a Group variable, then select a group variable by clicking the arrow below the "**Group by Variable**" button. This will result in a drop-down list of available variables. The user should select and click on an appropriate variable representing a group variable.

- Click on the "**Options**" button to get the options window.



- Specify the "**Regression Value.**" The default is "**4.**"

- Specify the "**Number of Regression Iterations.**" The default is "**10.**"

- Specify the "**Regression MDs Distribution.**" The default is "**Beta.**"

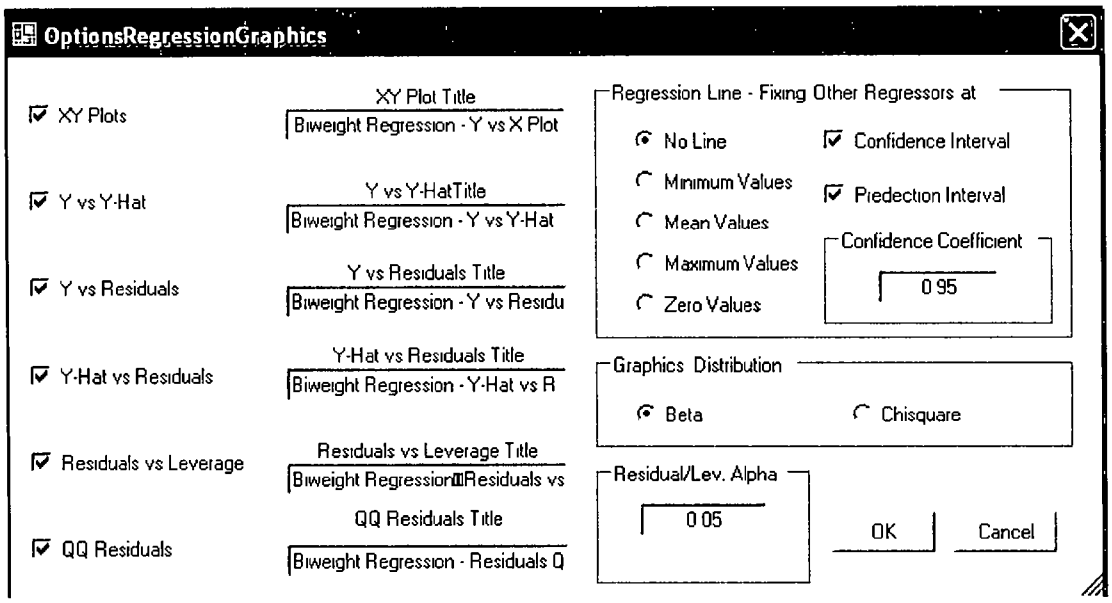
- Specify "**Identify Leverage Points.**" The default is "**On.**"

- Specify the "**Select Leverage Distance Method.**" The default is "**PROP.**"

- Specify the "**Number of Leverage Iterations.**" The default is"**10.**"

- Specify the "**Leverage Initial Distances**" The default is "**OKG (Maronna Zamar).**"
- Specify the "**Leverage Value.**" The default is "**0.05.**"
- Click "**OK**" to continue or "**Cancel**" to cancel the options

- Click on the "**Graphics**" button to get the options window.



- Specify the preferred plots and the input parameters.
- Click "**OK**" to continue or "**Cancel**" to cancel the options

- Click "**OK**" to continue or "**Cancel**" to cancel the computations.

Output example: The data set "**DEFINE.xls**" was used for Biweight regression. It has 1 predictor variables (p) and 26 observations. When the "**Leverage**" option is on, the leverage distances are calculated and outlying observations are obtained iteratively using initial estimates as median and OKG matrix and the leverage option as PROP (i.e., using PROP influence function). Then the weights are assigned to observations and those weights are used in the finding the regression outliers iteratively. When the leverage option is off, all observations are assigned one (1) as weights and then the regression outliers are found using the Biweight tuning constant iteratively. Finally, the estimated regression parameters are calculated.

Output for Biweight (Leverage ON).
Data Set Used: Define (predictor variables p = 1).

	Regression Analysis Output
Date/Time of Computation	3/4/2008 10 03 07 AM
User Selected Options	
From File	D \Narain\Scout_For_Windows\ScoutSource\WorkDatInExcel\DEFINE
Full Precision	OFF
Selected Regression Method	Biweight
Residual Biweight Tuning Constant	4 (Used to Identify Vertical Regression Outliers)
Number of Regression Iterations	10 (Maximum Number if doesn't Converge)
Leverage	Identify Leverage Points (Outliers in X-Space)
Selected Leverage Method	PROP
Initial Leverage Distance Method	OKG (Maronna Zamar) Matrix
Squared MDs	Beta Distribution used for Leverage Distances based upon Selected Regression (Leverage) Variables
Leverage Distance Alpha	0 05 (Used to Identify Leverage Points)
Number of Leverage Iterations	10 (Maximum Number if doesn't Converge)
Y vs Y-hat Plot	Not Selected
Y vs Residual Plot	Not Selected
Y-hat vs Residual Plot	Not Selected
Title For Y vs X Plots	Biweight Regression - Y vs X Plot
Title for Residual QQ Plot	Biweight Regression - Residuals QQ Plot
Residual Band Alpha	0 05 (Used in Graphics Residual Bands)
Title Residual vs Distance Plot	Biweight Regression - Residuals vs Unsquared Leverage Distance Plot
Show Intermediate Results	Do Not Display Intermediate Results
	Intermediate Results Shown on Another Output Sheet

Leverage Points are Outliers in X-Space of Selected Regression Variables.

Number of Selected Regression Variables	1
Number of Observations	26
Dependent Variable	Y

Residual Values used with Graphics Display	
Upper Residual Indvidual (0 05) MD	1 903
Lower Residual Indvidual (0 05) MD	-1 903

Correlation Matrix

	Y	X
Y	1	0 218
X	0 218	1

Output for Biweight (Leverage ON) (continued).

Eigenvalues of Correlation Matrix	
Eval 1	Eval 2
0.782	1.218

Ordinary Least Squares (OLS) Regression Results

Estimates of Regression Parameters	
Intercept	X
22.06	0 256

Stdv of Estimated Regression Parameters	
Intercept	X
4 107	0 233

ANOVA Table

Source of Variation	SS	DOF	MS	F-Value	P-Value
Regression	224 9	1	224 9	1.202	0.2838
Error	4490	24	187.1		
Total	4715	25			

R Square Estimate	0.0477
MAD Based Scale Estimate	8.862
Weighted Scale Estimate	13 68
IQR Estimate of Residuals	25.79
Det. of COV[Regression Coefficients] Matrix	0.391

Initial Weighted Regression Iteration with Identified Leverage Points

Estimates of Regression Parameters	
Intercept	X
22.06	0.256

Stdv of Estimated Regression Parameters	
Intercept	X
4.107	0.233

Output for Biweight (Leverage ON) (continued).

ANOVA Table					
Source of Variation	SS	DOF	MS	F-Value	P-Value
Regression	224 9	1	224.9	1.202	0 2838
Error	4490	24	187 1		
Total	4715	25			

R Square Estimate	0 0477
MAD Based Scale Estimate	8 862
Weighted Scale Estimate	13.68
Unsquared Leverage Distance Indiv-MD(0.05)	1.903
IQR Estimate of Residuals	25 79
Determinant of Leverage S Matrix	137.6

Regression Table with Leverage Option

Obs	Y Vector	Yhat	Residuals	Hat[i,i]	Res/Scale	Stude~ Res	Wts[i,i]	Res Dist	Lev Dist.	OLS R~ist
1	6	22.32	-16 32	0 0827	-1 193	-1.246	1	1.193	1 052	1.193
2	8.5	22.32	-13.82	0.0827	-1 01	-1.055	1	1.01	1 052	1.01
3	8.5	22.58	-14 08	0.0758	-1 029	-1.07	1	1.029	0 966	1.029
4	10	22 58	-12 58	0 0758	-0 919	-0 956	1	0.919	0 966	0 919
5	12	23.09	-11 09	0.0638	-0 811	-0.838	1	0.811	0.796	0 811
6	40	23.09	16 91	0 0638	1 237	1 278	1	1 237	0.796	1 237
7	42.5	23 09	19 41	0 0638	1.419	1 467	1	1 419	0.796	1 419
8	45	23 34	21.66	0 0587	1.583	1.632	1	1.583	0.711	1.583
9	50	23.34	26 66	0.0587	1 949	2.009	1	1.949	0 711	1.949
10	13	23 09	-10.09	0 0638	-0 737	-0 762	1	0 737	0 796	0 737
11	14	23.09	-9.086	0 0638	-0.664	-0 687	1	0.664	0.796	0.664
12	17	23 34	-6 342	0 0587	-0 464	-0.478	1	0.464	0 711	0.464
13	17 4	23 34	-5 942	0 0587	-0 434	-0 448	1	0 434	0.711	0 434
14	22	24.62	-2 62	0 0417	-0.192	-0 196	1	0 192	0.284	0.192
15	24	24 62	-0 62	0.0417	-0 0454	-0.0463	1	0.0454	0.284	0 0454
16	25	24.62	0 38	0.0417	0 0277	0 0283	1	0 0277	0.284	0 0277
17	42.5	27.18	15.32	0.0514	1.12	1.15	1	1.12	0.568	1.12
18	43	27.18	15 82	0 0514	1.157	1 188	1	1.157	0 568	1 157
19	44 1	27 69	16.41	0 0603	1 2	1.238	1	1 2	0.739	1 2
20	45 3	27 82	17.48	0 0629	1 278	1 32	1	1 278	0 781	1.278
21	20	29 73	-9 734	0.119	-0.712	-0 758	1	0.712	1 421	0.712
22	22	29.73	-7.734	0 119	-0 565	-0.602	1	0.565	1.421	0.565
23	21	29 99	-8 99	0 129	-0 657	-0 704	1	0 657	1.506	0.657
24	23	30 25	-7 245	0 14	-0 53	-0 571	1	0 53	1.591	0 53

(The complete regression table is not shown.)

Output for Biweight (Leverage ON) (continued).

Final Reweighted Regression Results

Estimates of Regression Parameters

Intercept	X
11 64	0 358

Stdv of Estimated Regression Parameters

Intercept	X
1 372	0 0696

ANOVA Table

Source of Variation	SS	DOF	MS	F-Value	P-Value
Regression	346	1	346	26 45	0 0002
Error	173 3	13 25	13 08		
Total	519.3	14 25			

R Square Estimate	0 666
MAD Based Scale Estimate	7 695
Weighted Scale Estimate	3 617
IQR Estimate of Residuals	25 53
Det of COV[Regression Coefficients] Matrix	0 00416

Regression Table

Obs	Y Vector	Yhat	Residuals	Hat[i,i]	Res/Scale	Stude~ Res	Wts[i,i]	Res Dist.
1	6	11 99	-5 993	0 0827	-1 657	-1 73	0 678	1 657
2	8 5	11 99	-3 493	0 0827	-0 966	-1 008	0 883	0 966
3	8 5	12 35	-3 852	0 0758	-1.065	-1.108	0 859	1 065
4	10	12 35	-2 352	0 0758	-0 65	-0 676	0.946	0 65
5	12	13 07	-1 068	0 0638	-0 295	-0 305	0 989	0 295
6	40	13 07	26 93	0.0638	7.446	7 696	0	7.446
7	42 5	13 07	29.43	0 0638	8.138	8 41	0	8 138
8	45	13 43	31 57	0 0587	8 73	8.998	0	8 73
9	50	13 43	36.57	0 0587	10.11	10 42	0	10 11
10	13	13 07	-0 0681	0 0638	-0 0188	-0 0195	1	0 0188
11	14	13 07	0 932	0 0638	0 258	0 266	0 992	0 258

(The complete regression table is not shown.)

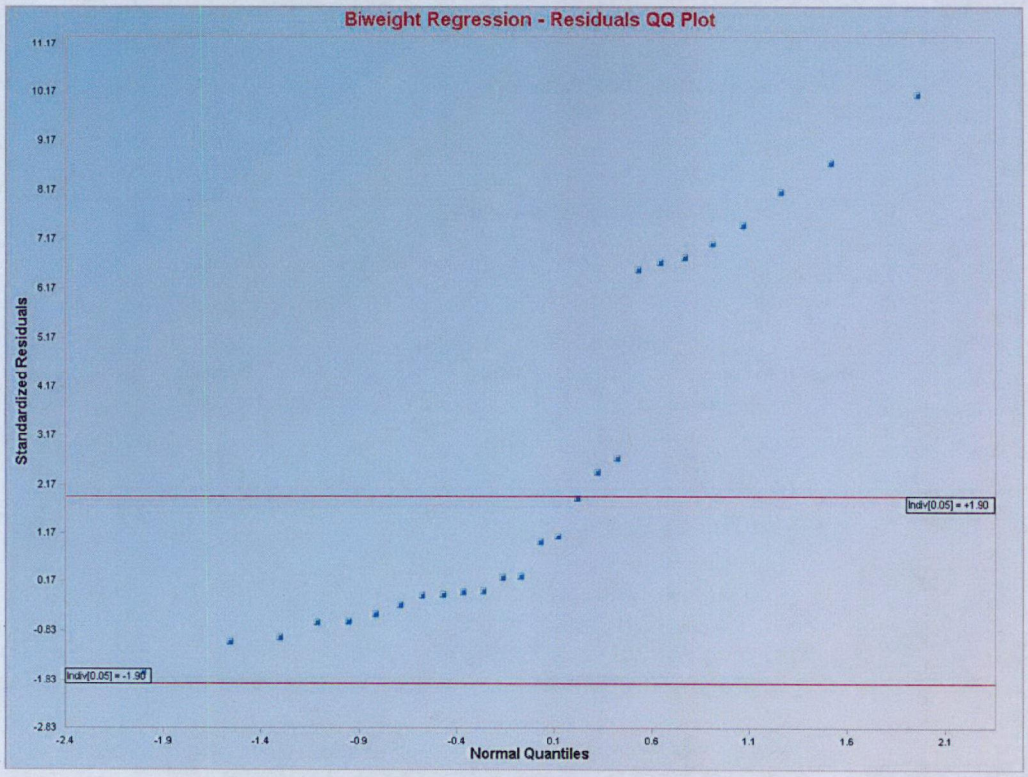
Final Weighted Correlation Matrix

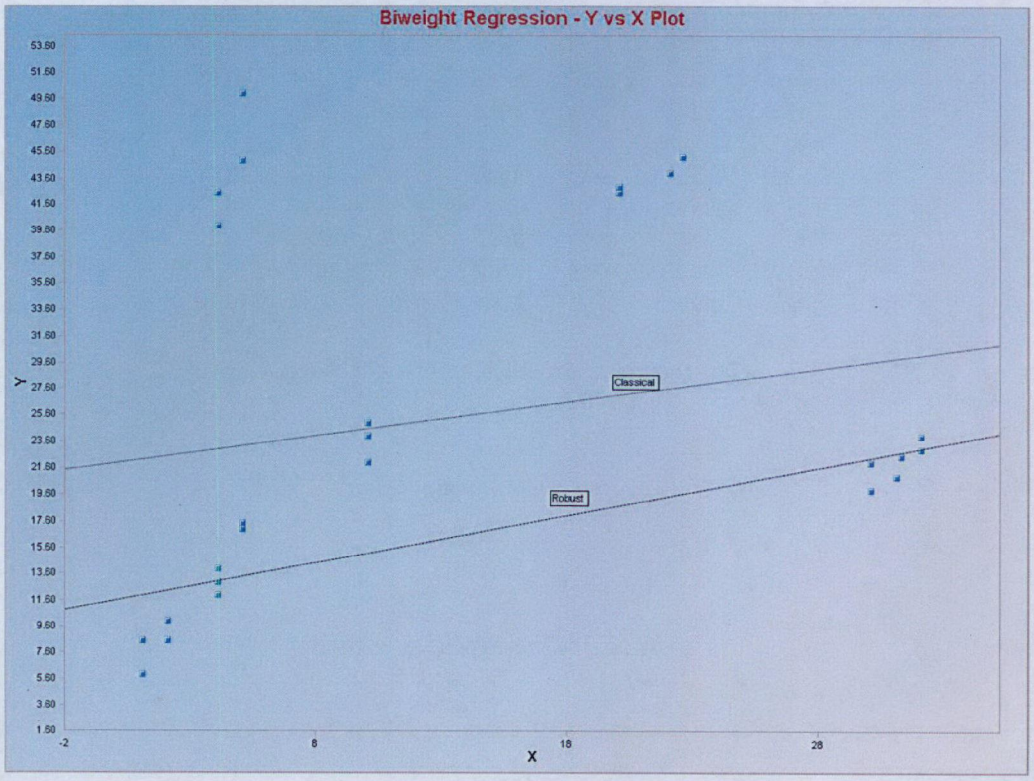
	Y	X
Y	1	0 867
X	0 867	1

Eigenvalues of Final Weighted Correlation Matrix

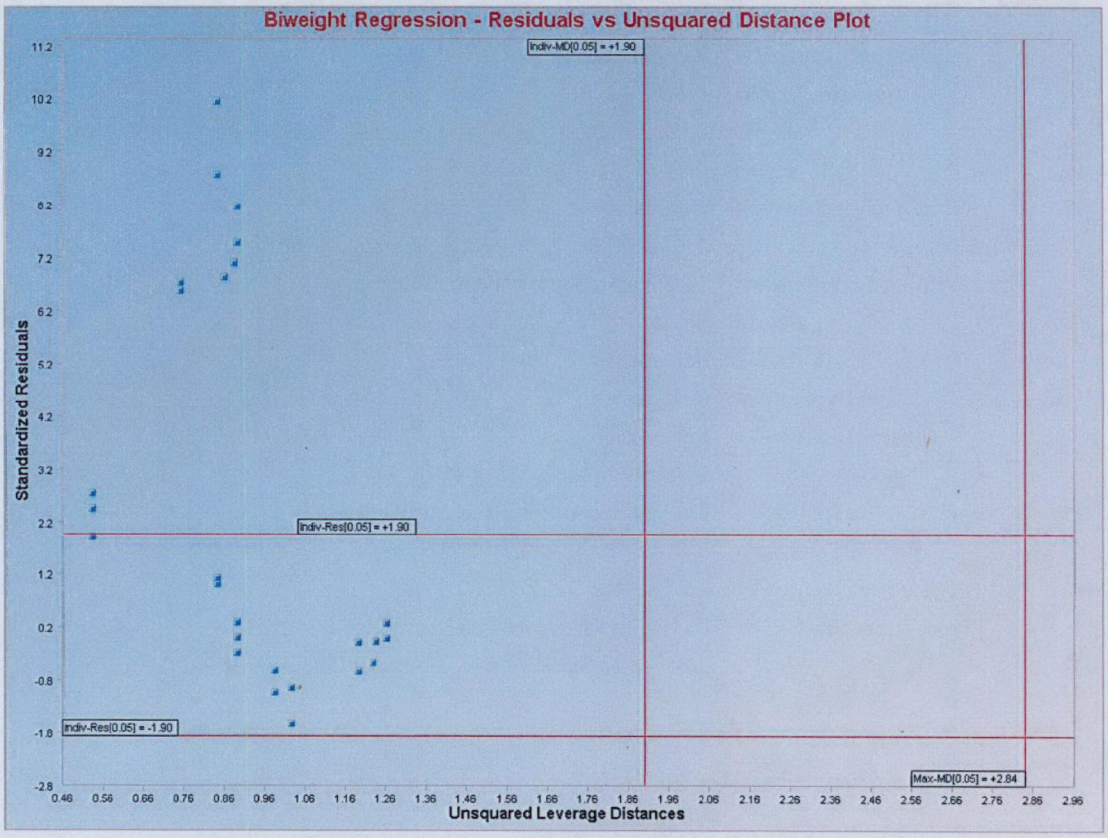
Eval 1	Eval 2
0 133	1 867

Output for Biweight (Leverage ON) (continued).





Output for Biweight (Leverage ON) (continued).



Interpretation of Graphs: Observations which are outside of the horizontal lines in the graphs are considered to be regression outliers. The observations to the right of the vertical lines are considered to be leverage outliers. The regression lines are produced since there is only one predictor variable.

Output for Biweight (Leverage OFF).

	Regression Analysis Output
Date/Time of Computation	3/5/2008 7:38:34 AM
User Selected Options	
From File	D:\Narain\Scout_For_Windows\ScoutSource\WorkDatInExcel\DEFINE
Full Precision	OFF
Selected Regression Method	Biweight
Residual Biweight Tuning Constant	4 (Used to Identify Vertical Regression Outliers)
Number of Regression Iterations	10 (Maximum Number if doesn't Converge)
Leverage	Off
Y vs Y-hat Plot	Not Selected
Y vs Residual Plot	Not Selected
Y-hat vs Residual Plot	Not Selected
Title For Y vs X Plots	Biweight Regression - Y vs X Plot
Title for Residual QQ Plot	Biweight Regression - Residuals QQ Plot
Residual Band Alpha	0.05 (Used in Graphics Residual Bands)
Residual vs Distance Plot	Not Selected
Show Intermediate Results	Do Not Display Intermediate Results
	Intermediate Results Shown on Another Output Sheet

Number of Selected Regression Variables	1
Number of Observations	26
Dependent Variable	Y

Residual Values used with Graphics Display	
Upper Residual Indvidual (0.05) MD	1.903
Lower Residual Indvidual (0.05) MD	-1.903

Correlation Matrix

	Y	X
Y	1	0.218
X	0.218	1

Eigenvalues of Correlation Matrix

Eval 1	Eval 2
0.782	1.218

Output for Biweight (Leverage OFF) (continued).

Ordinary Least Squares (OLS) Regression Results

Estimates of Regression Parameters

Intercept	X
22 06	0 256

Stdv of Estimated Regression Parameters

Intercept	X
4 107	0 233

ANOVA Table

Source of Variation	SS	DOF	MS	F-Value	P-Value
Regression	224.9	1	224.9	1.202	0.2838
Error	4490	24	187 1		
Total	4715	25			

R Square Estimate	0 0477
MAD Based Scale Estimate	8.862
Weighted Scale Estimate	13.68
IQR Estimate of Residuals	25 79
Det. of COV[Regression Coefficients] Matrix	0.391

Final Reweighted Regression Results

Estimates of Regression Parameters

Intercept	X
11 64	0 358

Stdv of Estimated Regression Parameters

Intercept	X
1.372	0.0696

ANOVA Table

Source of Variation	SS	DOF	MS	F-Value	P-Value
Regression	346	1	346	26 45	0.0002
Error	173 3	13 25	13.08		
Total	519.3	14 25			

R Square Estimate	0.666
MAD Based Scale Estimate	7 695
Weighted Scale Estimate	3 617
IQR Estimate of Residuals	25.53
Det of COV[Regression Coefficients] Matrix	0 00416

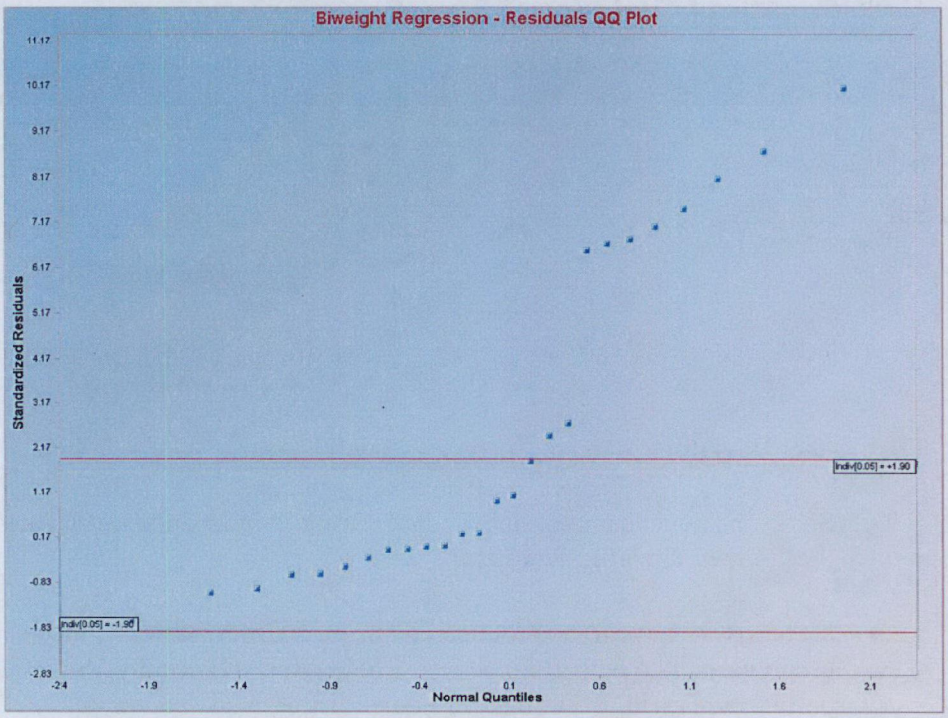
Output for Biweight (Leverage OFF) (continued).

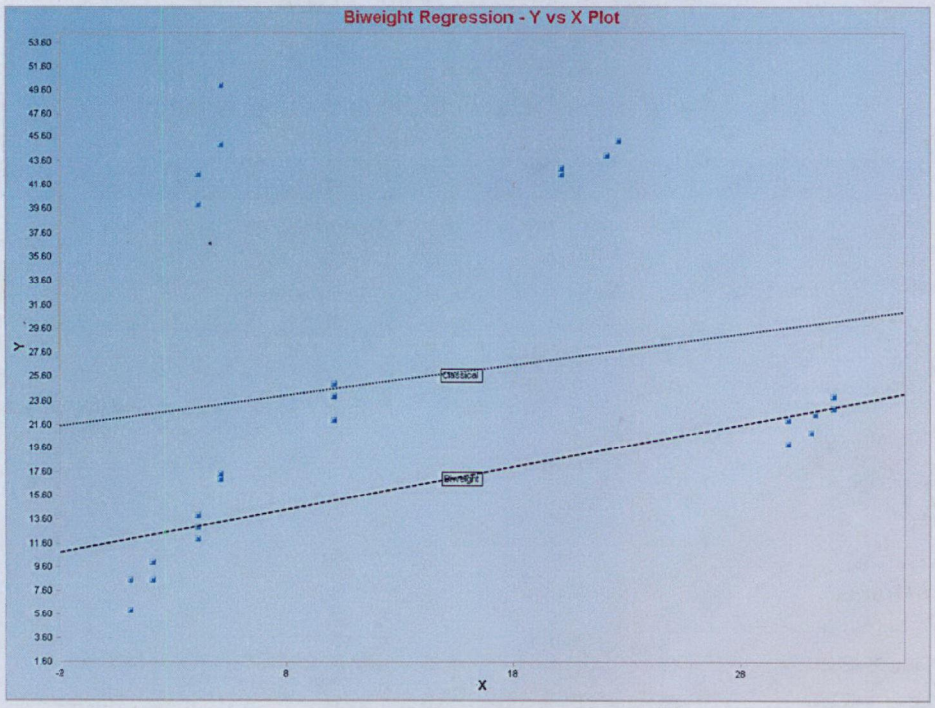
Regression Table								
Obs	Y Vector	Yhat	Residuals	Hat[i,i]	Res/Scale	Stude~ Res	Wts[i,i]	Res Dist.
1	6	11.99	-5.993	0.0827	-1.657	-1.73	0.678	1 657
2	8.5	11.99	-3.493	0.0827	-0.966	-1 008	0 883	0.966
3	8.5	12.35	-3.852	0.0758	-1.065	-1.108	0.859	1.065
4	10	12 35	-2.352	0.0758	-0 65	-0.676	0.946	0.65
5	12	13.07	-1.068	0.0638	-0.295	-0.305	0.989	0 295
6	40	13.07	26 93	0.0638	7.446	7.696	0	7.446
7	42.5	13.07	29.43	0.0638	8.138	8.41	0	8 138
8	45	13.43	31.57	0.0587	8.73	8.998	0	8 73
9	50	13.43	36.57	0.0587	10.11	10.42	0	10.11
10	13	13.07	-0.0681	0.0638	-0.0188	-0 0195	1	0.0188
11	14	13.07	0.932	0.0638	0.258	0 266	0.992	0.258
12	17	13.43	3.574	0.0587	0 988	1.018	0.879	0.988
13	17.4	13.43	3.974	0.0587	1 099	1.132	0.851	1.099
14	22	15.22	6.783	0.0417	1.875	1.916	0.599	1.875
15	24	15.22	8 783	0.0417	2.428	2.481	0.386	2.428
16	25	15 22	9.783	0.0417	2.705	2.763	0.281	2.705
17	42.5	18.8	23.7	0.0514	6.553	6.728	0	6.553
18	43	18.8	24 2	0.0514	6.691	6.87	0	6 691
19	44.1	19.52	24.58	0.0603	6.797	7.012	0	6.797
20	45.3	19.7	25 6	0.0629	7.079	7.313	0	7.079
21	20	22.38	-2.382	0.119	-0.659	-0.702	0.945	0.659
22	22	22.38	-0.382	0.119	-0.106	-0.113	0.999	0 106
23	21	22.74	-1.74	0.129	-0.481	-0.516	0 97	0.481
24	23	23.1	-0.0984	0.14	-0.0272	-0.0293	1	0 0272
25	22.5	22.81	-0.312	0.131	-0 0862	-0 0925	0.999	0.0862
26	24	23.1	0.902	0.14	0 249	0.269	0.992	0.249

Final Weighted Correlation Matrix		
	Y	X
Y	1	0.867
X	0 867	1

Eigenvalues of Final Weighted Correlation Matrix	
Eval 1	Eval 2
0.133	1.867

Output for Biweight (Leverage OFF) (continued).

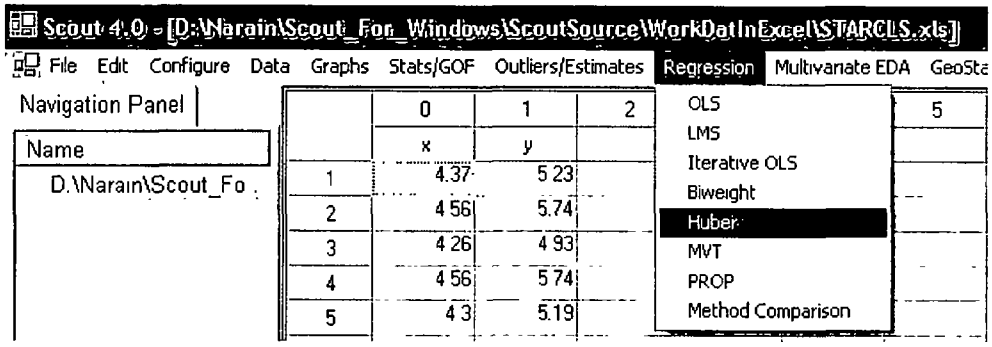




Interpretation of Graphs: Observations which are outside of the horizontal lines in the graph are considered to be regression outliers. The Leverage Distances vs. Standardized residuals plot is not produced even if checked on. The regression lines are produced since there is only one predictor variable.

9.6 Huber Regression Method

1. Click **Regression ▶ Huber.**

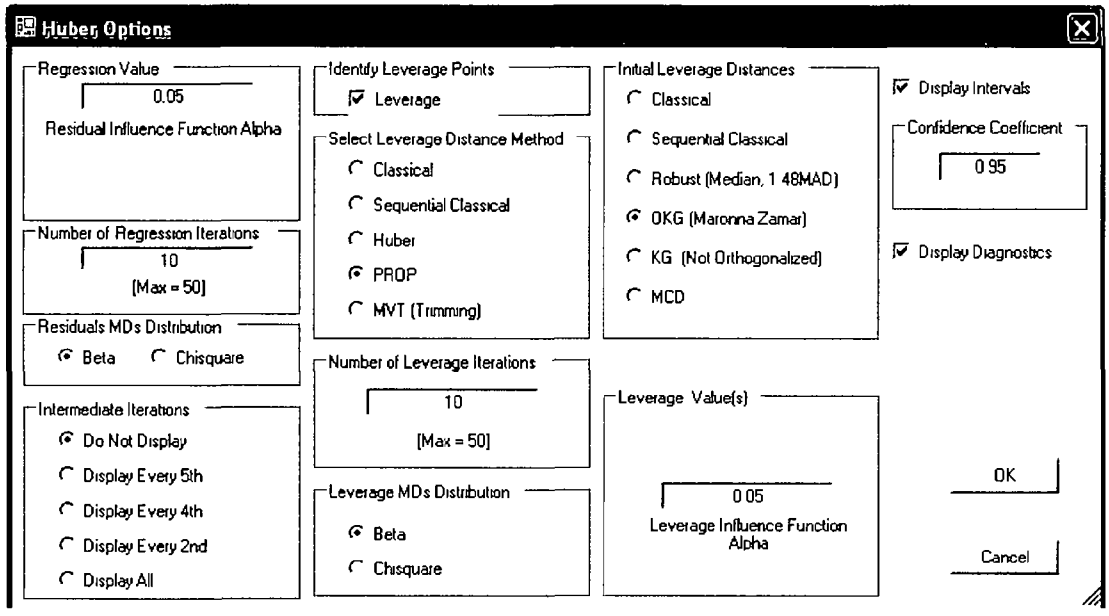


2. The "**Select Variables**" screen (Section 3.3) will appear.

   - Select the dependent variable and one or more independent variables from the "**Select Variables**" screen.

   - If the results have to be produced by using a Group variable, then select a group variable by clicking the arrow below the "**Group by Variable**" button. This will result in a drop-down list of available variables. The user should select and click on an appropriate variable representing a group variable.

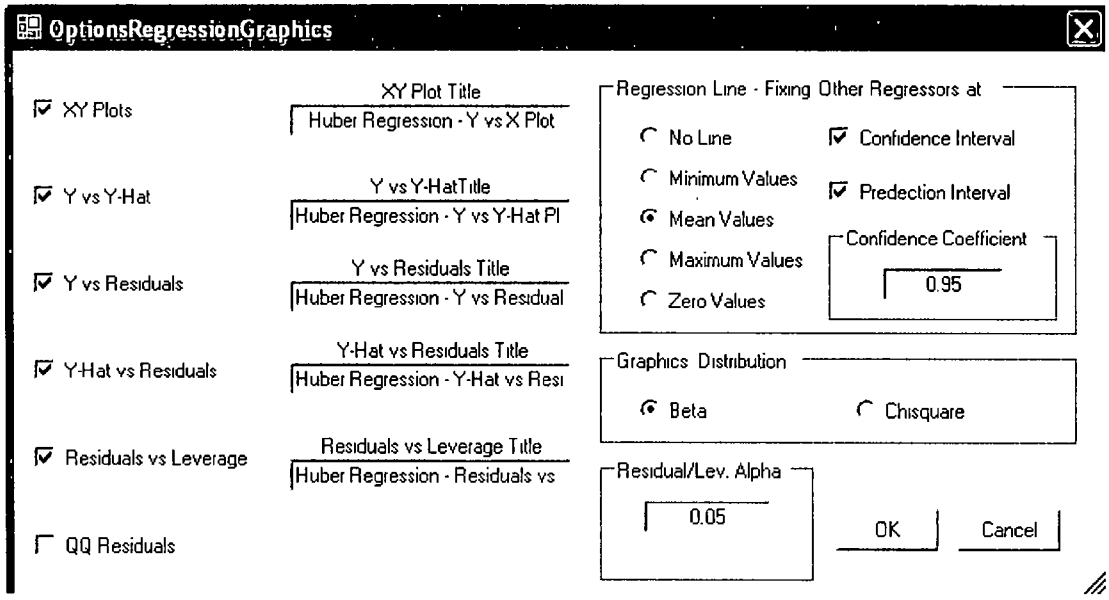
   - Click on the "**Options**" button to get the options window.



   - Specify the "**Regression Value.**" The default is "**0.05.**"

- Specify the "**Number of Regression Iterations.**" The default is"**10.**"
- Specify the "**Regression MDs Distribution.**" The default is "**Beta.**"
- Specify the "**Identify Leverage Points.**" The default is "**On.**"
- Specify the "**Select Leverage Distance Method.**" The default is "**PROP.**"
- Specify the "**Number of Leverage Iterations.**" The default is"**10.**"
- Specify the "**Leverage Initial Distances.**" The default is "**OKG (Maronna Zamar).**"
- Specify the "**Leverage Value.**" The default is "**0.05.**"
- Click "**OK**" to continue or "**Cancel**" to cancel the options.

- Click on the "**Graphics**" button to get the options window.



- Specify the preferred plots and the input parameters.
- Click "**OK**" to continue or "**Cancel**" to cancel the options.

- Click "**OK**" to continue or "**Cancel**" to cancel the computations.

Output example: The data set "**BRADU.xls**" was used for Huber regression. It has 3 predictor variables (p) and 75 observations. When the "**Leverage**" option is on, the leverage distances are calculated and outlying observations are obtained iteratively using initial estimates as median and OKG matrix and the leverage option as PROP (i.e., using PROP influence function). Then the weights are assigned to observations and those weights are used in the finding the regression outliers iteratively. When the leverage option is off, all observations are assigned one (1) as weights and then the regression outliers are found using the Huber function iteratively. Finally, the estimated regression parameters are calculated.

Output for Huber (Leverage ON).
Data Set Used: Bradu (predictor variables p = 3).

Date/Time of Computation	3/5/2008 7 51 39 AM
User Selected Options	
From File	D \Narain\Scout_For_Windows\ScoutSource\WorkDatInExcel\BRADU
Full Precision	OFF
Selected Regression Method	Huber
Residual Influence Function Alpha	0 05 (Used to Identify Vertical Regression Outliers)
Number of Regression Iterations	10 (Maximum Number if doesn't Converge)
Leverage	Identify Leverage Points (Outliers in X-Space)
Selected Leverage Method	PROP
Initial Leverage Distance Method	OKG (Maronna Zamar) Matrix
Squared MDs	Beta Distribution used for Leverage Distances based upon Selected Regression (Leverage) Variables
Leverage Distance Alpha	0.05 (Used to Identify Leverage Points)
Number of Leverage Iterations	10 (Maximum Number if doesn't Converge)
Y vs Y-hat Plot	Not Selected
Y vs Residual Plot	Not Selected
Y-hat vs Residual Plot	Not Selected
Title For Y vs X Plots	Huber Regression - Y vs X Plot
Title for Residual QQ Plot	Huber Regression - Residuals QQ Plot
Residual Band Alpha	0 05 (Used in Graphics Residual Bands)
Title Residual vs Distance Plot	Huber Regression - Residuals vs Unsquared Leverage Distance Plot
Show Intermediate Results	Do Not Display Intermediate Results
	Intermediate Results Shown on Another Output Sheet

Leverage Points are Outliers in X-Space of Selected Regression Variables.

Number of Selected Regression Variables	3
Number of Observations	75
Dependent Variable	y

Residual Values used with Graphics Display

Upper Residual Individual (0 05) MD	1.94
Lower Residual Individual (0 05) MD	-1 94

Correlation Matrix

	y	x1	x2	x3
y	1	0 946	0 962	0 743
x1	0 946	1	0 979	0 708
x2	0 962	0 979	1	0 757
x3	0 743	0 708	0 757	1

Output for Huber (Leverage ON) (continued).

Eigenvalues of Correlation Matrix

Eval 1	Eval 2	Eval 3	Eval 4
0.0172	0.0556	0.368	3.559

Ordinary Least Squares (OLS) Regression Results

Estimates of Regression Parameters

Intercept	x1	x2	x3
-0.388	0 239	-0 335	0.383

Stdv of Estimated Regression Parameters

Intercept	x1	x2	x3
0.416	0.262	0 155	0.129

ANOVA Table

Source of Variation	SS	DOF	MS	F-Value	P-Value
Regression	543.3	3	181.1	35.77	0.0000
Error	359.5	71	5.063		
Total	902.8	74			

R Square Estimate	0 602
MAD Based Scale Estimate	1.067
Weighted Scale Estimate	2.25
IQR Estimate of Residuals	1.468
Det. of COV[Regression Coefficients] Matrix	5.5107E-8

Initial Weighted Regression Iteration with Identified Leverage Points

Estimates of Regression Parameters

Intercept	x1	x2	x3
-0.0105	0 0624	0 0119	-0 107

Stdv of Estimated Regression Parameters

Intercept	x1	x2	x3
0.197	0.0689	0.0684	0.0713

Output for Huber (Leverage ON) (continued).

ANOVA Table								
Source of Variation	SS	DOF	MS	F-Value	P-Value			
Regression	0.898	3	0 299	0.94	0 4272			
Error	18.14	57	0.318					
Total	19.04	60						
R Square Estimate		0 0472						
MAD Based Scale Estimate		0.902						
Weighted Scale Estimate		0 564						
Unsquared Leverage Distance Indiv-MD(0 05)		2 743						
IQR Estimate of Residuals		1.236						
Determinant of Leverage S Matrix		1.357						

Regression Table with Leverage Option

Obs	Y Vector	Yhat	Residuals	Hat[i,i]	Res/Scale	Stude~ Res	Wts[i,i]	Res Dist	Lev Dist	OLS R~ist
1	9.7	-2.174	11 87	0 063	21.05	21.74	2.364E-13	21 05	29.44	1.502
2	10 1	-2 265	12.36	0.0599	21 92	22.61	1.074E-13	21 92	30 21	1 775
3	10 3	-2 418	12.72	0.0857	22 54	23 58	1 886E-14	22 54	31 89	1 334
4	9 5	-2.528	12.03	0 0805	21 32	22.23	6 931E-15	21 32	32 86	1.138
5	10	-2 443	12.44	0 0729	22 06	22 91	1 266E-14	22 06	32 28	1.36
6	10	-2 217	12.22	0 0756	21 66	22.52	7.228E-14	21 66	30.59	1.527
7	10 8	-2 219	13.02	0 068	23 08	23 9	6 576E-14	23 08	30 68	2 006
8	10.3	-2 24	12.54	0 0631	22 23	22.97	1 634E-13	22 23	29 8	1.705
9	9 6	-2 475	12.07	0 08	21 4	22 32	1 768E-14	21 4	31 95	1 204
10	9 9	-2 437	12.34	0.0869	21 87	22 89	5 017E-14	21 87	30.94	1 35
11	-0 2	-2 782	2 582	0 0942	4.577	4.809	1.424E-16	4 577	36 64	3.48
12	-0.4	-2.946	2 546	0 144	4.513	4.877	3.684E-17	4.513	37.96	4.165
13	0 7	-2.589	3.289	0.109	5.83	6.177	1 069E-16	5 83	36.92	2.719
14	0.1	-2.556	2.656	0.564	4.708	7.127	1.478E-18	4.708	41.09	1.69
15	-0 4	0 0115	-0.412	0 0579	-0.73	-0.752	1	0.73	2 002	0 294
16	0 6	0 177	0.423	0 0759	0.75	0.78	1	0 75	2 165	0.385
17	-0.2	-0.0128	-0.187	0.0393	-0.332	-0 339	1	0.332	1.938	0.287
18	0	-0 0619	0.0619	0.0231	0.11	0.111	1	0 11	0 786	0.175
19	0 1	-0 0971	0 197	0 0312	0 349	0.355	1	0 349	1 287	0 29
20	0 4	-0 0119	0.412	0 0476	0 73	0 748	1	0 73	2 067	0.151
21	0 9	-0 0253	0 925	0 0294	1 64	1 665	1	1.64	1 059	0 299
22	0 3	-0.151	0.451	0 0457	0 799	0 818	1	0 799	1 746	0 415
23	-0 8	0 0561	-0 856	0 0293	-1 518	-1 54	1	1 518	1.163	0.19

(The complete regression table is not shown.)

Output for Huber (Leverage ON) (continued).

Final Reweighted Regression Results

Estimates of Regression Parameters

Intercept	x1	x2	x3
-0.413	0 237	-0.382	0.44

Stdv of Estimated Regression Parameters

Intercept	x1	x2	x3
0 378	0 238	0 142	0.118

ANOVA Table

Source of Variation	SS	DOF	MS	F-Value	P-Value
Regression	610	3	203.3	48.96	0 0000
Error	290.9	70 05	4.153		
Total	900.9	73 05			

R Square Estimate	0.677
MAD Based Scale Estimate	1.158
Weighted Scale Estimate	2.038
IQR Estimate of Residuals	1 579
Det of COV[Regression Coefficients] Matrix	2.8216E-8

Regression Table

Obs	Y Vector	Yhat	Residuals	Hat[i,i]	Res/Scale	Stude~ Res	Wts[i,i]	Res Dist
1	9.7	6.944	2.756	0.063	1.352	1.397	1	1.352
2	10 1	6.722	3 378	0 0599	1 657	1.709	1	1.657
3	10.3	8 045	2 255	0 0857	1 106	1.157	1	1.106
4	9.5	7.667	1 833	0.0805	0.9	0.938	1	0.9
5	10	7.651	2 349	0.0729	1.153	1.197	1	1.153
6	10	7.201	2 799	0 0756	1 374	1.429	1	1 374

(The complete regression table is not shown.)

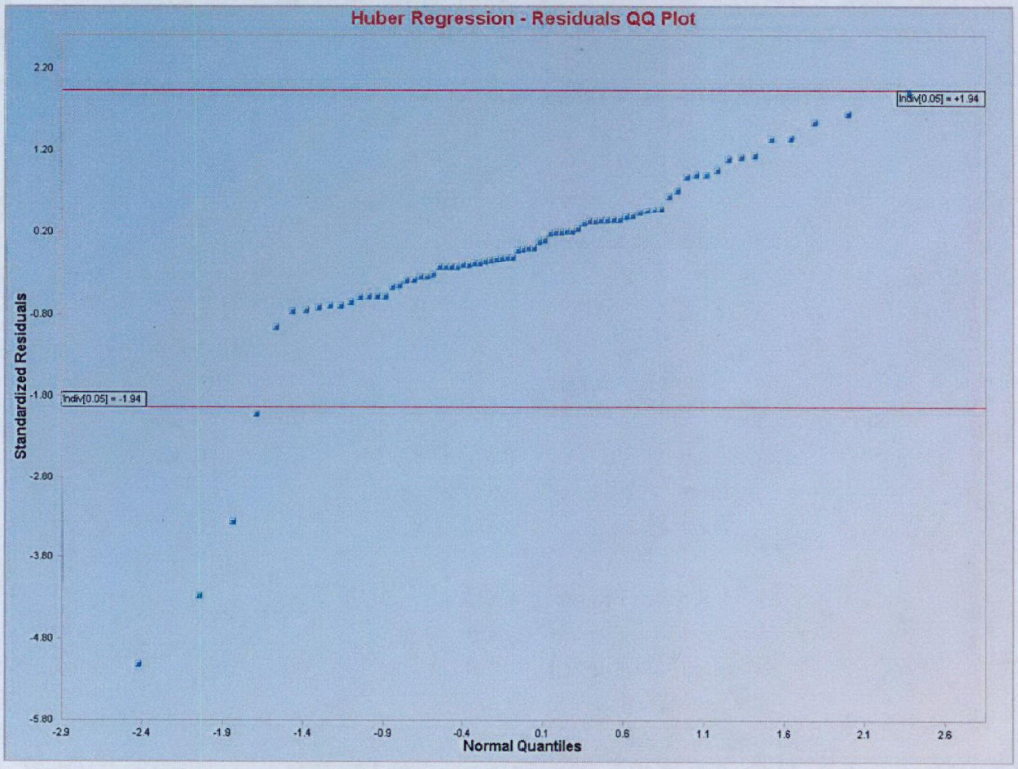
Final Weighted Correlation Matrix

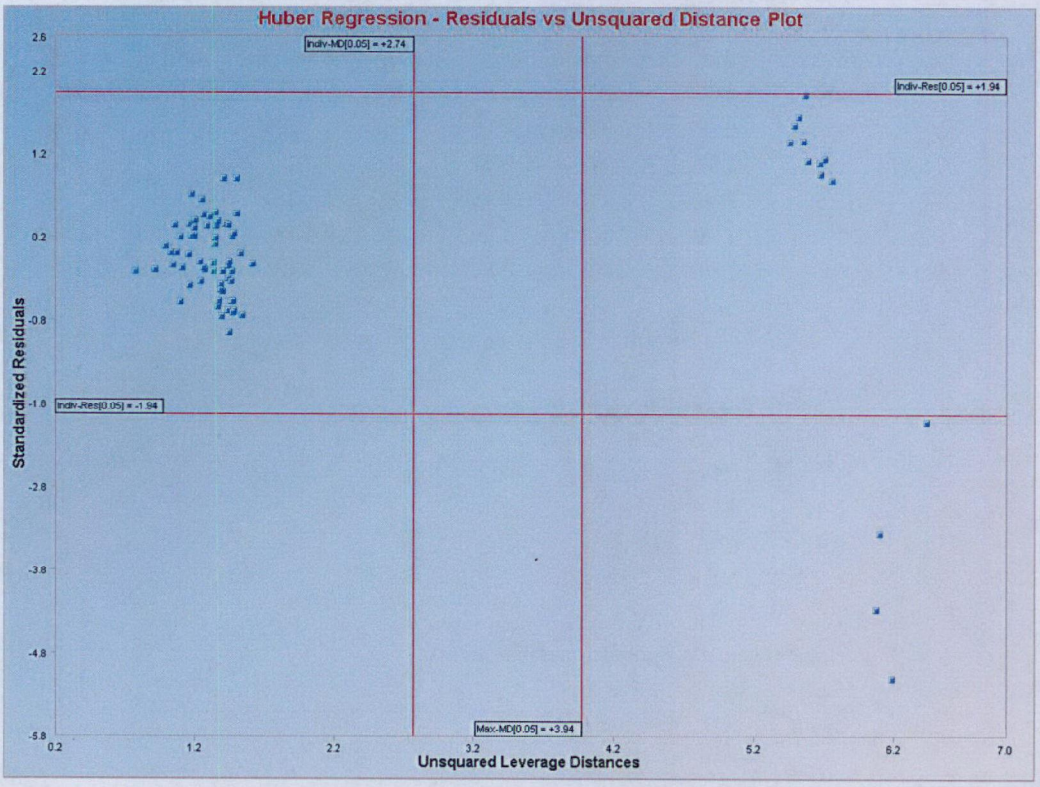
	y	x1	x2	x3
y	1	0 94	0 957	0.813
x1	0 94	1	0 977	0 774
x2	0 957	0.977	1	0.833
x3	0 813	0.774	0 833	1

Eigenvalues of Final Weighted Correlation Matrix

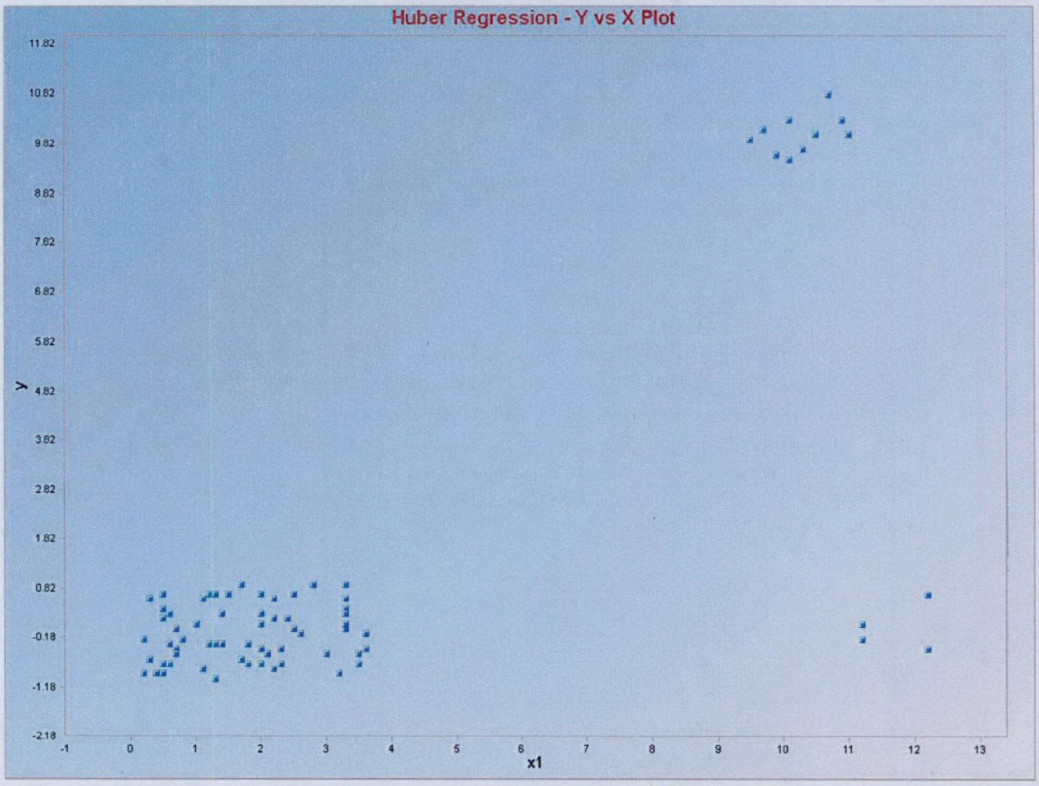
Eval 1	Eval 2	Eval 3	Eval 4
0.0165	0.0618	0 27	3 652

Output for Huber (Leverage ON) (continued).





Output for Huber (Leverage ON) (continued).



Interpretation of Graphs: Observations which are outside of the horizontal lines in the graphs are considered to be regression outliers. The observations to the right of the vertical lines are considered to be leverage outliers. Regression lines are not produced since there are three predictor variables. Select other "**X**" variables by using the drop-down bar in the graphics panel and click on "**Redraw**."

Output for Huber (Leverage OFF).

Regression Analysis Output	
Date/Time of Computation	3/5/2008 8 15 51 AM
User Selected Options	
From File	D \Narain\Scout_For_Windows\ScoutSource\WorkDatInExcel\BRADU
Full Precision	OFF
Selected Regression Method	Huber
Residual Influence Function Alpha	0 05 (Used to Identify Vertical Regression Outliers)
Number of Regression Iterations	10 (Maximum Number if doesn't Converge)
Leverage	Off
Y vs Y-hat Plot	Not Selected
Y vs Residual Plot	Not Selected
Y-hat vs Residual Plot	Not Selected
Title For Y vs X Plots	Huber Regression - Y vs X Plot
Title for Residual QQ Plot	Huber Regression - Residuals QQ Plot
Residual Band Alpha	0 05 (Used in Graphics Residual Bands)
Title Residual vs Distance Plot	Huber Regression - Residuals vs Unsquared Leverage Distance Plot
Show Intermediate Results	Do Not Display Intermediate Results
	Intermediate Results Shown on Another Output Sheet

Number of Selected Regression Variables	3
Number of Observations	75
Dependent Variable	y

Residual Values used with Graphics Display	
Upper Residual Individual (0 05) MD	1.94
Lower Residual Individual (0 05) MD	-1 94

Correlation Matrix

	y	x1	x2	x3
y	1	0 946	0 962	0 743
x1	0 946	1	0.979	0 708
x2	0 962	0 979	1	0 757
x3	0 743	0 708	0 757	1

Eigenvalues of Correlation Matrix

Eval 1	Eval 2	Eval 3	Eval 4
0.0172	0 0556	0 368	3 559

Ordinary Least Squares (OLS) Regression Results

Estimates of Regression Parameters

Intercept	x1	x2	x3
-0 388	0 239	-0.335	0 383

Stdv of Estimated Regression Parameters

Intercept	x1	x2	x3
0 416	0 262	0 155	0 129

ANOVA Table

Source of Variation	SS	DOF	MS	F-Value	P-Value
Regression	543 3	3	181.1	35 77	0 0000
Error	359 5	71	5 063		
Total	902 8	74			

Output for Huber (Leverage OFF) (continued).

R Square Estimate	0 602
MAD Based Scale Estimate	1 067
Weighted Scale Estimate	2 25
IQR Estimate of Residuals	1.468
Det of COV[Regression Coefficients] Matrix	5 5107E-8

Final Reweighted Regression Results

Estimates of Regression Parameters

Intercept	x1	x2	x3
-0 413	0 237	-0 382	0 44

Stdv of Estimated Regression Parameters

Intercept	x1	x2	x3
0 378	0 238	0 142	0 118

ANOVA Table

Source of Variation	SS	DOF	MS	F-Value	P-Value
Regression	610	3	203 3	48 96	0.0000
Error	290 9	70.05	4.153		
Total	900 9	73 05			

R Square Estimate	0 677
MAD Based Scale Estimate	1.158
Weighted Scale Estimate	2 038
IQR Estimate of Residuals	1 578
Det. of COV[Regression Coefficients] Matrix	2 8216E-8

Regression Table

Obs	Y Vector	Yhat	Residuals	Hat[i,i]	Res/Scale	Stude~ Res	Wts[i,i]	Res Dist.
1	9 7	6.944	2.756	0 063	1 352	1 397	1	1 352
2	10 1	6 722	3 378	0 0599	1 657	1 709	1	1 657
3	10.3	8 045	2 255	0 0857	1 106	1 157	1	1 106
4	9 5	7 667	1 833	0 0805	0 9	0 938	1	0 9
5	10	7 651	2 349	0 0729	1 153	1 197	1	1 153
6	10	7 201	2 799	0 0756	1 374	1 429	1	1 374
7	10 8	6 895	3 905	0 068	1 916	1 985	1	1.916

(The complete regression table is not shown.)

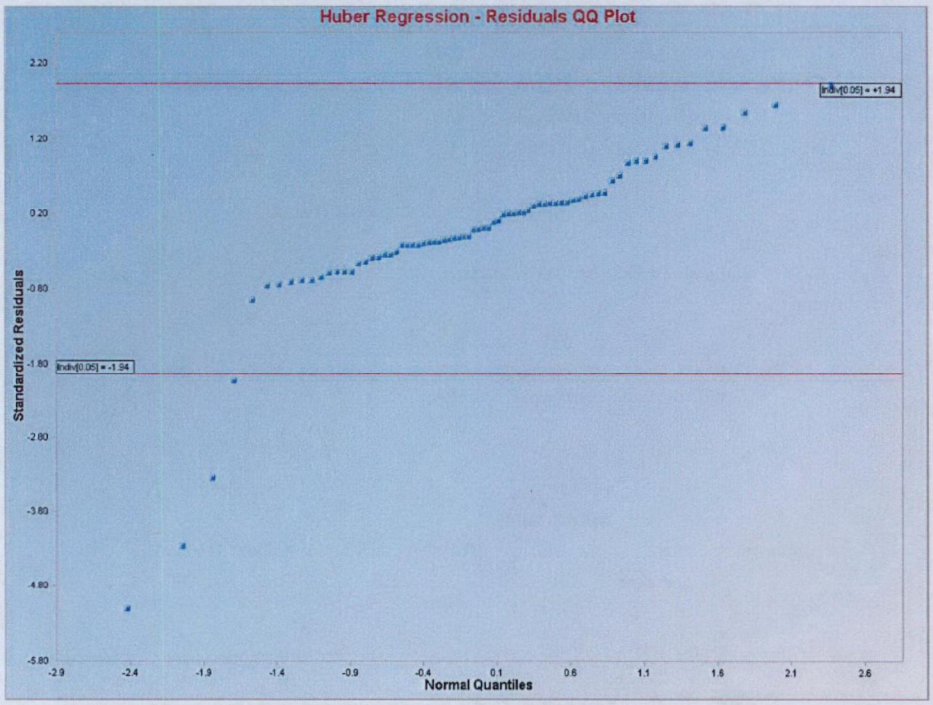
Final Weighted Correlation Matrix

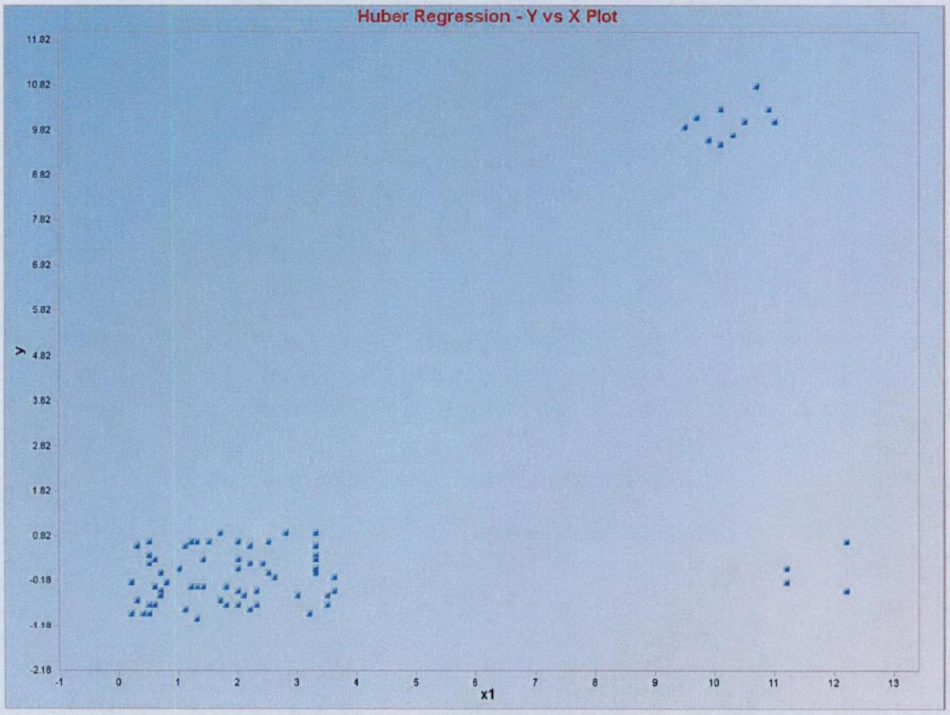
	y	x1	x2	x3
y	1	0 94	0 957	0 813
x1	0 94	1	0 977	0 774
x2	0 957	0 977	1	0 833
x3	0 813	0 774	0 833	1

Eigenvalues of Final Weighted Correlation Matrix

Eval 1	Eval 2	Eval 3	Eval 4
0 0165	0 0618	0 27	3 652

Output for Huber (Leverage OFF) (continued).

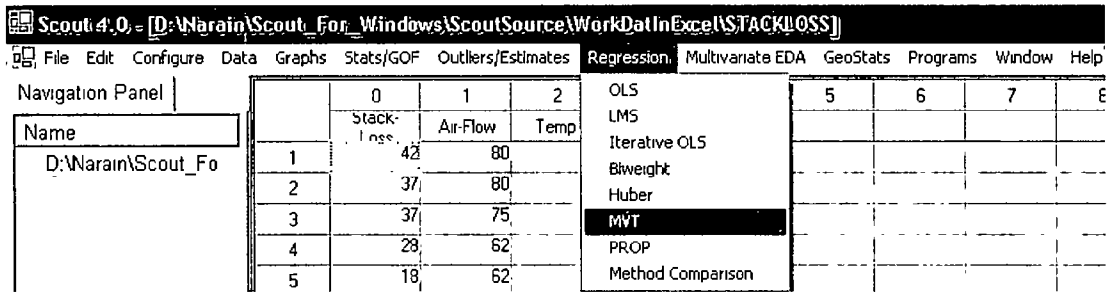




Interpretation of Graphs: Observations which are outside of the horizontal lines in the graph are considered to be regression outliers. The Leverage Distances vs. Standardized residuals plot is not produced even if checked on. Regression lines are not produced since there are three predictor variables. Select other "**X**" variables by using the drop-down bar in the graphics panel and click on "**Redraw**."

9.7 MVT Regression Method

1. Click **Regression ▶ MVT.**

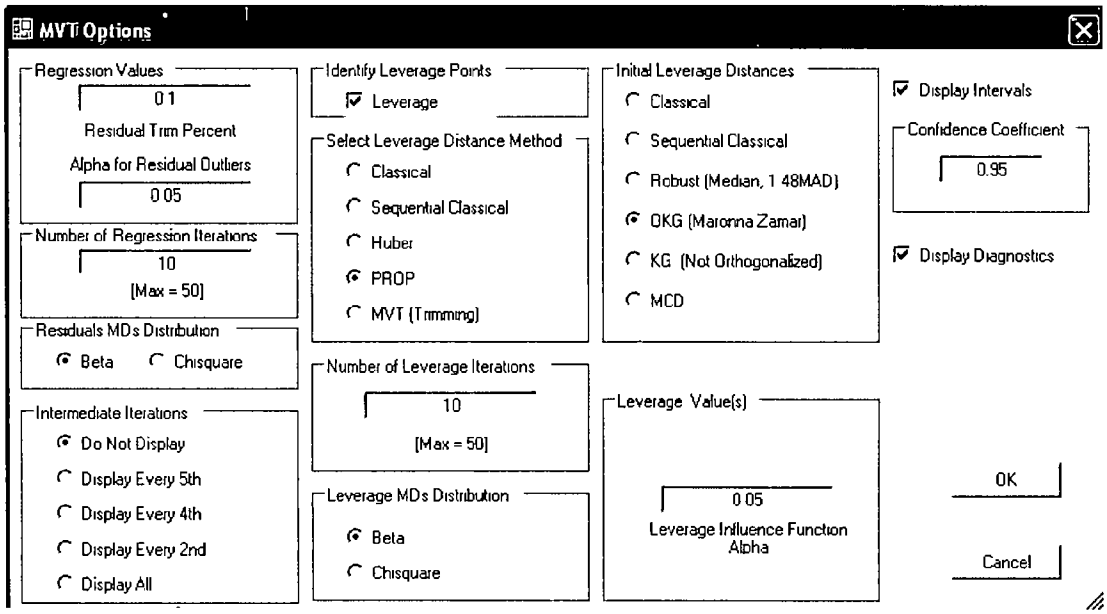


2. The "**Select Variables**" screen (Section 3.3) will appear.

   - Select the dependent variable and one or more independent variables from the "**Select Variables**" screen.

   - If the results have to be produced by using a Group variable, then select a group variable by clicking the arrow below the "**Group by Variable**" button. This will result in a drop-down list of available variables. The user should select and click on an appropriate variable representing a group variable.

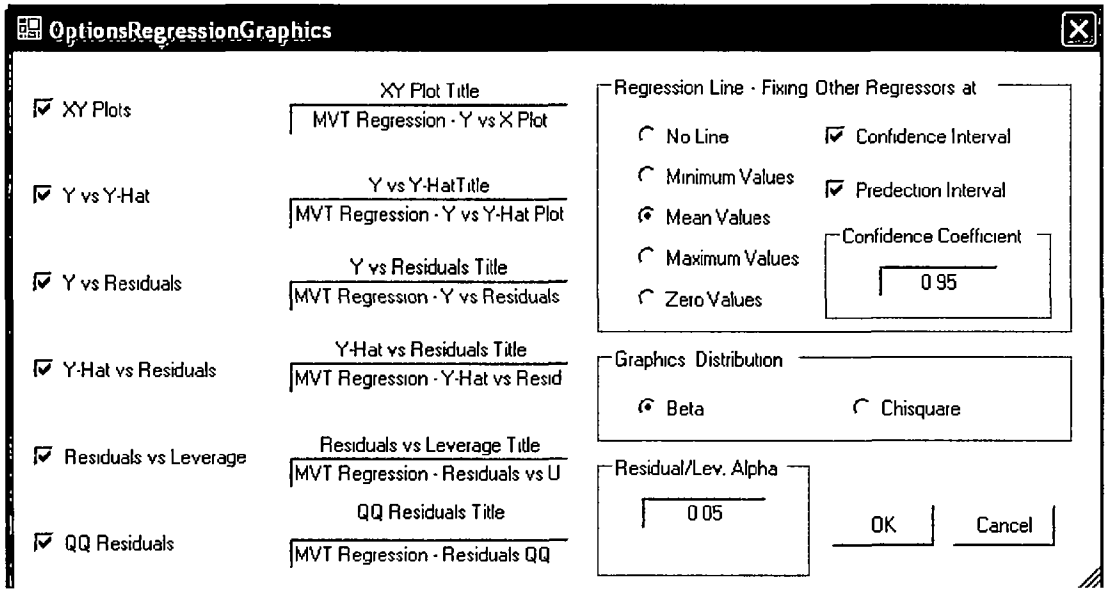
   - Click on the "**Options**" button to get the options window.



   - Specify the "**Regression Value.**" The default is "**0.05.**"

- Specify the "**Number of Regression Iterations.**" The default is"**10.**"
- Specify the "**Regression MDs Distribution.**" The default is "**Beta.**"
- Specify the "**Identify Leverage Points.**" The default is "**On.**"
- Specify the "**Select Leverage Distance Method.**" The default is "**PROP.**"
- Specify the "**Number of Leverage Iterations.**" The default is"**10.**"
- Specify the "**Leverage Initial Distances**" The default is "**OKG (Maronna Zamar).**"
- Specify the "**Leverage Value.**" The default is "**0.05.**"
- Click "**OK**" to continue or "**Cancel**" to cancel the options.

- Click on the "**Graphics**" button to get the options window.



- Specify the preferred plots and the input parameters.
- Click "**OK**" to continue or "**Cancel**" to cancel the options.

- Click "**OK**" to continue or "**Cancel**" to cancel the computations.

Output example: The data set "**STACKLOSS.xls**" was used for MVT regression. It has 3 predictor variables (p) and 21 observations. When the "**Leverage**" option is on, the leverage distances are calculated and outlying observations are obtained iteratively using initial estimates as median and OKG matrix and the leverage option as PROP (i.e., using PROP influence function). Then the weights are assigned to observations and those weights are used in the finding the regression outliers iteratively. When the leverage option is off, all observations are assigned one (1) as weights and then the regression outliers are found using the trimming percentage and a critical alpha iteratively. Finally, the estimated regression parameters are calculated.

Output for MVT (Leverage ON).
Data Set Used: Stackloss (predictor variables p = 3).

	Regression Analysis Output
Date/Time of Computation	3/5/2008 8 22 37 AM
User Selected Options	
From File	D:\Narain\Scout_For_Windows\ScoutSource\WorkDatInExcel\STACKLOSS
Full Precision	OFF
Selected Regression Method	Multivariate Trimming (MVT)
Residual MVT Trimming Percentage	0 1 (Used to Identify Vertical Regression Outliers)
Alpha for Residual Outliers	0 05 (Planned Future Modification Used to Compare Residual MVT MDs)
Number of Regression Iterations	10 (Maximum Number if doesn't Converge)
Leverage	Identify Leverage Points (Outliers in X-Space)
Selected Leverage Method	PROP
Initial Leverage Distance Method	OKG (Maronna Zamar) Matrix
Squared MDs	Beta Distribution used for Leverage Distances based upon Selected Regression (Leverage) Variables
Leverage Distance Alpha	0 05 (Used to Identify Leverage Points)
Number of Leverage Iterations	10 (Maximum Number if doesn't Converge)
Y vs Y-hat Plot	Not Selected
Y vs Residual Plot	Not Selected
Y-hat vs Residual Plot	Not Selected
Title For Y vs X Plots	MVT Regression - Y vs X Plot
Title for Residual QQ Plot	MVT Regression - Residuals QQ Plot
Residual Band Alpha	0 05 (Used in Graphics Residual Bands)
Title Residual vs Distance Plot	MVT Regression - Residuals vs Unsquared Leverage Distance Plot
Show Intermediate Results	Do Not Display Intermediate Results
	Intermediate Results Shown on Another Output Sheet

Leverage Points are Outliers in X-Space of Selected Regression Variables.

Number of Selected Regression Variables	3
Number of Observations	21
Dependent Variable	Stack-Loss

Residual Values used with Graphics Display

Upper Residual Individual (0.05) MD	1 889
Lower Residual Individual (0 05) MD	-1 889

Correlation Matrix

	Stack-Loss	Air-Flow	Temp	Acid-Conc
Stack-Loss	1	0 782	0 5	0 92
Air-Flow	0.782	1	0 391	0 876
Temp	0 5	0 391	1	0 4
Acid-Conc	0 92	0 876	0 4	1

Eigenvalues of Correlation Matrix

Eval 1	Eval 2	Eval 3	Eval 4
0 0532	0 215	0 734	2.997

Output for MVT (Leverage ON) (continued).

Ordinary Least Squares (OLS) Regression Results

Estimates of Regression Parameters

Intercept	Air-Flow	Temp	Acid-Conc
-39 92	0.716	1 295	-0 152

Stdv of Estimated Regression Parameters

Intercept	Air-Flow	Temp	Acid-Conc
11 9	0.135	0 368	0 156

ANOVA Table

Source of Variation	SS	DOF	MS	F-Value	P-Value
Regression	1890	3	630 1	59 9	0 0000
Error	178 8	17	10.52		
Total	2069	20			

R Square Estimate	0 914
MAD Based Scale Estimate	2 768
Weighted Scale Estimate	3 243
IQR Estimate of Residuals	4 313
Det of COV[Regression Coefficients] Matrix	1 0370E-5

Initial Weighted Regression Iteration with Identified Leverage Points

Estimates of Regression Parameters

Intercept	Air-Flow	Temp	Acid-Conc
-39 54	0 709	1 291	-0.151

Stdv of Estimated Regression Parameters

Intercept	Air-Flow	Temp	Acid-Conc
12.1	0 143	0 373	0 162

ANOVA Table

Source of Variation	SS	DOF	MS	F-Value	P-Value
Regression	1421	3	473 5	44 06	0 0000
Error	172 2	16 03	10 75		
Total	1593	19.03			

R Square Estimate	0 892
MAD Based Scale Estimate	2 738
Weighted Scale Estimate	3 278
Unsquared Leverage Distance Indiv-MD(0 05)	2 619
IQR Estimate of Residuals	4 169
Determinant of Leverage S Matrix	5200

Regression Table with Leverage Option

Obs	Y Vector	Yhat	Residuals	Hat[i,i]	Res/Scale	Stude~ Res	Wts[i,i]	Res Dist	Lev Dist.	OLS R~ist
1	42	38 58	3 417	0 302	1 042	1 247	0 562	1 042	2 931	0 997
2	37	39 73	-1 734	0 318	-0 529	-0 641	0 497	0 529	3 02	0 591
3	37	32 31	4 694	0 175	1 432	1 576	1	1 432	2 073	1 405

(The complete regression table is not shown.)

Output for MVT (Leverage ON) (continued).

Final Reweighted Regression Results

Estimates of Regression Parameters

Intercept	Air-Flow	Temp	Acid-Conc
-42.45	0.957	0 556	-0.109

Stdv of Estimated Regression Parameters

Intercept	Air-Flow	Temp	Acid-Conc
7.385	0.0945	0 264	0.0968

ANOVA Table

Source of Variation	SS	DOF	MS	F-Value	P-Value
Regression	1890	3	630	158.1	0 0000
Error	59 78	15	3.986		
Total	1950	18			

R Square Estimate	0 969
MAD Based Scale Estimate	2.069
Weighted Scale Estimate	1.996
IQR Estimate of Residuals	2.995
Det. of COV[Regression Coefficients] Matrix	3.4468E-7

Regression Table

Obs	Y Vector	Yhat	Residuals	Hat[i,i]	Res/Scale	Stude~ Res	Wts[i,i]	Res Dist.
1	42	39 4	2.604	0 302	1.305	1 561	1	1.305
2	37	39 5	-2.504	0 318	-1.254	-1 519	1	1.254
3	37	33 39	3.607	0 175	1.807	1 989	1	1.807
4	28	20 73	7.273	0 129	3.643	3 902	0	3.643
5	18	19 62	-1.616	0 0522	-0 81	-0.832	1	0.81
6	18	20 17	-2.172	0.0775	-1.088	-1.133	1	1.088

(The complete regression table is not shown.)

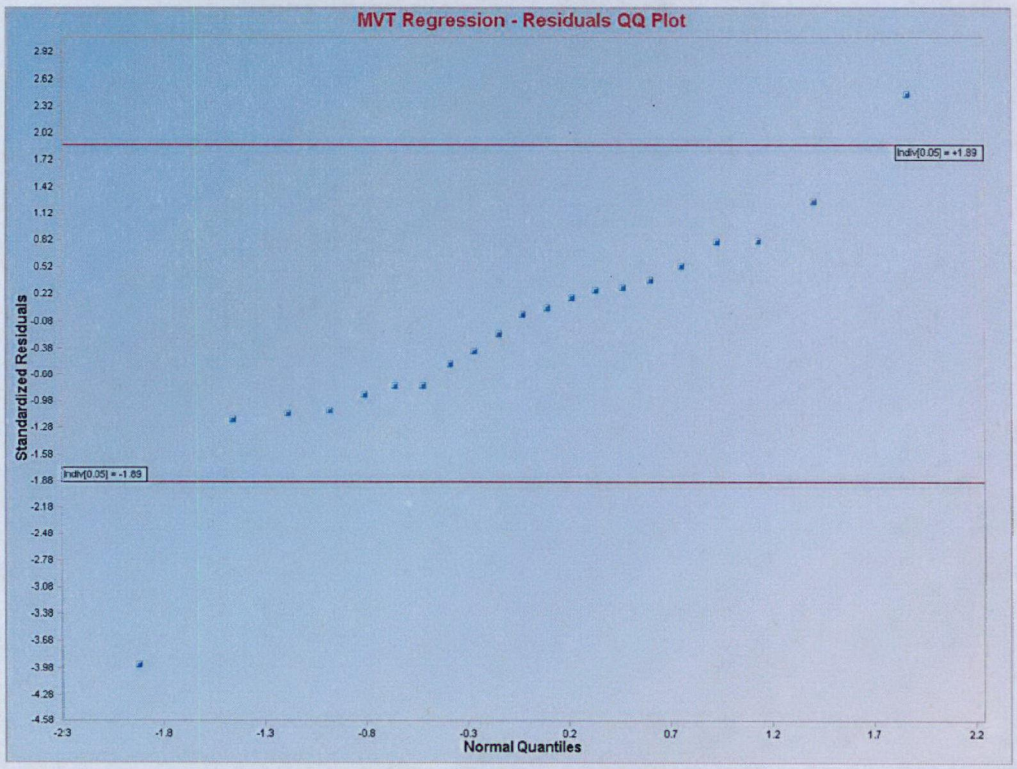
Final Weighted Correlation Matrix

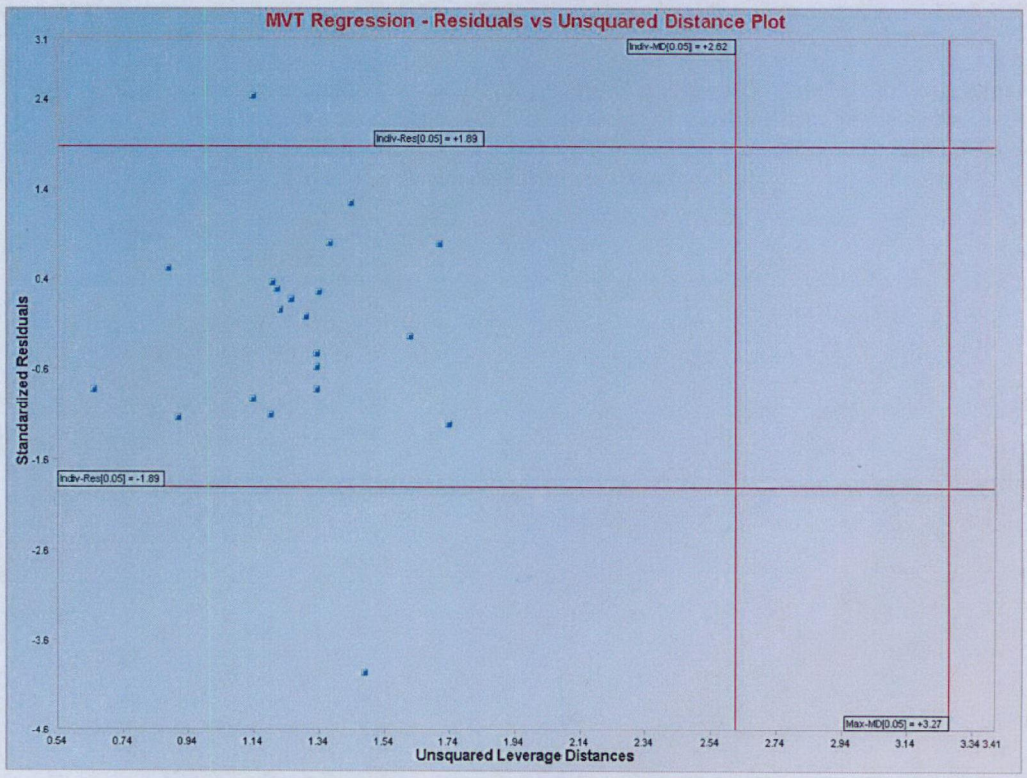
	Stack-Loss	Air-Flow	Temp	Acid-Conc
Stack-Loss	1	0.836	0 474	0 979
Air-Flow	0 836	1	0 418	0 869
Temp	0 474	0.418	1	0 423
Acid-Conc	0.979	0 869	0 423	1

Eigenvalues of Final Weighted Correlation Matrix

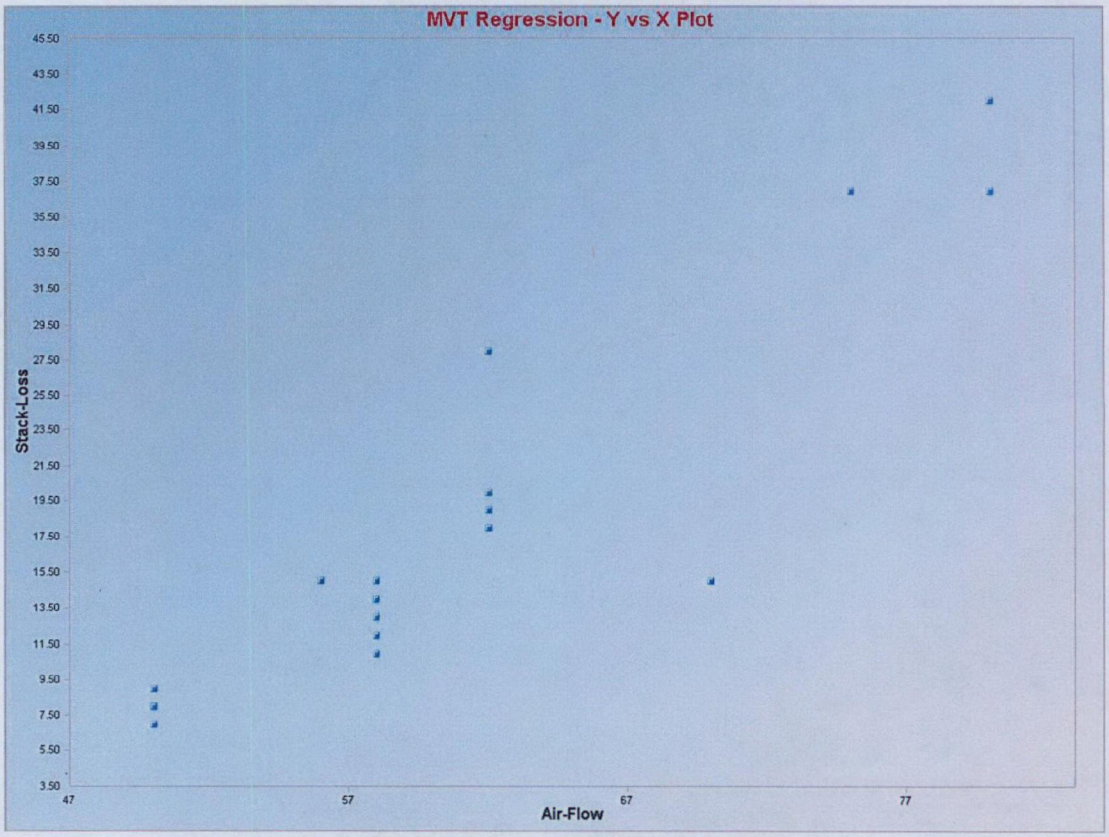
Eval 1	Eval 2	Eval 3	Eval 4
0 0169	0 19	0.724	3.069

Output for MVT (Leverage ON) (continued).





Output for MVT (Leverage ON) (continued).



Interpretation of Graphs: Observations which are outside of the horizontal lines in the graphs are considered to be regression outliers. The observations to the right of the vertical lines are considered to be leverage outliers. The regression lines are not produced since there are three predictor variables. Select other "**X**" variables by using the drop-down bar in the graphics panel and click on "**Redraw**."

Output for MVT (Leverage OFF).

Regression Analysis Output	
Date/Time of Computation	1/10/2008 8 47 07 AM
User Selected Options	
From File	D \Narain\Scout_For_Windows\ScoutSource\WorkDatInExcel\STACKLOSS
Full Precision	OFF
Selected Regression Method	Multivariate Triming (MVT)
MVT Trim Percentage	0.05 (Used to Identify Regression Outliers)
Number of Regression Iterations	10 (Maximum Number if doesn't Converge)
Leverage	Off
Res-Lev Rectangle Alpha	0 05 (Used with Graphics Confidence Bands)
Title for Residual QQ Plot	MVT Regression - Residuals QQ Plot
Title Residual vs Distance Plot	MVT Regression - Residuals vs Unsquared Distance Plot
Title For Y vs X Plots)	MVT Regression - Y vs X Plot
Residual QQ Plot	Not Selected
Y vs Residual Plot	Not Selected
Y-hat vs Residual Plot	Not Selected
Show Intermediate Results	Do Not Display Intermediate Results
	Intermediate Results Shown on Another Output Sheet

Number of Selected Regression Variables	3
Number of Observations	21
Dependant Variable	Stack-Loss

Ordinary Least Squares (OLS) Regression Results

Regression Parameters Vector Estimates

Intercept	Air-Flow	Temp	Acid-Conc
-39 92	0 716	1.295	-0.152

Stdv of Regression Estimates Vector

Intercept	Air-Flow	Temp.	Acid-Conc
11 9	0 135	0 368	0 156

ANOVA Table

Source of Variation	SS	DOF	MS	F-Value	P-Value
Regression	1890	3	630 1	59.9	0 0000
Error	178 8	17	10 52		
Total	2069	20			

Output for MVT (Leverage OFF) (continued).

R Square Estimates	0.914
MAD Based Scale Estimates	2.768
Weighted Scale Estimates	3.243
IQR Estimates	4 313
Det. of COV[Regression Coefficients] Matrix	1.0370E-5

Results From the Regression Operation

Regression Parameters Vector Estimates

Intercept	Air-Flow	Temp.	Acid-Conc
-43 7	0.889	0.817	-0.107

Stdv of Regression Estimates Vector

Intercept	Air-Flow	Temp.	Acid-Conc
9.492	0 119	0.325	0.125

ANOVA Table

Source of Variation	SS	DOF	MS	F-Value	P-Value
Regression	1957	3	652 3	98.82	0.0000
Error	105.6	16	6.601		
Total	2063	19			

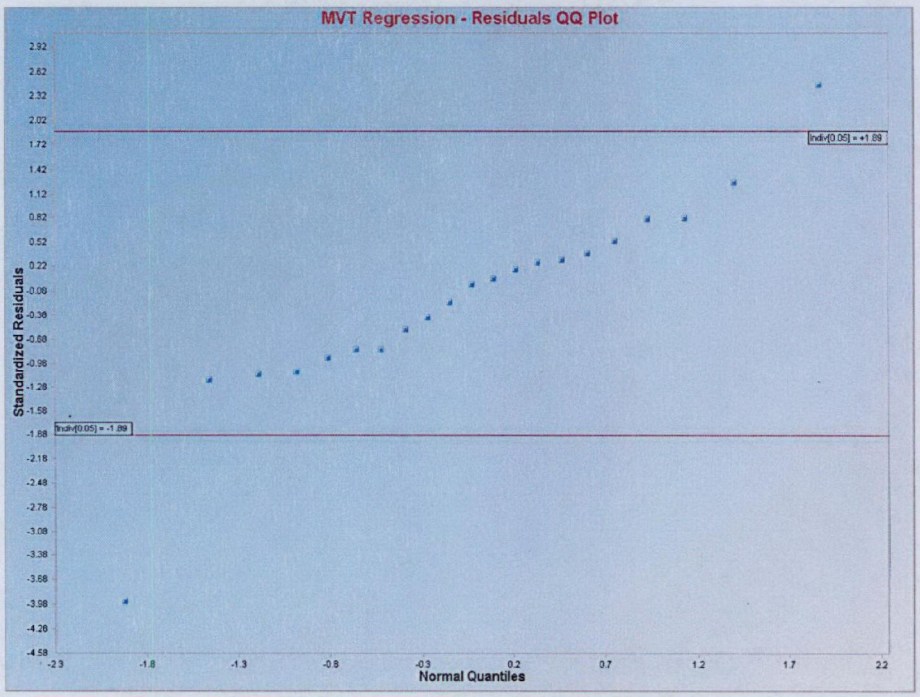
R Square Estimates	0.949
MAD Based Scale Estimates	3.046
Weighted Scale Estimates	2.569
IQR Estimates	3.365
Det of COV[Regression Coefficients] Matrix	2.2471E-6

Regression Table

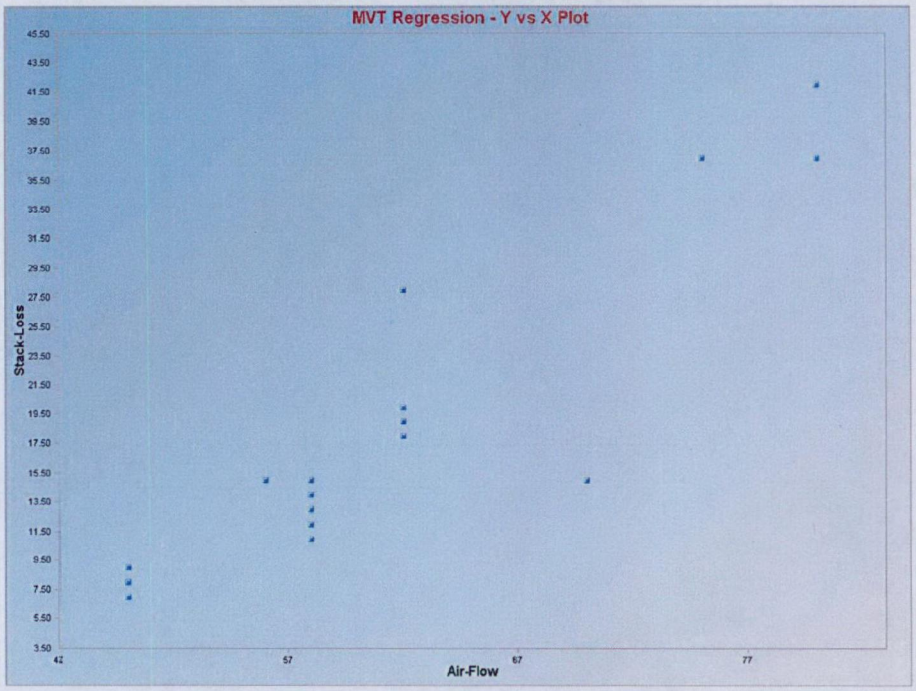
Obs	Y Vector	Yhat	Residuals	Hat[i,i]	Res/Scale	Stude~ Res	Wts[i,i]	Res Dist.
1	42	39.94	2.062	0.302	0.803	0 96	1	0.803
2	37	40.04	-3.045	0 318	-1 185	-1.435	1	1.185
3	37	33 75	3 248	0 175	1 264	1 392	1	1 264
4	28	21.7	6.302	0.129	2.453	2 627	1	2.453
5	18	20.07	-2.065	0.0522	-0 804	-0.826	1	0 804
6	18	20.88	-2 882	0.0775	-1 122	-1.168	1	1.122
7	19	21 06	-2 055	0.219	-0.8	-0.905	1	0.8
8	20	21.06	-1 055	0.219	-0 411	-0 465	1	0.411
9	15	17.33	-2.325	0.14	-0.905	-0.976	1	0.905

(The complete regression table is not shown.)

Output for MVT (Leverage OFF) (continued).



Output for MVT (Leverage OFF) (continued).

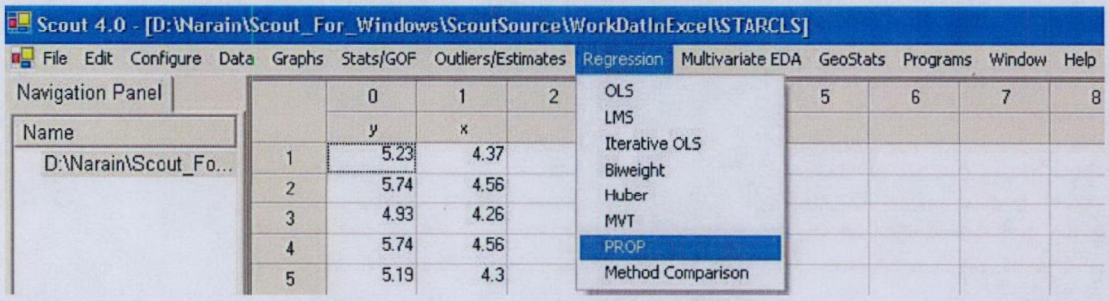


Interpretation of Graphs: Observations which are outside of the horizontal lines in the graph are considered to be regression outliers. The Leverage Distances vs. Standardized residuals plot is not produced even if checked on. Regression lines are not produced since there are three predictor variables. Select other "**X**" variables by using the drop-down bar in the graphics panel and click on "**Redraw**."

Note: There are at least four regression outliers (1, 3, 4, and 21) in the data set of size 21on the previous page. However, the trimming percentage selected is only 5%, which is equivalent to one outlier in the data set of size 21. The user may want to use the MVT method with a higher trimming percentage to identify all of the outliers.

9.8 PROP Regression Method

1. Click **Regression ▶ PROP.**

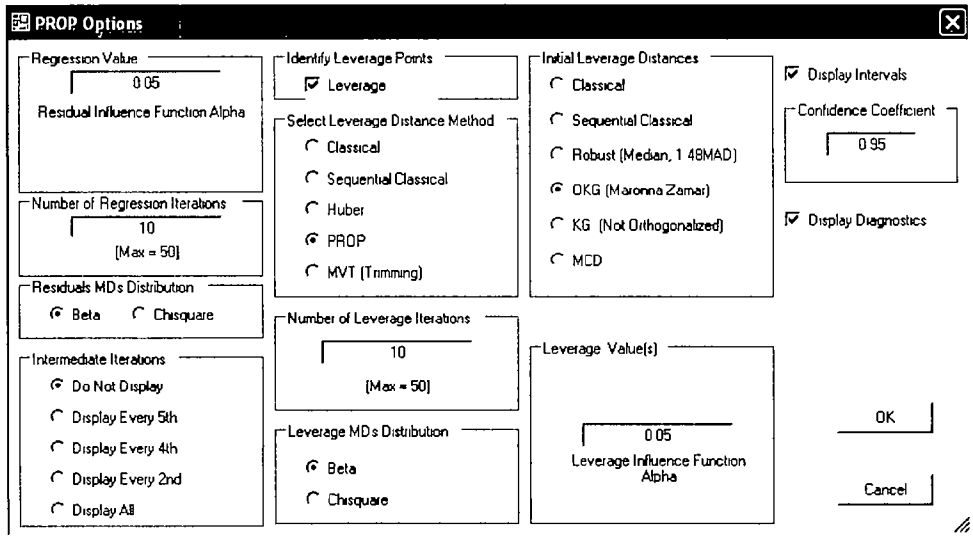


2. The "**Select Variables**" screen (Section 3.3) will appear.

- Select the dependent variable and one or more independent variables from the "**Select Variables**" screen.

- If the results have to be produced by using a Group variable, then select a group variable by clicking the arrow below the "**Group by Variable**" button. This will result in a drop-down list of available variables. The user should select and click on an appropriate variable representing a group variable.

- Click on the "**Options**" button to get the options window.



- Specify the "**Regression Value.**" The default is "**0.05.**"

- Specify the "**Number of Regression Iterations.**" The default is"**10.**"

- Specify the "**Regression MDs Distribution.**" The default is "**Beta.**"

- Specify the "**Identify Leverage Points.**" The default is "**On.**"

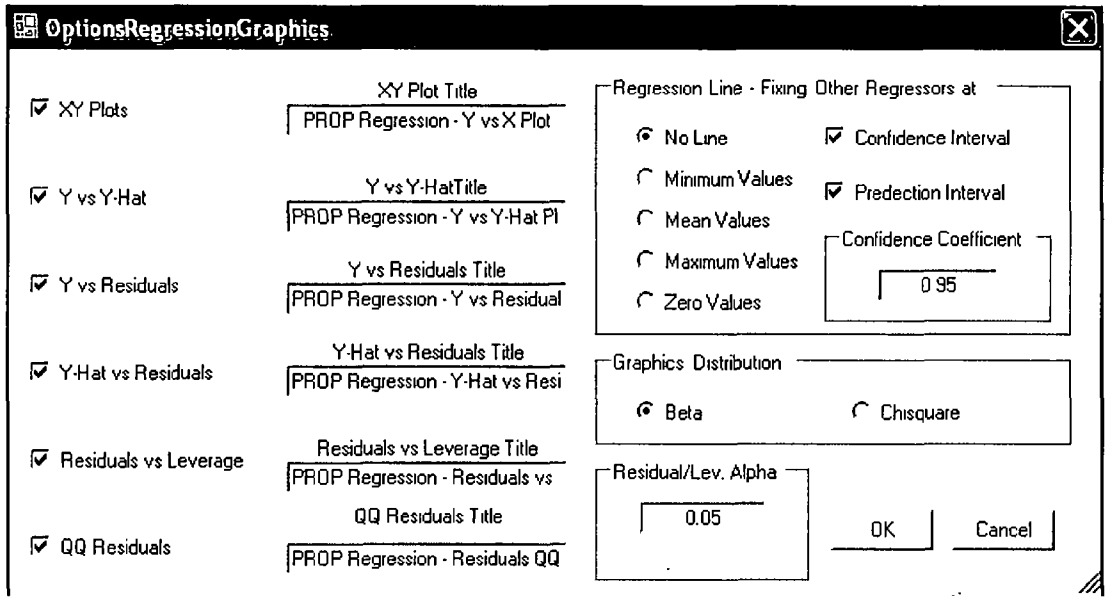
- Specify the "**Select Leverage Distance Method.**" The default is "**PROP.**"

- Specify the "**Number of Leverage Iterations.**" The default is"**10.**"

- Specify the "**Leverage Initial Distances.**" The default is "**OKG (Maronna Zamar).**"

- Specify the "**Leverage Value.**" The default is "**0.05.**"
- Click "**OK**" to continue or "**Cancel**" to cancel the options.

- Click on the "**Graphics**" button to get the options window.



- Specify the preferred plots and the input parameters.
- Click "**OK**" to continue or "**Cancel**" to cancel the options.

- Click "**OK**" to continue or "**Cancel**" to cancel the computations.

Output example: The data set "**STARCLS.xls**" was used for PROP regression. It has 1 predictor variables (p) and 47 observations. When the "**Leverage**" option is on, the leverage distances are calculated and outlying observations are obtained iteratively using initial estimates as median and OKG matrix and the leverage option as PROP (i.e., using PROP influence function). Then the weights are assigned to observations and those weights are used in the finding the regression outliers iteratively. When the leverage option is off, all observations are assigned one (1) as weights and then the regression outliers are found using the PROP function iteratively. Finally the estimated regression parameters are calculated.

Output for PROP (Leverage ON).
Data Set Used: Star Cluster (predictor variables p = 1).

Regression Analysis Output	
Date/Time of Computation	3/12/2008 8 09 44 AM
User Selected Options	
From File	D \Narain\Scout_For_Windows\ScoutSource\WorkDatInExcel\STARCLS
Full Precision	OFF
Selected Regression Method	PROP
Residual Influence Function Alpha	0 05 (Used to Identify Vertical Regression Outliers)
Number of Regression Iterations	10 (Maximum Number if doesn't Converge)
Leverage	Identify Leverage Points (Outliers in X-Space)
Selected Leverage Method	PROP
Initial Leverage Distance Method	OKG (Maronna Zamar) Matrix
Squared MDs	Beta Distribution used for Leverage Distances based upon Selected Regression (Leverage) Variables
Leverage Distance Alpha	0 05 (Used to Identify Leverage Points)
Number of Leverage Iterations	10 (Maximum Number if doesn't Converge)
Y vs Y-hat Plot	Not Selected
Y vs Residual Plot	Not Selected
Y-hat vs Residual Plot	Not Selected
Title For Y vs X Plots	PROP Regression - Y vs X Plot
Title for Residual QQ Plot	PROP Regression - Residuals QQ Plot
Residual Band Alpha	0 05 (Used in Graphics Residual Bands)
Title Residual vs Distance Plot	PROP Regression - Residuals vs Unsquared Leverage Distance Plot
Show Intermediate Results	Do Not Display Intermediate Results
	Intermediate Results Shown on Another Output Sheet

Leverage Points are Outliers in X-Space of Selected Regression Variables.

Number of Selected Regression Variables	1
Number of Observations	47
Dependent Variable	*y*

Residual Values used with Graphics Display

Upper Residual Individual (0 05) MD	1 929
Lower Residual Individual (0 05) MD	-1 929

Correlation Matrix

	y	x
y	1	-0 21
x	0 21	1

Output for PROP (Leverage ON) (continued).

Eigenvalues of Correlation Matrix	
Eval 1	Eval 2
0.79	1.21

Ordinary Least Squares (OLS) Regression Results

Estimates of Regression Parameters	
Intercept	x
6.793	-0.413

Stdv of Estimated Regression Parameters	
Intercept	x
1.237	0.286

ANOVA Table

Source of Variation	SS	DOF	MS	F-Value	P-Value
Regression	0.665	1	0.665	2.085	0.1557
Error	14.35	45	0.319		
Total	15.01	46			

R Square Estimate	0 0443
MAD Based Scale Estimate	0 651
Weighted Scale Estimate	0.565
IQR Estimate of Residuals	1.025
Det. of COV[Regression Coefficients] Matrix	5.5584E-4

Initial Weighted Regression Iteration with Identified Leverage Points

Estimates of Regression Parameters	
Intercept	x
-7.97	2.93

Stdv of Estimated Regression Parameters	
Intercept	x
2.396	0.543

(The complete regression table is not shown.)

Output for PROP (Leverage ON) (continued).

ANOVA Table

Source of Variation	SS	DOF	MS	F-Value	P-Value
Regression	4 205	1	4 205	29 09	0 0000
Error	5 647	39.06	0.145		
Total	9.852	40 06			

R Square Estimate	0.427
MAD Based Scale Estimate	0 45
Weighted Scale Estimate	0 38
Unsquared Leverage Distance Indiv-MD(0 05)	1 929
IQR Estimate of Residuals	0 606
Determinant of Leverage S Matrix	0 0118

Regression Table with Leverage Option

Obs	Y Vector	Yhat	Residuals	Hat[i,i]	Res/Scale	Stude~ Res	Wts[i,i]	Res Dist	Lev Dist.	OLS R~ist
1	5 23	4 836	0 394	0 0222	1.037	1.049	1	1 037	0 348	0 43
2	5.74	5 393	0 347	0 0373	0.914	0.931	1	0 914	1 405	1 472
3	4 93	4 513	0 417	0 0219	1 096	1.108	1	1 096	1 362	0 182
4	5 74	5.393	0 347	0.0373	0.914	0 931	1	0 914	1 405	1.472
5	5.19	4 631	0.559	0 0213	1.471	1 487	1	1 471	0 993	0 308
6	5.46	5 1	0 36	0 0271	0 948	0 961	1	0 948	0 483	0 903
7	4.65	3 283	1 367	0 0781	3.596	3.745	0 0135	3 596	5 237	0 985
8	5 27	5 422	-0 152	0 0387	-0.399	-0.407	1	0 399	1.497	0.647
9	5 57	4 513	1 057	0.0219	2.779	2 81	1	2 779	1.362	0.951
10	5.12	4 836	0 284	0 0222	0.748	0 756	1	0 748	0 348	0.235
11	5.73	2 257	3.473	0 194	9.134	10.17	3 3033E-4	9 134	8 465	0.671
12	5.45	5 012	0 438	0 025	1 153	1 168	1	1 153	0 206	0 863
13	5.42	5 158	0 262	0.0287	0 689	0 699	1	0 689	0.667	0 847
14	4 05	3 781	0 269	0 0444	0 708	0.724	0 0923	0 708	3.668	1 924
15	4 26	4.601	-0 341	0 0214	-0 898	-0 908	1	0.898	1 086	1 347
16	4 58	4.982	-0.402	0 0244	-1 058	-1 071	1	1.058	0 114	0 685
17	3 94	4 426	-0.486	0 0229	-1 277	-1 292	1	1 277	1 639	1.957
18	4.18	4 982	-0 802	0 0244	-2 11	-2 136	1	2 11	0 114	1 393
19	4.18	4.426	-0 246	0 0229	-0 646	-0 653	1	0 646	1 639	1.532
20	5.89	2.257	3.633	0.194	9.555	10 64	3.3033E-4	9.555	8 465	0 955
21	4 38	4.601	-0 221	0 0214	-0 582	-0 589	1	0 582	1 086	1 134
22	4 22	4.601	-0 381	0 0214	-1 003	-1 014	1	1 003	1 086	1.418
23	4 42	4.982	-0.562	0 0244	-1 479	-1.497	1	1 479	0 114	0.968

(The complete regression table is not shown.)

Output for PROP (Leverage ON) (continued).

Final Reweighted Regression Results

Estimates of Regression Parameters

Intercept	x
-7 955	2 926

Stdv of Estimated Regression Parameters

Intercept	x
1 911	0.434

ANOVA Table

Source of Variation	SS	DOF	MS	F-Value	P-Value
Regression	5.401	1	5 401	45 44	0 0000
Error	4 614	38.83	0 119		
Total	10 01	39.83			

R Square Estimate	0 539
MAD Based Scale Estimate	0 45
Weighted Scale Estimate	0.345
IQR Estimate of Residuals	0 607
Det. of COV[Regression Coefficients] Matrix	5 4829E-4

Regression Table

Obs	Y Vector	Yhat	Residuals	Hat[i,i]	Res/Scale	Stude~ Res	Wts[i,i]	Res Dist
1	5 23	4 831	0.399	0 0222	1 157	1 17	1	1.157
2	5 74	5.387	0.353	0 0373	1.023	1 043	1	1 023
3	4.93	4.509	0 421	0.0219	1 22	1 234	1	1 22
4	5 74	5 387	0 353	0.0373	1 023	1 043	1	1 023
5	5 19	4.626	0 564	0 0213	1 635	1 652	1	1 635
6	5 46	5.095	0.365	0 0271	1.06	1.075	1	1.06
7	4 65	3.281	1 369	0 0781	3 972	4.137	0 0628	3 972
8	5 27	5.416	-0.146	0 0387	-0 425	-0.433	1	0 425

(The complete regression table is not shown.)

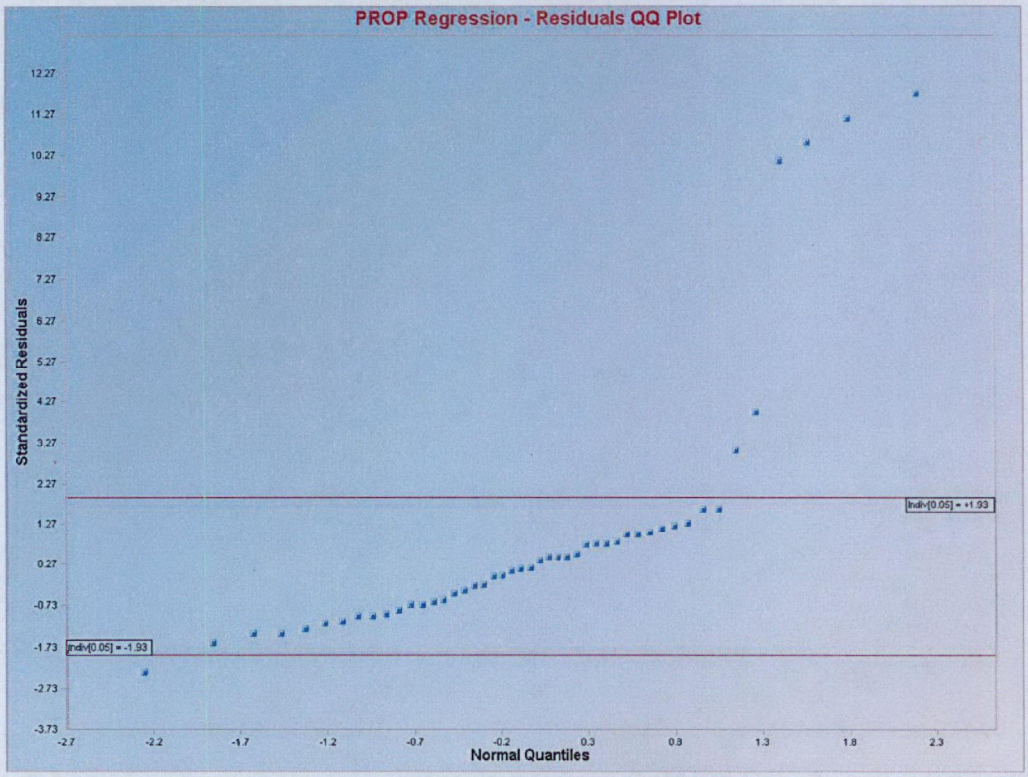
Final Weighted Correlation Matrix

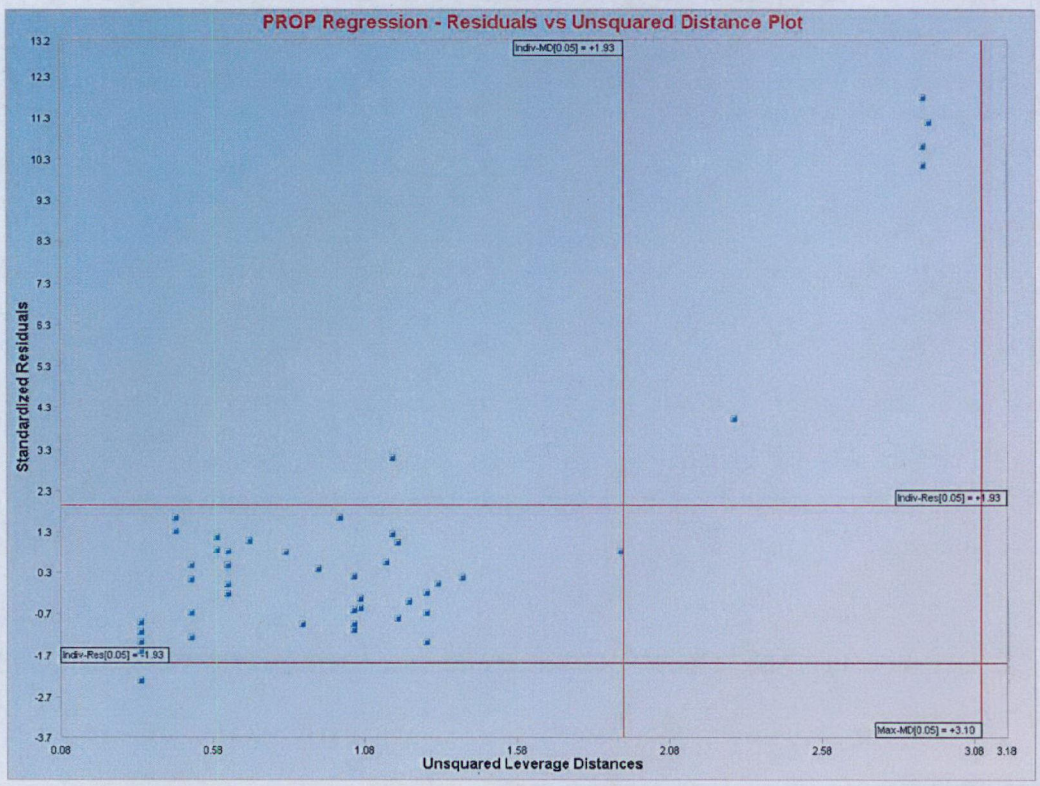
	y	x
y	1	0 759
x	0 759	1

Eigenvalues of Final Weighted Correlation Matrix

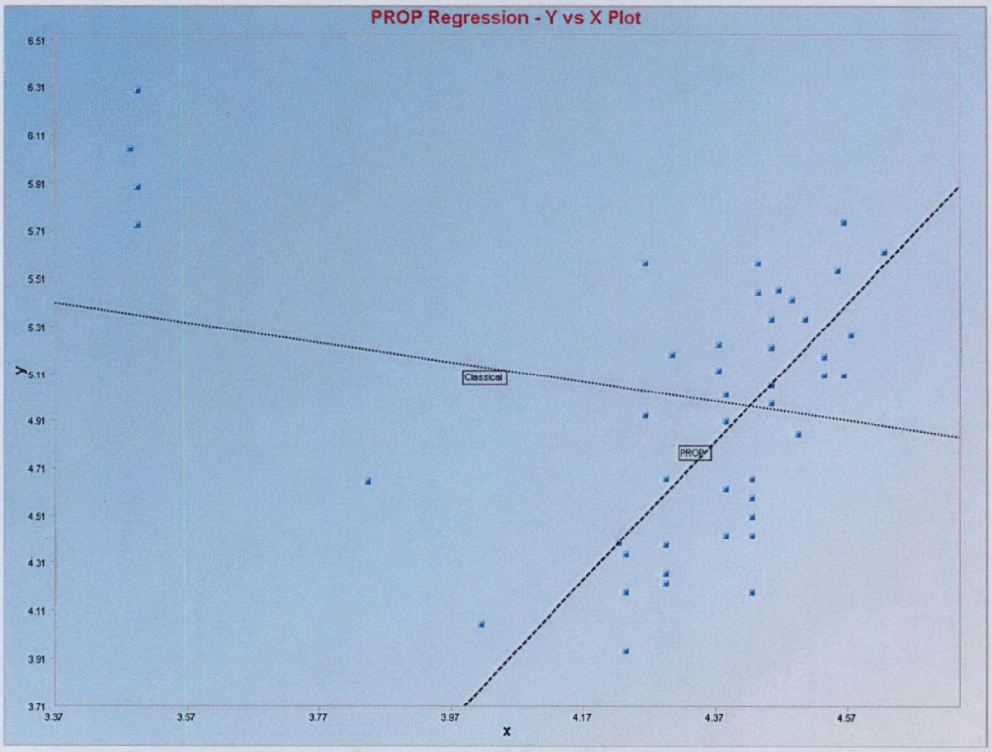
Eval 1	Eval 2
0 241	1 759

Output for PROP (Leverage ON) (continued).





Output for PROP (Leverage ON) (continued).



Interpretation of Graphs: Observations which are outside of the horizontal lines on the residual Q-Q plot or on the residual versus unsquared leverage distances represent regression outliers. Observations lying to the right of the vertical lines represent leverage outliers, leverage points lying between the two horizontal lines represent good leverage points, and the rest of the leverage points represent bad leverage points. Both the classical and robust regression lines are also shown on the y vs. x scatter plot.

Output for PROP (Leverage OFF).

In order to demonstrate the usefulness of the leverage options (when several leverage points may be present), the Star cluster data is considered again with the leverage option off.

The output thus obtained is given as follows.

Regression Analysis Output	
Date/Time of Computation	3/12/2008 8 15 48 AM
User Selected Options	
From File	D \Narain\Scout_For_Windows\ScoutSource\WorkDatInExcel\STARCLS
Full Precision	OFF
Selected Regression Method	PROP
Residual Influence Function Alpha	0.05 (Used to Identify Vertical Regression Outliers)
Number of Regression Iterations	10 (Maximum Number if doesn't Converge)
Leverage	Off
Y vs Y-hat Plot	Not Selected
Y vs Residual Plot	Not Selected
Y-hat vs Residual Plot	Not Selected
Title For Y vs X Plots	PROP Regression - Y vs X Plot
Title for Residual QQ Plot	PROP Regression - Residuals QQ Plot
Residual Band Alpha	0 05 (Used in Graphics Residual Bands)
Title Residual vs Distance Plot	PROP Regression - Residuals vs Unsquared Leverage Distance Plot
Show Intermediate Results	Do Not Display Intermediate Results
	Intermediate Results Shown on Another Output Sheet

Number of Selected Regression Variables	1
Number of Observations	47
Dependent Variable	y

Residual Values used with Graphics Display	
Upper Residual Indvidual (0 05) MD	1 929
Lower Residual Indvidual (0.05) MD	-1.929

Correlation Matrix

	y	x
y	1	-0 21
x	-0 21	1

Eigenvalues of Correlation Matrix

Eval 1	Eval 2
0 79	1 21

Output for PROP (Leverage OFF) (continued).

Ordinary Least Squares (OLS) Regression Results

Estimates of Regression Parameters

Intercept	x
6.793	-0.413

Stdv of Estimated Regression Parameters

Intercept	x
1.237	0.286

ANOVA Table

Source of Variation	SS	DOF	MS	F-Value	P-Value
Regression	0.665	1	0.665	2.085	0.1557
Error	14.35	45	0.319		
Total	15.01	46			

R Square Estimate	0.0443
MAD Based Scale Estimate	0.651
Weighted Scale Estimate	0.565
IQR Estimate of Residuals	1.025
Det. of COV[Regression Coefficients] Matrix	5.5584E-4

Final Reweighted Regression Results

Estimates of Regression Parameters

Intercept	x
6.799	-0.414

Stdv of Estimated Regression Parameters

Intercept	x
1.235	0.286

ANOVA Table

Source of Variation	SS	DOF	MS	F-Value	P-Value
Regression	0.668	1	0.668	2.102	0.1540
Error	14.29	44.95	0.318		
Total	14.95	45.95			

Output for PROP (Leverage OFF) (continued).

R Square Estimate	0.0447
MAD Based Scale Estimate	0 651
Weighted Scale Estimate	0 564
IQR Estimate of Residuals	1.025
Det of COV[Regression Coefficients] Matrix	5 5303E-4

Regression Table

Obs	Y Vector	Yhat	Residuals	Hat[i,i]	Res/Scale	Stude~ Res	Wts[i,i]	Res Dist.
1	5 23	4 988	0.242	0.0222	0.429	0 433	1	0.429
2	5.74	4 91	0.83	0 0373	1.473	1.501	1	1.473
3	4 93	5 034	-0 104	0 0219	-0.184	-0 187	1	0 184
4	5 74	4 91	0 83	0 0373	1 473	1 501	1	1 473
5	5.19	5 017	0.173	0 0213	0 306	0 309	1	0 306
6	5 46	4 951	0.509	0 0271	0 903	0 915	1	0.903
7	4.65	5 208	-0 558	0 0781	-0 99	-1.031	1	0 99
8	5 27	4 906	0 364	0 0387	0 646	0.659	1	0 646
9	5.57	5.034	0 536	0 0219	0 951	0 961	1	0.951
10	5.12	4.988	0.132	0.0222	0 233	0.236	1	0.233
11	5 73	5 353	0.377	0.194	0.669	0 745	1	0 669
12	5.45	4 964	0.486	0 025	0.863	0 874	1	0.863
13	5.42	4 943	0.477	0 0287	0 846	0 859	1	0 846
14	4 05	5 138	-1 088	0 0444	-1.929	-1 974	1	1 929
15	4 26	5 022	-0.762	0 0214	-1 351	-1 366	1	1.351
16	4.58	4.968	-0.388	0 0244	-0 688	-0.696	1	0.688
17	3.94	5.046	-1.106	0.0229	-1.963	-1.985	0 951	1.963
18	4.18	4.968	-0.788	0.0244	-1.397	-1.415	1	1 397
19	4.18	5.046	-0 866	0 0229	-1.537	-1.555	1	1.537
20	5.89	5.353	0 537	0 194	0.952	1.061	1	0 952
21	4.38	5.022	-0.642	0.0214	-1.138	-1.15	1	1.138

(The complete regression table is not shown.)

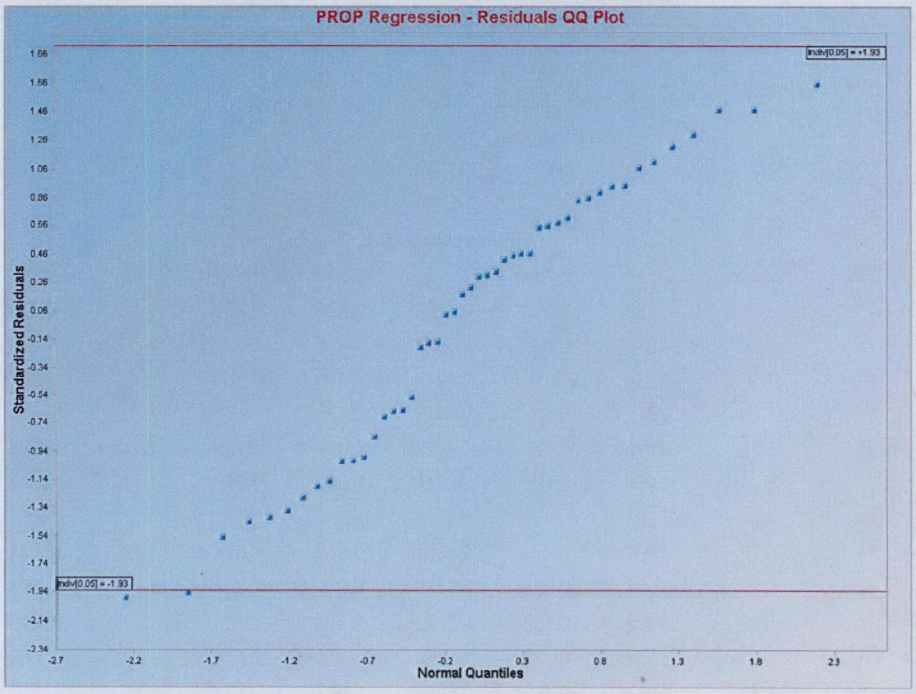
Final Weighted Correlation Matrix

	y	x
y	1	-0 212
x	-0 212	1

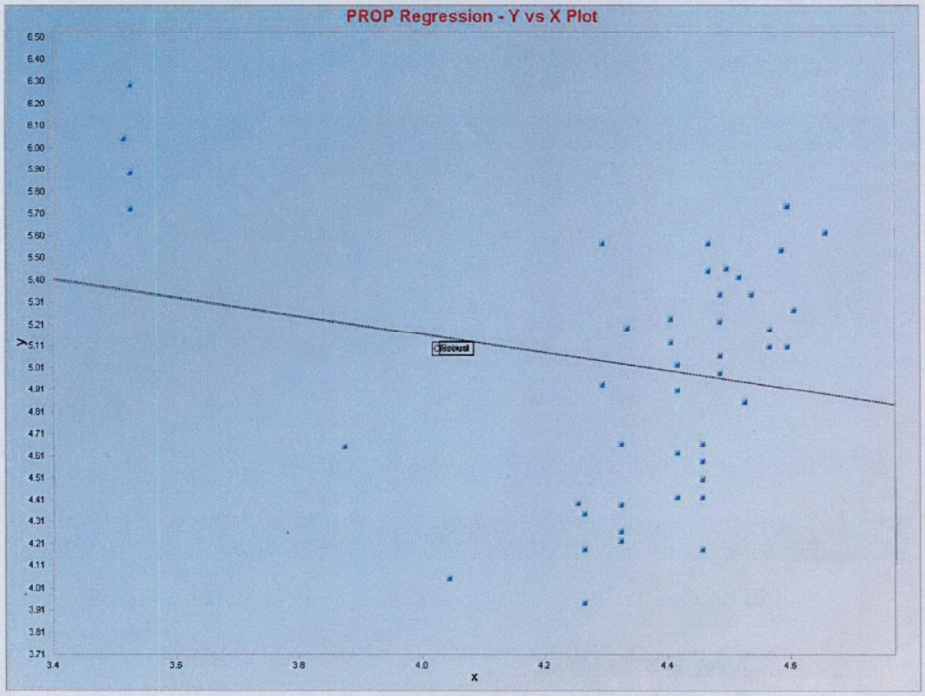
Eigenvalues of Final Weighted Correlation Matrix

Eval 1	Eval 2
0.788	1 212

Q-Q plot of Standardized Output for PROP (Leverage OFF).



Q-Q plot of Standardized Output for PROP (Leverage OFF).



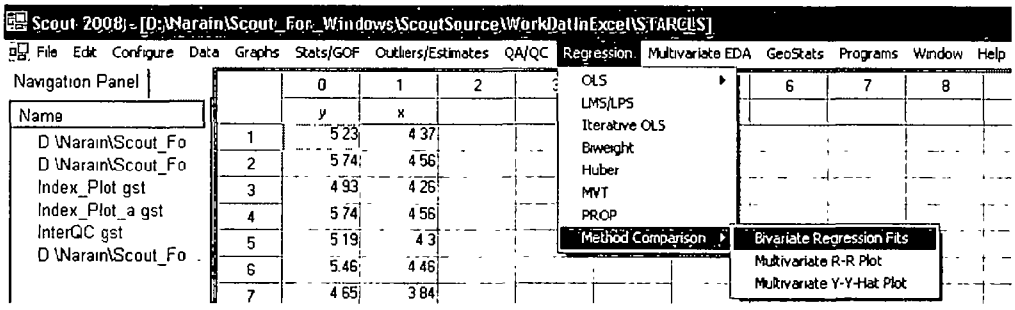
Interpretation of Graphs: Observations (if any) lying outside of the horizontal lines in the Q-Q plot are considered to be regression outliers. The Leverage Distances vs. standardized residuals plot is not produced as the leverage option was not activated. Regression lines are produced since there is only one predictor variable. It is easy to see from the above graph (where both the classical and robust regression lines are overlapping and attracted toward the outliers) that one should use the leverage option to properly identify all of the leverage points. Once the leverage points are identified, the robust regression method should be used to distinguish between the good and bad leverage points.

9.9 Method Comparison in Regression Module

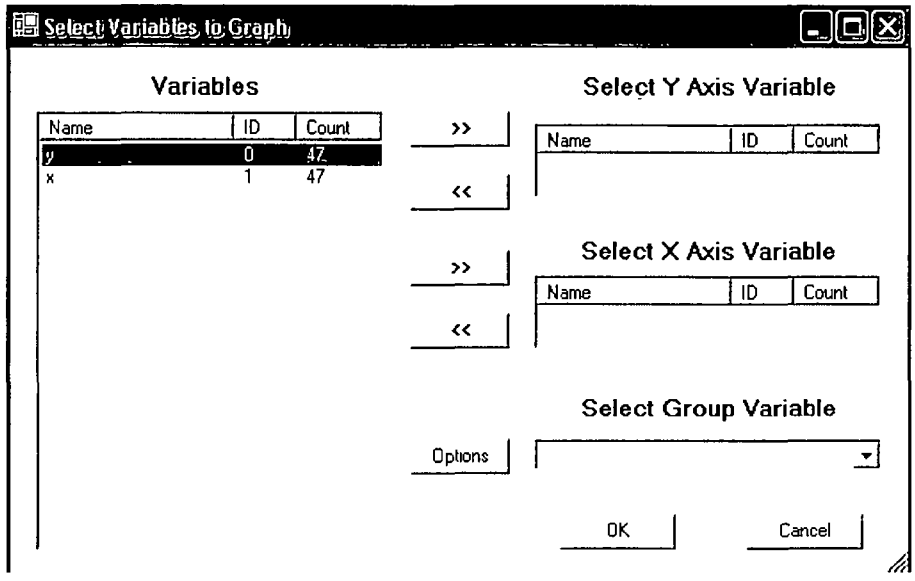
The "**Method Comparison**" option in the "**Regression**" drop-down menu can be used to compare the regression estimates of bivariate data obtained using various classical and robust regression methods. Regression lines for the selected regression methods are drawn on two-dimensional scatter plots. These comparisons are done in the "**Bivariate Regression Fits**" drop-down menu. The method comparison module also compares the residuals obtained by a single regression method against residuals obtained from one or more methods. A comparison of fits (Y-hat) from one method against fits from the other methods is done in a similar way. These comparisons of the residuals and fits from the various regression methods are done in "**R-R Plots**" and "**Y-Y-hat Plots**," respectively.

9.9.1 Bivariate Fits

1. Click **Regression ▶ Method Comparison ▶ Bivariate Regression Fits.**



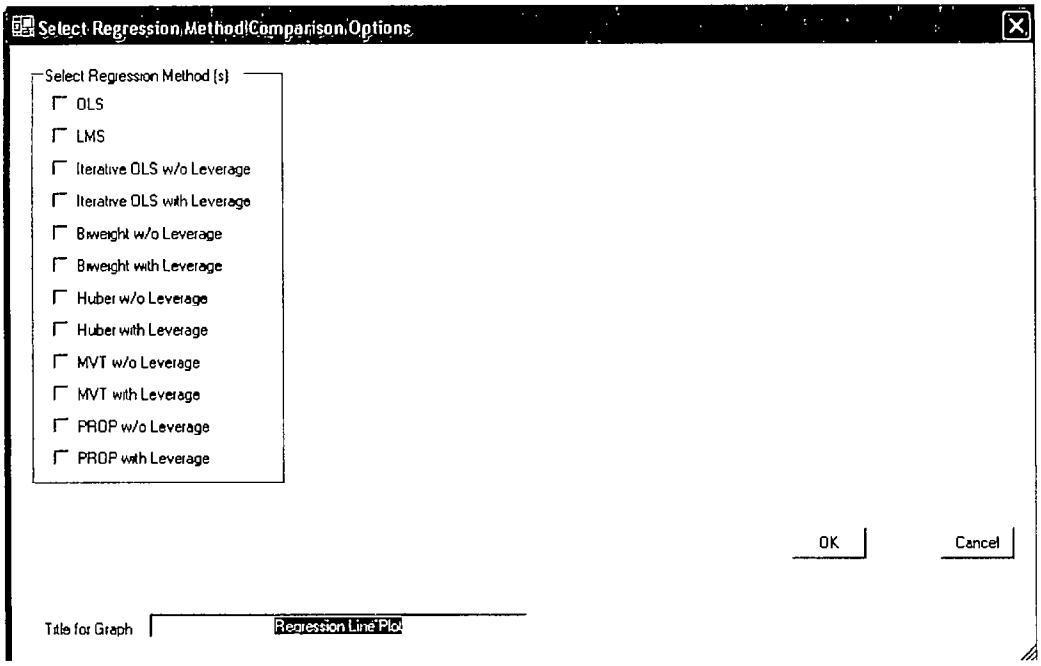
2. The "**Select Variables**" screen (Section 3.3) will appear.



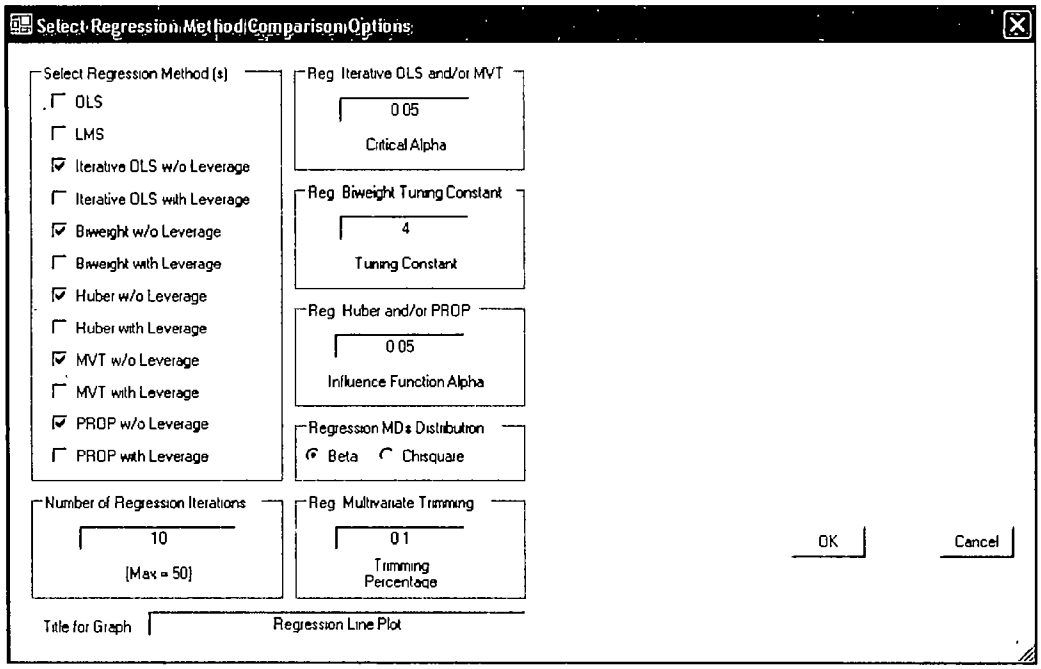
- Select the Y axis variable and the X axis variable from the "**Select Variables to Graph**" screen.

- If the results have to be produced by using a Group variable, then select a group variable by clicking the arrow below the "**Group by Variable**" button. This will result in a drop-down list of available variables. The user should select and click on an appropriate variable representing a group variable.

- Click on "**Options**" for method comparison options.

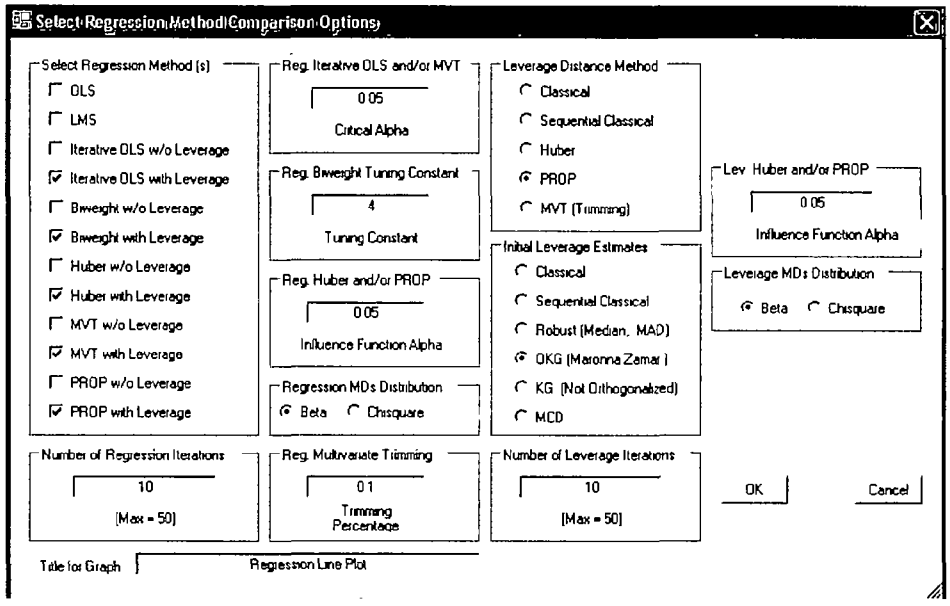


- The options in the window shown above represent the different options.



- The options selected in the window shown above are the options for the regression methods without the leverage option.

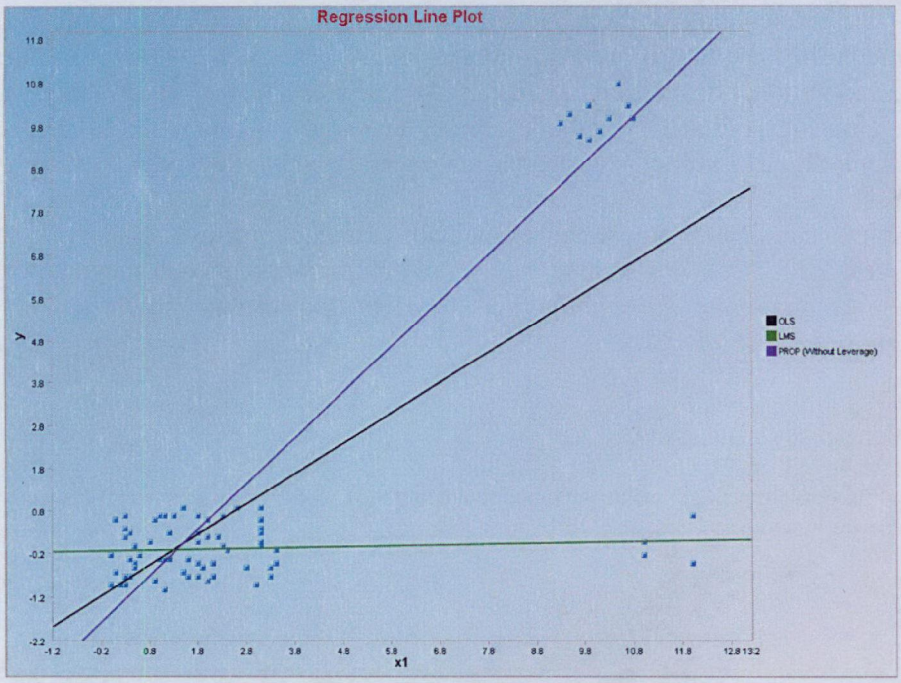
- The "**Iterative OLS w/o Leverage**" requires the input of a "**Critical Alpha.**"
- The "**Biweight w/o Leverage**" requires the input of a "**Tuning Constant.**"
- The "**Huber w/o Leverage**" requires the input of an "**Influence Function Alpha.**"
- The "**MVT w/o Leverage**" requires the input of a "**Critical Alpha**" and a "**Trimming Percentage.**"
- The "**PROP w/o Leverage**" requires the input of an "**Influence Function Alpha.**"



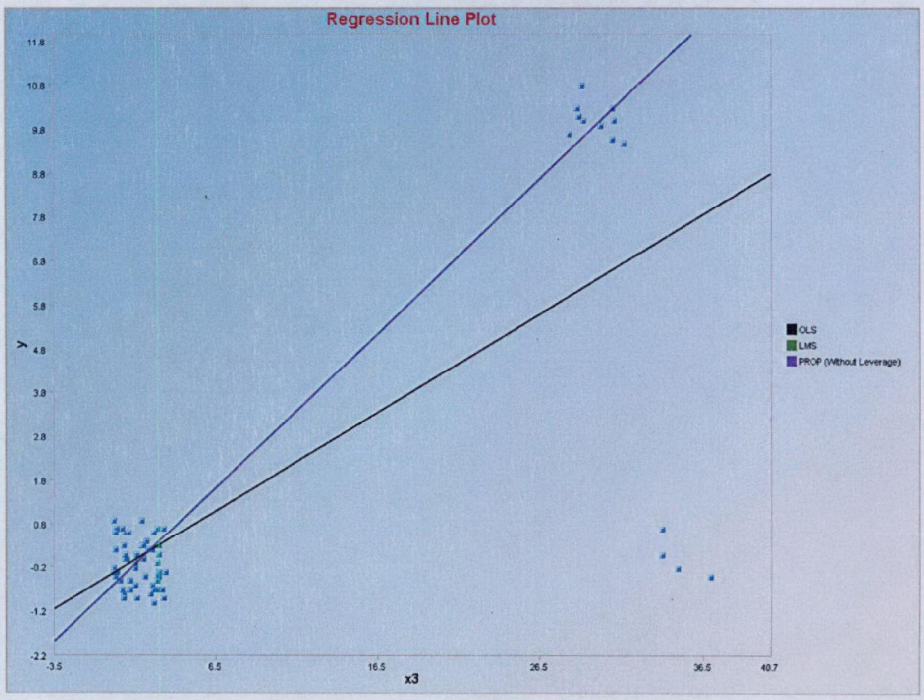
- Options in the window selected above represent options for the regression methods with leverage.

  - The "**Leverage Distance Method**" remains the same for any of the regression methods.
  - The "**Classical**" and "**Sequential Classical**" requires the input of a "**Critical Alpha**"
  - The "**Huber**" and "**PROP**" requires the input of an "**Influence Function Alpha**" and the "**Leverage MDs Distribution.**"
  - The "**MVT**" requires the input of a "**Critical Alpha**" and a "**Trimming Percentage.**"
  - The Leverage Distance Method requires an "**Initial Leverage Estimates**" selection to start the computations.

Graphical Display for Method Comparisons Option.
Data Set: Bradu (X1 vs. Y).
Methods: OLS, LMS, PROP w/o Leverage (PROP influence function alpha for regression outliers = 0.2).

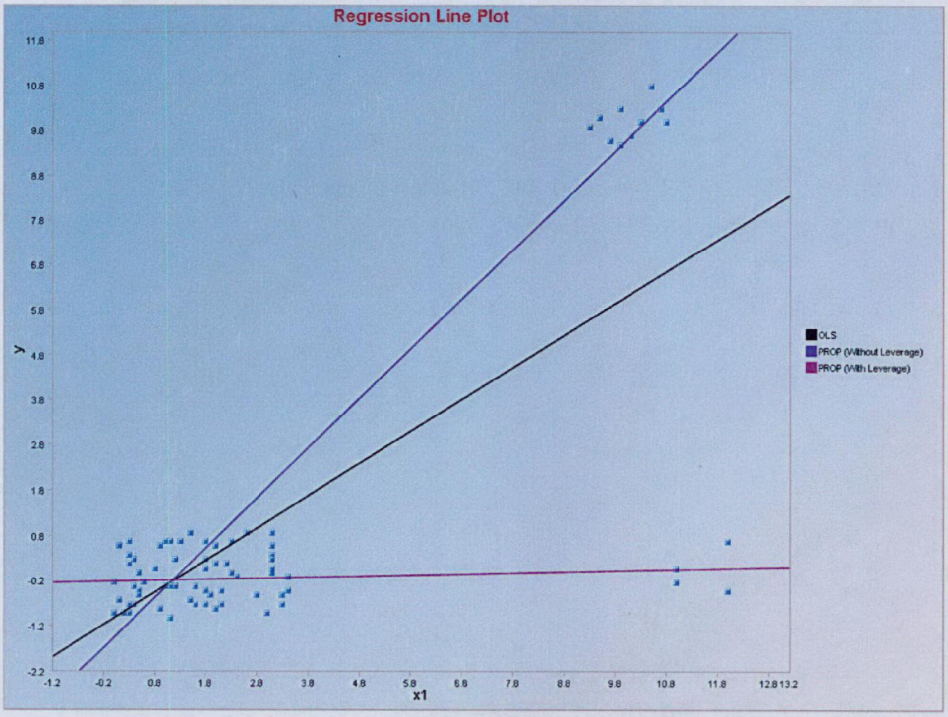


Data Set: Bradu (X3 vs. Y).
Methods: OLS, LMS, PROP w/o Leverage (PROP influence function alpha = 0.2).

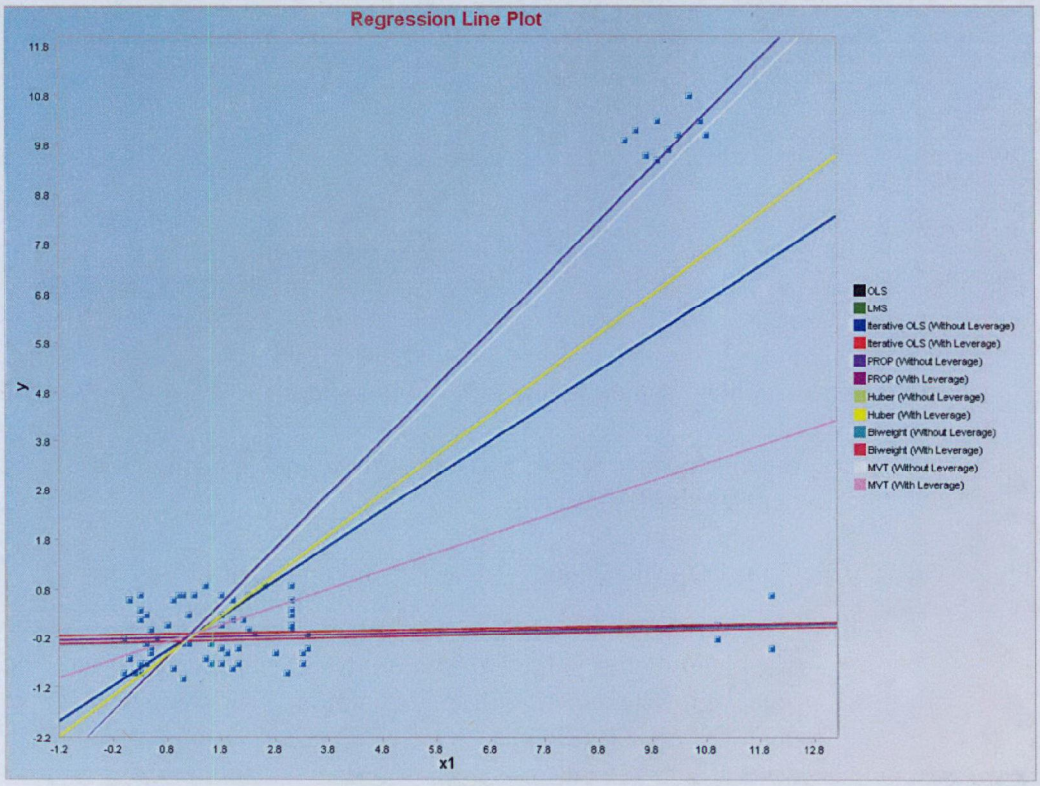


It is noted that the LMS (green line) method finds different sets of outliers when compared to the PROP (violet line) method. As shown earlier, in multiple linear LMS regression of y on x1, x2, and x3, observations 1 through 10 were identified as regression outliers (and bad leverage points). Here the LMS regression of y on x1 (and also of y on x2) also identified the first 10 points as regression outliers; whereas the LMS regression of y on x3 identified observations 11, 12, 13, and 14 as bad leverages and regression outliers. However, the PROP method, without the leverage option, identified observations 11, 12, 13, and 14 as regression outliers and bad leverage points for all of the regression models: y vs. x1, x2, and x3; y vs. x1; y vs. x2; and y vs. x3. In practice, it is desirable to supplement statistical results with graphical displays. In the present context, graphical displays also help the user to determine points that may represent good (or bad) leverage points. *Without the first 10 points, this data set should be used to obtain any regression model.*

Output for Method Comparisons.
Data Set: Bradu (X1 vs. Y).
Methods: PROP w/o Leverage (influence function alpha = 0.2), PROP with Leverage (initial estimate as OKG, influence function as 0.05), and OLS.



Output for Method Comparisons.
Data Set: Bradu (X1 vs. Y).
Methods: All (12 methods).

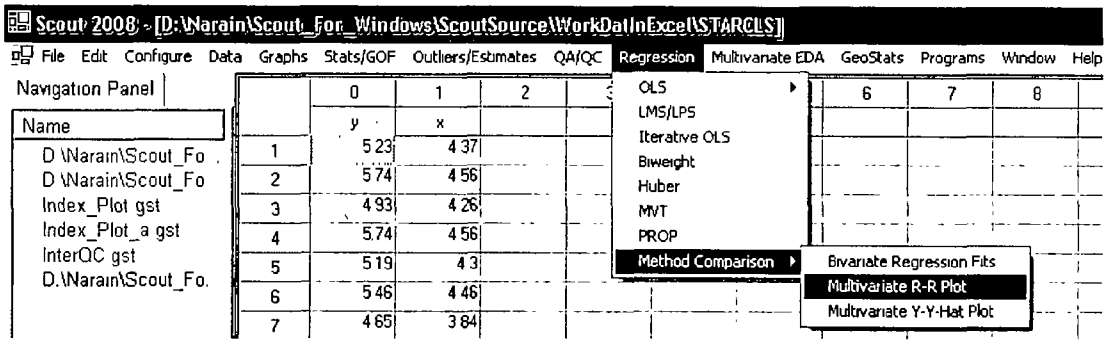


As mentioned before, the user should select the various options carefully. It is suggested not to select all of the available options to generate a single graph. Such a graph will be cluttered with many regression lines. This is illustrated in the above figure.

Note: *Sometimes a line will be outside the frame of the graph. In such cases, a warning message (in orange) will be printed in the Log Panel.*

9.9.2 Multivariate R-R Plots

1. Click **Regression ⊳ Method Comparison ⊳ Multivariate R-R Plot.**

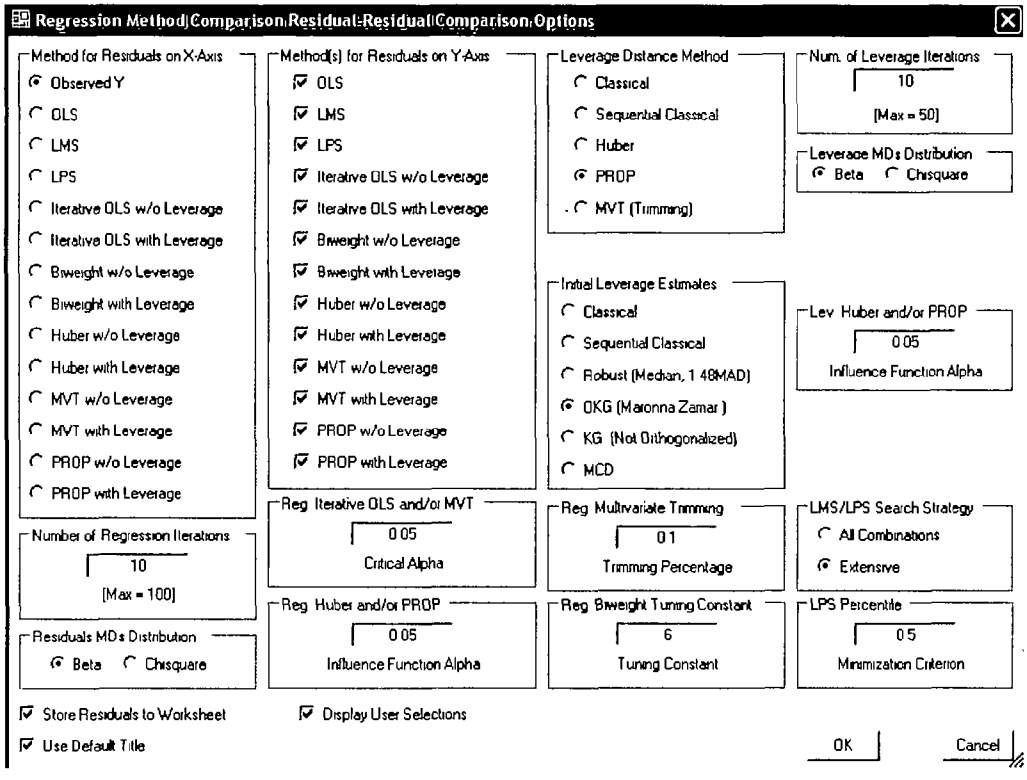


2. The "**Select Variables**" screen (Section 3.3) will appear.

   - Select the dependent variable and one or more independent variables from the "**Select Variables**" screen.

   - Click on the "**Options**" button to get the options window.

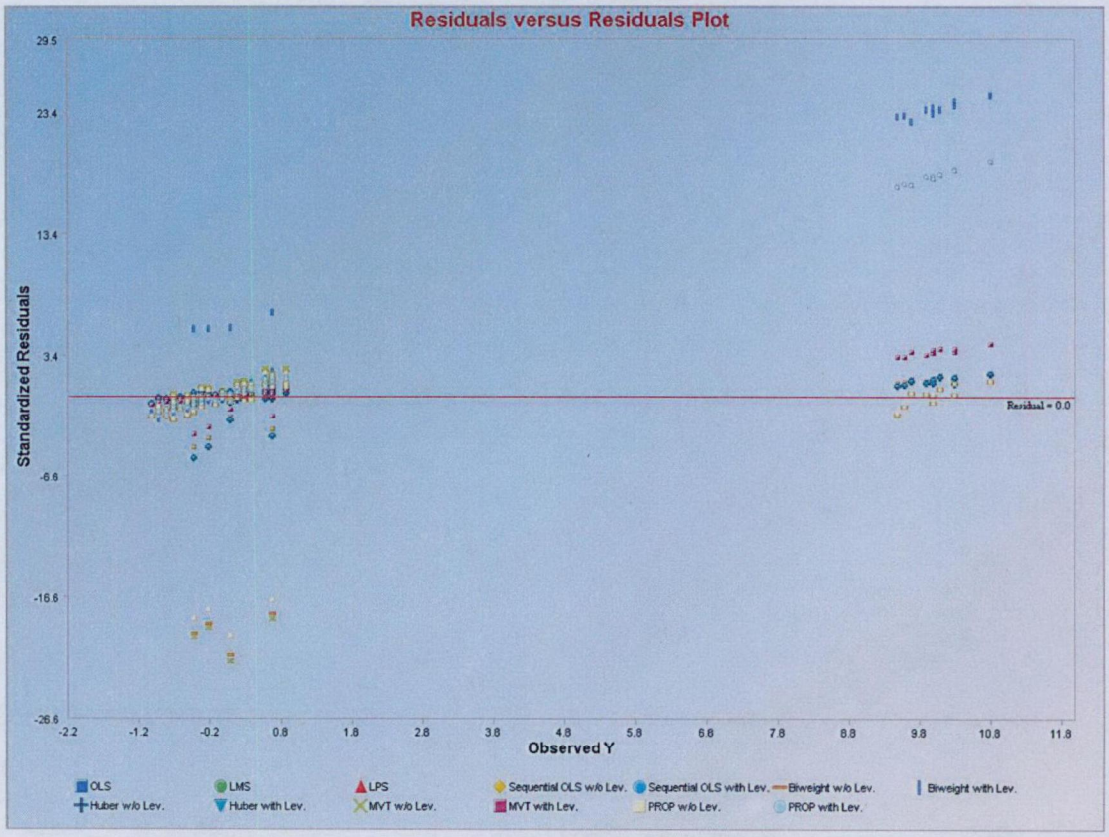
     - The options in the window shown below are the options when all the check-boxes in the "**Method(s) for Residuals on Y-Axis**" are checked. The default option is of plotting the "**Observed Y**" against "**OLS**" residuals.



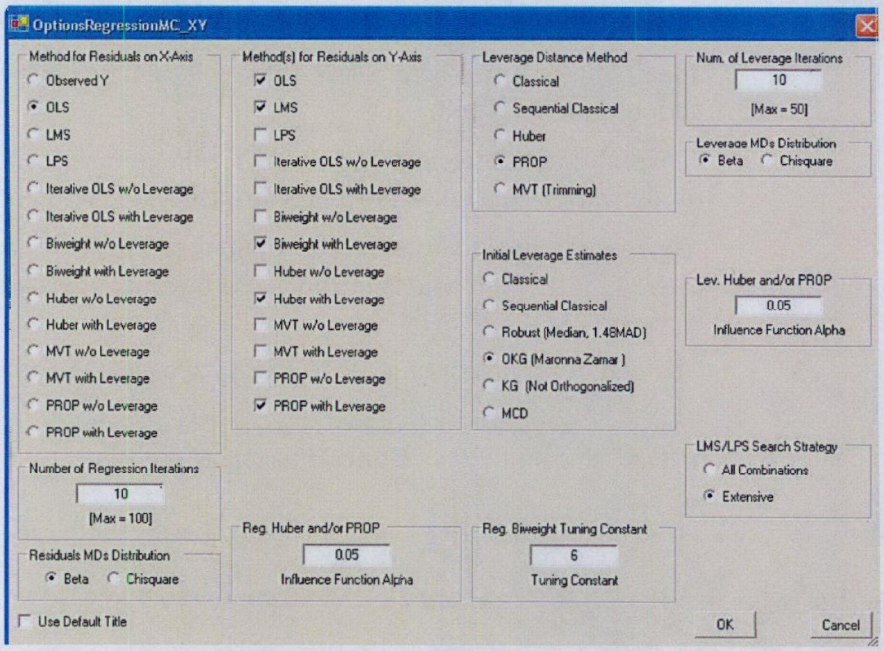
- The options required for the various regression methods are discussed in the previous sections of this chapter.
- Select a method for X-axis and one or more methods for the Y-axis.
- Specify the required parameters of the selected methods in the various options boxes.
- "**Display User Selections**" option stores the user selected options for the various methods into a new worksheet for reference.
- "**Store Residuals to Worksheet**" options stores the residuals of each of the selected y-axis methods and the x-axis method in a new worksheet.
- Click on "**OK**" to continue or "**Cancel**" to cancel the options window.

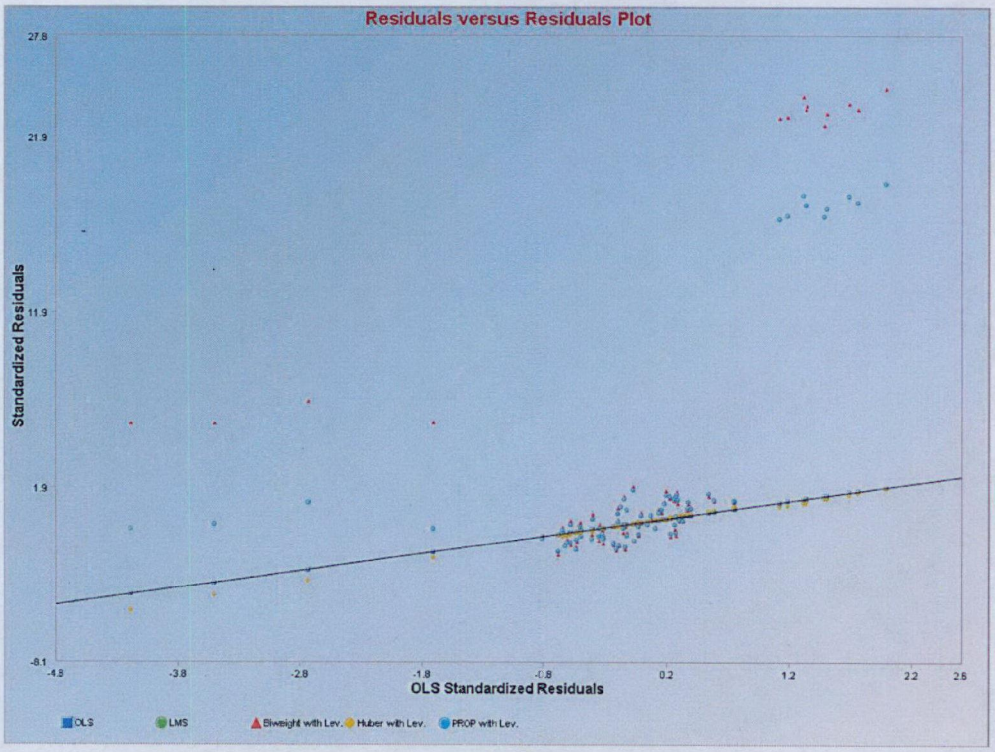
- Click on "**OK**" to continue or "**Cancel**" to cancel the generation of R-R Plots.

Output for R-R Plots.
Data Set: Bradu.
Methods: 13 (All) methods on Y-axis vs. Observed Y on X-axis.



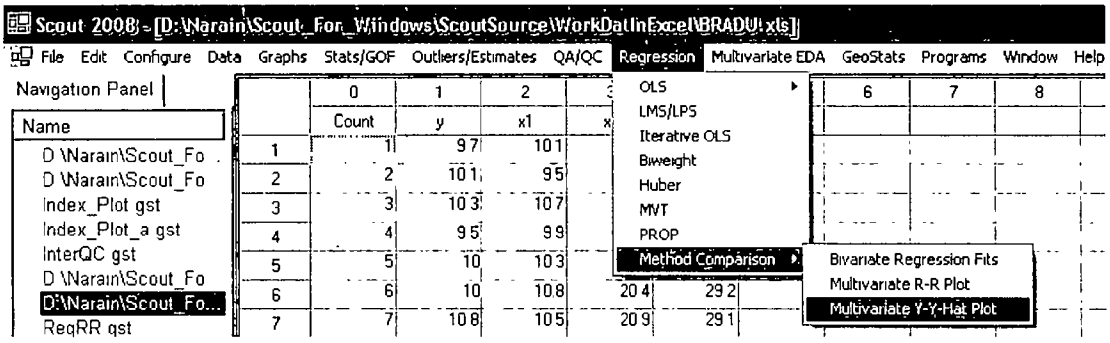
Data Set: Bradu.
Methods: 5 (OLS, LMS, Biweight, Huber and PROP with leverages) methods on Y-axis vs. OLS on X-axis.





9.9.3 Multivariate Y-Y-hat Plots

1. Click **Regression ▶ Method Comparison ▶ Multivariate Y-Y-hat Plot.**

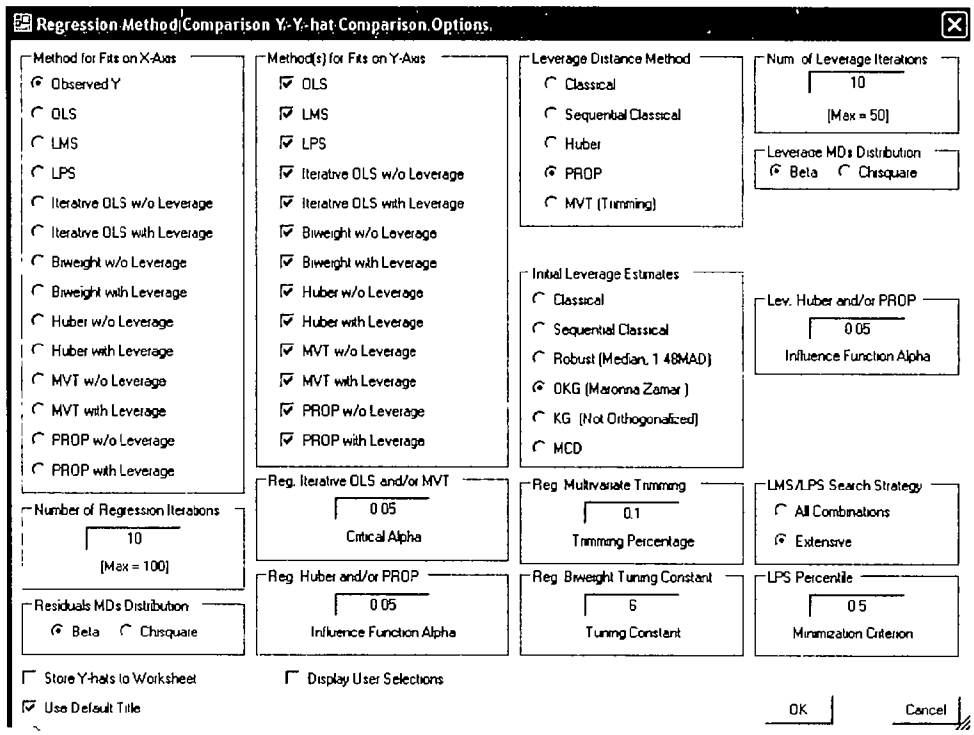


2. The "**Select Variables**" screen (Section 3.3) will appear.

   - Select the dependent variable and one or more independent variables from the "**Select Variables**" screen.

   - Click on the "**Options**" button to get the options window.

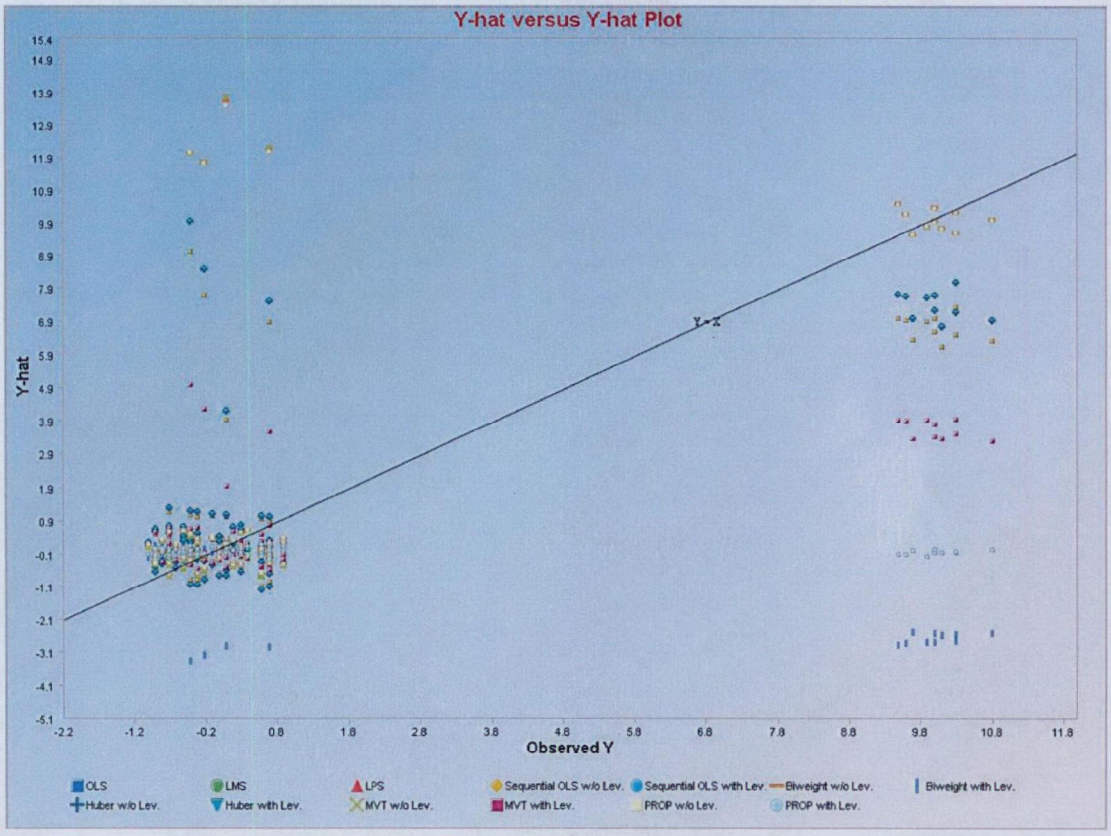
     - The options in the window shown below are the options when all the check-boxes in the "**Method(s) for Fits on Y-Axis**" are checked. The default option is of plotting the "**Observed Y**" against "**OLS**" fits.



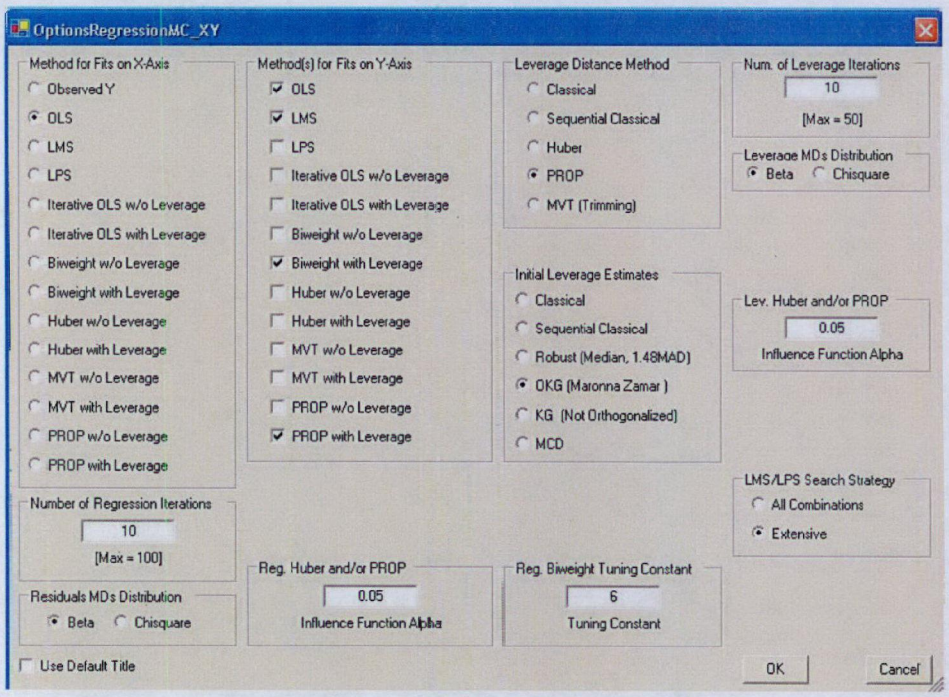
- The options required for the various regression methods are discussed in the previous sections of this chapter.
- Select a method for X-axis and one or more methods for the Y-axis.
- Specify the required parameters of the selected methods in the various options boxes.
- "**Display User Selections**" option stores the user selected options for the various methods into a new worksheet for reference.
- "**Store Residuals to Worksheet**" options stores the residuals of each of the selected y-axis methods and the x-axis method in a new worksheet.
- Click on "**OK**" to continue or "**Cancel**" to cancel the options window.

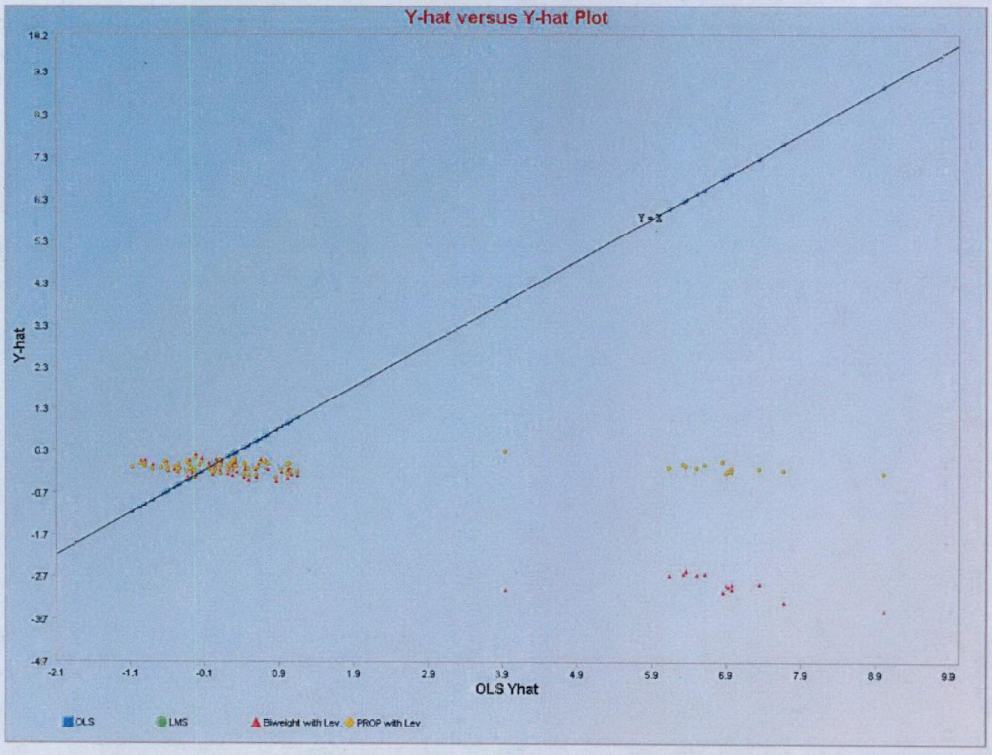
- Click on "**OK**" to continue or "**Cancel**" to cancel the generation of Y-Y-hat Plots.

Output for Y-Y-hat Plots.
Data Set: Bradu.
Methods: 13 (All) methods on Y-axis vs. Observed Y on X-axis.

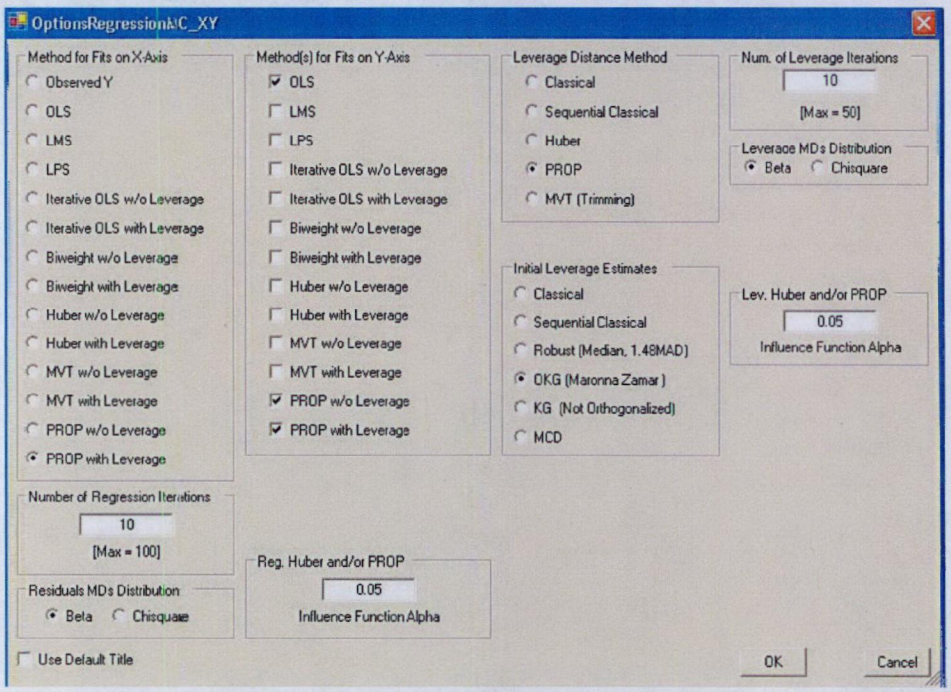


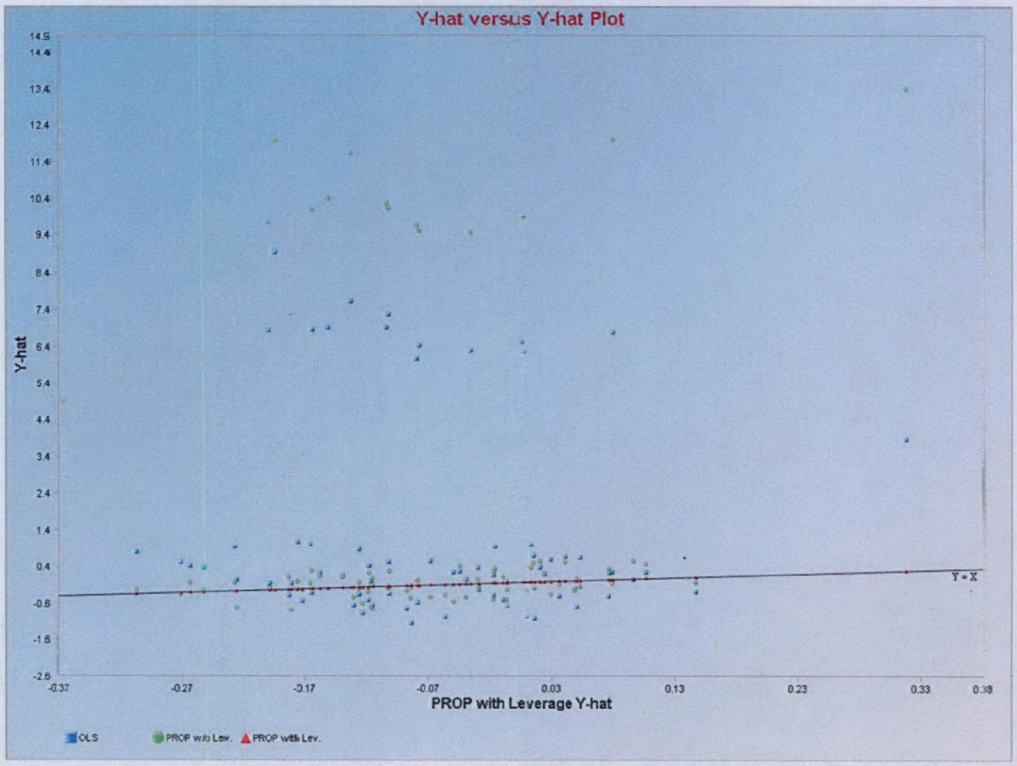
Data Set: Bradu.
Methods: 5 (OLS, LMS, Biweight, and PROP with leverages) methods on Y-axis vs. OLS on X-axis.





Data Set: Bradu.
Methods: 3 (OLS, PROP with and without leverage) methods on Y-axis vs. PROP with leverage on X-axis.





References

Agullo, J. (1997). "Exact Algorithms to Compute the Least Median of Squares Estimate in Multiple Linear Regression," in Statistical Procedures and Related Topics, ed. Dodge, Y., Institute of Mathematical Statistics, Hayward, CA, 133-146.

Belsley, D.A., Kuh, E., and Welsch, R.E. (1980). Regression Diagnostics: Identifying Influential Data and Sources of Collinearity, John Wiley and Sons, NY.

Chatterjee, S. and Machler, M. (1997). "Robust regression: A weighted least squares approach," Communications in Statistics Theory and Methods, 26, 1381–1394.

Cook, R.D., and Weisberg, S. (1982). Residuals and Influence in Regression, Chapman & Hall, London.

Dollinger, M.B., and Staudte, R.G. (1991). "Influence Functions of Iteratively Reweighted Least Squares Estimators," Journal of the American Statistical Association, 86, 709-716.

Draper, N.R., and Smith, H. (1984). Applied Regression Analysis, 2nd ed., John Wiley and Sons, NY.

Gnanadesikan, R., and Kettenring, J.R. (1972). "Robust Estimates, Residuals, and Outlier Detection with Multi-response Data," Biometrics, 28, 81-124.

Hadi, A.S., and Simonoff, J.S. (1993). "Procedures for the Identification of Multiple Outliers in Linear Models," Journal of the American Statistical Association, 88, 1264-1272.

Hawkins, D.M., Bradu, D., and Kass, G.V. (1984). "Location of Several Outliers in Multiple Regression Data Using Elemental Sets," Technometrics, 26, 197-208.

Hawkins, D.M., and Simonoff, J.S. (1993). "High Break Down Regression and Multivariate Estimation," Applied Statistics, 42, 423-432.

Hettmansperger, T.P., and Sheather, S.J. (1992). "A Cautionary Note on the Method of Least Median Squares," The American Statistician, 46, 79-83.

Neter, J., Kutner, M.H., Nachtsheim, C.J., and Wasserman W. (1996). Applied Linear Statistical Models, 4th ed., McGraw-Hill, Boston.

Olive, D.J. (2002). "Applications of Robust Distances for Regression," Technometrics, 44, 64-71.

Rousseeuw, P.J. (1984). "Least Median of Squares Regression," Journal of the American Statistical Association, 79, 871-880.

Rousseeuw, P.J., and Leroy, A.M. (1987). Robust Regression and Outlier Detection, John Wiley and Sons, NY.

Ruppert, D. (1992). "Computing S-Estimators for Regression and Multivariate Location/Dispersion," Journal of Computational and Graphical Statistics, 1, 253-270.

Ruppert, D., and Carroll, R.J. (1980). "Trimmed Least Squares Estimation in the Linear Model," Journal of the American Statistical Association, 75, 828-838.

Simpson, D.G., Ruppert, D., and Carroll, R.J. (1992). "On One-Step GM Estimates and Stability of Inferences in Linear Regression," Journal of the American Statistical Association, 87, 439-450.

Simpson, J.R., and Montgomery, D.C. (1998a). "The development and evaluation of alternative generalized M estimation techniques," Communications in Statistics — Simulation and Computation, 27, 999–1018.

Singh, A. and Nocerino, J.M. (1995). Robust Procedures for the Identification of Multiple Outliers, Handbook of Environmental Chemistry, Statistical Methods, Vol. 2. G, pp. 229-277, Springer Verlag, Germany.

Stromberg, A.J. (1993). "Computing the Exact Least Median of Squares Estimate and Stability Diagnostics in Multiple Linear Regression," SIAM Journal of Scientific and Statistical Computing, 14, 12891299.

Welsh, A.H. (1986). "Bahadur Representations for Robust Scale Estimators Based on Regression Residuals," The Annals of Statistics, 14, 1246-1251.

Yohai, V.J., and Zamar, R.H. (1988). "High break down point estimates of regression by means of the minimization of an efficient scale," Journal of the American Statistical Association, 83, 406–413.