REGRESSION ANALYSIS

ESTIMASI PARAMETER REGRESI LINIER BERGANDA DENGAN METODE ROBUST S

SKRIPSI

NUR KUMALA

**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS AIRLANGGA
SURABAYA**

2005

ESTIMASI PARAMETER REGRESI LINIER BERGANDA DENGAN METODE ROBUST S

SKRIPSI

Sebagai Salah Satu Syarat untuk Memperoleh Gelar Sarjana Sains Bidang Matematika di Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Airlangga

Oleh :

NUR KUMALA
NIM : 080012118

Tanggal Lulus : 10 Mei 2005

Disetujui Oleh :

Pembimbing I	Pembimbing II
Drs. SULIYANTO, M.Si.	**NUR CHAMIDAH, S.Si, M.Si.**
NIP 131 933 016	**NIP 132 205 653**

LEMBAR PENGESAHAN SKRIPSI

Judul : **ESTIMASI PARAMETER REGRESI LINIER BERGANDA DENGAN METODE ROBUST S**

Penyusun : **NUR KUMALA**

NIM : **080012118**

Tanggal Ujian : **10 Mei 2005**

Disetujui Oleh :

Pembimbing I

Drs. SULIYANTO, M.Si.
NIP 131 933 016

Pembimbing II

NUR CHAMIDAH, S.Si, M.Si.
NIP 132 205 653

Mengetahui :

Ketua Jurusan Matematika
FMIPA Universitas Airlangga

DEPARTEMEN PENDIDIKAN NASIONAL UNIVERSITAS AIRLANGGA JURUSAN MATEMATIKA FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

Drs. MOH. IMAM UTOYO, M.Si
NIP 131 801 397

PEDOMAN PENGGUNAAN SKRIPSI

Skripsi ini tidak dipublikasikan, namun tersedia di perpustakaan dalam lingkungan Universitas Airlangga. Diperkenankan untuk dipakai sebagai referensi kepustakaan, tetapi pengutipan seijin penulis dan harus menyebutkan sumbernya sesuai kebiasaan ilmiah.

Dokumen skripsi ini merupakan hak milik Universitas Airlangga.

KATA PENGANTAR

Puji syukur yang sedalam–dalamnya penulis panjatkan kehadirat Allah SWT atas karunia dan rahmat-Nya dan disadari dengan keteguhan hati akhirnya penulis dapat menyelesaikan skripsi yang berjudul *"Estimasi Parameter Regresi Linier Berganda Dengan Metode Robust S"*.

Pada kesempatan ini kami ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada Bapak Drs. Suliyanto, M.Si. selaku dosen pembimbing I dan Ibu Nur Chamidah, S.Si., M.Si. selaku dosen pembimbing II yang telah meluangkan waktu untuk memberikan nasehat dan bimbingan serta rekan-rekan yang telah banyak membantu penyelesaian skripsi ini.

Akhir kata kami mengharapkan semoga skripsi ini dapat memberikan sesuatu yang bermanfaat bagi pembaca pada umumnya dan jurusan matematika pada khususnya.

Surabaya, Juni 2005

Penyusun

بسم الله الرحمن الرحيم

" Sesungguhnya sesudah kesulitan itu ada kemudahan "

" Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya "

(QS. Al-Baqarah : 286)

Ku persembahkan Skripsi ini

Sebagai Ungkapan Terima kasih dan Baktiku

Kepada Bapak dan Ibuku Tercinta

Atas segala Do'a, Kasih sayang dan Perhatian yang

Diberikan dengan Tulus dan Ikhlas

UCAPAN TERIMA KASIH

Syukur Alhamdulillah kehadirat Allah SWT karena hanya dengan rahmat dan hidayah-Nya, sehingga penulisan skripsi dengan judul *Estimasi Parameter Regresi Linier Berganda Dengan Metode Robust S* ini dapat diselesaikan dengan baik. Serta shalawat dan salam senantiasa tercurahkan kepada junjungan kita Nabi Muhammad SAW.

Dalam kesempatan ini penulis mengucapkan terima kasih yang sebesar-besarnya kepada :

1. Drs. H. A. Latief Burhan, M. S selaku Dekan Fakultas MIPA Unair.
2. Drs. Suliyanto, M.Si selaku dosen pembimbing/penguji I dan Nur Chamidah, S.Si, M.Si selaku dosen pembimbing/penguji II, atas bimbingan dan kesabarannya.
3. Drs. Eko Chahyono, M.Si dan Toha Saifudin, S.Si, M.Si selaku dosen penguji IV dan III.
4. Ir. Elly Anna, M.Si selaku dosen wali.
5. Bapak dan Ibu dosen jurusan Matematika atas bimbingannya.
6. Ibu, Bapak dan Saudara-Saudaraku atas do'a dan kasih sayangnya.
7. Keponakkanku Affan, Tia dan Farris yang dapat membuatku tertawa saat aku lagi bingung membuat program skripsi.
8. Sahabatku dalam suka dan duka "Yuli" dan "Laily", atas segala suport, nasehat dan do'anya, semoga sukses selalu menyertaimu.

9. Temanku Indri atas segala bantuannya dalam membuat program, semoga cepat dapat kerja yang sesuai dengan hatimu.
10. Teman-teman seperjuanganku Titian, Naniek, Mita, Wiwin, Yuli mulyana,mbak Herlina (jangan pesimis ya, aku selalu siap membantumu) dan semua teman-teman angkatan 2000.
11. Temanku sigit atas pinjaman kastu perpusnya.
12. Semua pihak yang telah membantu dalam penyelasaian skripsi ini yang tak bisa penulis sebutkan satu persatu.

Semoga skripsi ini dapat memberikan manfaat kepada penulis dan pihak lain yang menggunakannya.

TTD

NUR KUMALA

Nur kumala, 2005. *Estimasi regresi Linier Berganda Dengan Metode Robust S*. Skripsi ini dibawah bimbingan Drs. Suliyanto, M.Si, dan Nur Chamidah, S.Si., M.Si. Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam. Universitas Airlangga.

ABSTRAK

Estimasi parameter regresi linier berganda bertujuan untuk menjelaskan pengaruh satu atau lebih peubah bebas $\mathbf{x}_i$ terhadap peubah respon y_i. Penyelesaian yang sering digunakan adalah metode *Ordinary Least Squares* (OLS). Tujuan estimator OLS ini adalah meminimumkan jumlah kuadrat residual. Salah satu kelemahan metode ini adalah kurang bisa mendeteksi dan tidak robust terhadap kehadiran *outlier* .

Alternatif untuk mamperbaiki kelemahan OLS adalah menggunakan estimator yang bersifat *robust* yang mampu mendeteksi dan dapat bertahan terhadap kehadiran *outlier* dalam jumlah tertentu pada data pengamatan. Pada skripsi ini membahas metode estimasi *robust* yang berasal dari keluarga *High Breakdown Value* yaitu Estimasi Robust S dalam analisa regresi linier berganda. Hasil dari penelitian ini menunjukkan kemampuan estimator S yang mempunyai sifat lebih baik saat adanya *outlier*. Hasil penerapan terhadap data dengan model regresi linier berganda dengan estimator S adalah :

OLS : $\hat{y} = -299.515 + 0.059x_2 + 0.934x_3$ $\qquad R^2 = 0.476$

S : $\hat{y} = -299.518 + 0.056x_2 + 0.931x_3$ $\qquad R^2 = 0.527$

Dalam hal mendeteksi kehadiran *outlier* pada data , Jika digunakan estimator OLS dengan kriteria *Mahalobis Distance* (MD) tidak terdeteksi adanya *outlier*. Sedangkan estimator S dengan menggunakan metode *Mahalobis Distance* (MD) menghasilkan *outlier* sebanyak 2 pengamatan dan jika menggunakan kriteria *Robust Distance* (RD) menunjukkan kemampuan mendeteksi adanya *outlier* sebanyak 12 pengamatan, terdiri dari 2 pengamatan tergolong outlier vertikal,10 pengamatan tergolong titik *leverage* baik.

Kata Kunci : *Outlier*, Estimasi S, Regresi Linier Berganda

Nur kumala, 2005. *Parameter Estimation Multiple Linier Regression with Robust Method S.* This Skripsi in guided by Drs. Suliyanto, M.Si, and Nur Chamidah, S.Si., M.Si. Mathematical Departement Faculty of Mathematics and Natural Science. Airlangga University.

ABSTRAK

The purposed of estimating parameter of multiple linier regression models is to describe the effect of one or some variables $\mathbf{x}_i$ to the response variable y_i. *Ordinary Least Squares* (OLS) method is usualy utilised. The goal of the OLS is to minimise the residuals sum of squares. However, it has some weaknesses, i.e. sensitive to any violation from the model assumptions. For example, the presence of outliers can cause the result of the OLS estimator unfitted.

One alternative to recover the OLS weaknesses is using a robust method that can be resistant in presence of outliers in the data. This research studies about robust estimation from the family of *High Breakdown Point Estiator*, specifically using Estimator S. The result of this research shows S have some benefits when outliers presences in data, i.e. resistance of scale estimate value, utulity to detect and classify outliers from its source. Application OLS and S on Education Fee data that multilpe linier regression models and determination coefficients are as follows :

OLS : $\hat{y}$ = -229.515 + 0.059x_2 + 0.934x_3 R^2 = 0.476

S : $\hat{y}$ = -229.518 + 0.056x_2 + 0.931x_3 R^2 = 0.527

Comparing OLS estimator with Malahanobis Distance (MD) method did not detect any outlier presence in the data. If S estimator with Malahanobis Distance (MD) method can detect 2 outlier and with Robust Distance (RD) method can detect i.e. 11 observations consist of 2 observations classified as vertical outliers, 10 observations clasified as good leverage point .

Key Words : Outliers, S Estimation, Multiple Linier Regression.

DAFTAR ISI

DAFTAR TABEL

DAFTAR LAMPIRAN

BAB I

PENDAHULUAN

I.1 Latar Belakang

Model regresi merupakan suatu model yang mengambarkan hubungan antara peubah bebas dengan peubah respon. Analisi regresi linier sederhana mempelajari hubungan antara satu peubah bebas dengan satu peubah terikat, yang salah satu tujuannya untuk memperkirakan nilai rata-rata peubah respon, jika peubah bebasnya di ketahui. Menurut kenyataanya sering kali model persaman regresi linier sederhana ini tidak cukup mewakili untuk menyatakan data yang ada ke dalam model regresi, sehingga di perlukan perluasan model yang meliputi lebih dari dua peubah bebas. Hal ini akan membawa pada suatu pembahasan model regresi linier berganda yaitu peubah respon Y tergantung pada dua atau lebih peubah bebas X.

Metode statistik yang dapat digunakan untuk mengestimasi parameter dalam regresi linier berganda antara lain *ordinary least square* (OLS), metode kemungkinan maksimum atau *maximum likelihood estimator* (MLE), metode bootstrap, metode kuadrat terkecil terboboti (*weighted leastsquare*), atau banyak metode lain yang di gunakan berdasarkan pada keadaan datanya. Metode OLS banyak digunakan karena metode ini sangat sederhana dan mudah untuk dilakukan, tapi kurang baik dalam mendeteksi adanya *outlier* dalam data.

Metode estimasi robust dalam mengestimasi model regresi merupakan alat analisis data yang cukup baik jika datanya mengandung adanya *outlier* dan dapat mengatasi sisaan yang tidak harus berdistribusi normal. Metode estimasi robust ini mampu juga digunakan untuk mendeteksi adanya *outlier* dan mampu menunjukan resistensi atau kestabilan yang cukup baik walaupun ada *outlier*. Metode estimasi robust yang sering digunakan untuk mendeteksi *outlier* dalam data antara lain metode estimasi Huber M, estimasi *high breakdown value*, dan kombinasi dari dua metode ini. Estimasi Huber M kurang efektif jika digunakan pada data yang banyak memiliki *outlier*, karena estimasi Huber M tidak dapat membedakan *leverage points* baik dan *leverage points* buruk, maka dalam kasus-kasus seperti ini estimasi *high brekdown value* diperlukan. Estimasi *high brekdown value* meliputi estimasi *Least Trimmed Squaris* (LTS), estimasi *Least median squares* (LMS) dan estimasi S **(Colin chen, 2001).**

Estimasi LTS mengambil waktu perhitungan lebih sedikit dan lebih akurat, tetapi dalam prakteknya masih ditemui hambatan pada pengamatan dengan jumlah data yang besar **(Rousseeuw dan Leroy, 1997)**. Untuk mengatasi kelemahan pada estimasi LTS dalam hal jumlah data pengamatan yang besar dapat digunakan estimasi robust S, sehingga penulis tertarik untuk membahas estimasi robust S.

I.2 Rumusan Masalah

Berdasarkan latar belakang masalah tersebut diatas, maka rumusan masalah yang akan dibahas dalam tulisan ini adalah :

1. Bagaimana algoritma untuk mengestimasi parameter regresi linier berganda dengan metode estimasi robust S.
2. Bagaimana membuat program dengan menggunakan bahasa S-plus 2000 yang sesuai dengan algoritma metode estimasi robust S.
3. Bagaimana menerapkan program pada data *Biaya Pendidikan* (Rousseeuw dan Leroy) tahun 1986.

I.3 Tujuan

Berdasarkan rumusan masalah diatas, maka tujuan penulisan ini adalah:

1. Membuat algoritma estimasi parameter regresi linier berganda dengan metode estimasi robust S.
2. Membuat program dengan menggunakan bahasa S-plus 2000 yang sesuai dengan metode estimasi robust S.
3. Menerapkan pada data *Biaya Pendidikan* (Rousseeuw dan Leroy) tahun 1986 .

I.4. Manfaat

1. Menambah wawasan dan pengetahuan tentang metode estimasi robust S, pembuatan algoritma dan pemprograman statistik dengan bahasa S – Plus 2000.
2. Mengetahui kemampuan dan penerapan metode estimasi parameter regresi linier berganda salah satu keluarga *high breakdown value* yaitu S dalam mengatasi *outlier* pada data.

BAB II
TINJAUAN PUSTAKA

2.1 Regresi linier Berganda

Model regresi berbentuk linier yang terdiri dari beberapa peubah bebas dengan satu peubah respon disebut model regresi linier berganda. Secara umum model regresi linier berganda dengan p-1 peubah bebas dinyatakan sebagai:

$$y_i = \theta_1 x_{i1} + \theta_2 x_{i2} + \dots + \theta_p x_{ip} + \varepsilon_i , \; i = 1,2, \dots , n \tag{2.1}$$

dipilih $x_{i1} = 1$, sehingga didapat model regresi linier berganda dengan menggunakan intersep. Asumsi yang berlaku dalam model regresi linier berganda menurut **Draper, (1992)** adalah

1. $Cov(\varepsilon_i, \varepsilon_j) = 0, i \neq j = 1, \dots, n$
2. ε_i berdistribusi normal dengan $E(\varepsilon_i) = 0$ dan $var(\varepsilon_i) = \sigma^2$, untuk $i = 1,\dots,n$
3. Tidak ada multikolinier antara variabel X_i dan X_j, untuk $i \neq j = 1, \dots,p$.
4. Parameter $\theta_1, \theta_2, \dots, \theta_p$ berupa konstanta.

Jika $\mathbf{x} = (x_{ij})$ menyatakan suatu matriks nxp, $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$ adalah vektor peubah respon berukuran nx1 diberikan dan $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_p)^T$ adalah vektor parameter yang tidak diketahui berukuran px1, maka model regresi pada persamaan (2.1) dapat ditulis dalam bentuk matriks sebagai berikut :

$$\mathbf{Y} = \mathbf{X}\theta + \varepsilon \tag{2.2}$$

dimana $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T$ merupakan suatu vektor berukuran n dari *residual* yang tidak diketahui.

2.2 Estimasi *Ordinary Least Squares*

Teorema 2.1 (Myers dan Milton, 1991)

Jika model regresi linier berbentuk $\mathbf{y} = \mathbf{X}\theta + \varepsilon$, dimana $\mathbf{y}$ dan ε adalah suatu vektor berdimensi nx1, dan $\mathbf{X}$ adalah suatu matriks berdimensi nxp, yang mempunyai $\mathbf{X}^T\mathbf{X}$ *full-rank,* serta θ adalah suatu vektor parameter regresi berdimensi p x 1, maka dengan menggunakan estimator OLS diperoleh estimasi parameter regresi $\hat{\theta}$ sebagai berikut :

$$\hat{\theta} = (X^TX)^{-1}X^Ty \tag{2.3}$$

Bukti :

Berdasarkan persamaan (2.2), $\varepsilon_i = y_i - x_i^T\theta$, sehingga diperoleh :

$$\sum_{i=1}^{n} \varepsilon_i^2 = \varepsilon^T\varepsilon$$

dimana $\varepsilon^T = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$ dan dalam bentuk matriks dapat ditulis menjadi

$$\begin{aligned} Q = \varepsilon^T\varepsilon &= (y - x^T\theta)^T(y - x^T\theta) \\ &= (y^T - x\theta^T)(y - x^T\theta) \\ &= y^Ty - \theta x^Ty^T - \theta^Txy + \theta^Tx^Tx\theta \end{aligned} \tag{2.4}$$

dengan menurunkan persamaan (2.4) terdapat θ akan diperoleh persamaan normal

$$\frac{\partial Q}{\partial \theta} = -2x^Ty + 2x^Tx\theta$$

Nilai minimum Q dicapai pada saat

$$\frac{\partial Q}{\partial \theta} = 0$$

sehingga diperoleh

$$-2x^T y + 2x^T x\hat{\theta} = 0$$

$$2x^T x\hat{\theta} = 2x^T y$$

$$\hat{\theta} = (x^T x)^{-1} x^T y \qquad (\textit{terbukti})$$

2.2.1 *Mean Square Error* (MSE)

MSE diperoleh dari nilai rata – rata harapan dari kuadrat perbedaan estimator di sekitar nilai parameter populasi sebenarnya, dihitung melalui :

$$\hat{\sigma}^2 = \frac{\text{jumlah kuadrat residual}}{n-p}$$

$$= \frac{\sum_{i=1}^{n} \varepsilon_i^2}{n-p} \qquad (2.5)$$

Koefisien determinasi R^2, yang mempunyai nilai dalam batas $0 \leq R^2 \leq 1$, diperoleh melalui :

$$R^2 = 1 - \frac{\text{jumlah kuadrat residual}}{\text{jumlah kuadrat total}}$$

$$= 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \qquad (2.6)$$

koefisien determinasi digunakan untuk mengukur variasi peubah respon yang dijelaskan oleh peubah bebas dalam model regresi linier berganda.

(Sembiring, 1995)

2.2.2 Regresi Robust

Pengertian robust secara umum adalah kekar, sedangkan regresi robust adalah sebuah alat penting untuk menganalisa data terkontaminasi *outlier* dan memberikan hasil-hasil yang lebih fleksibel jika pada data terdapat *outlier*.

(Colin Chen, 2002)

2.2.3. Deteksi Titik *Leverage* Melalui Pengukuran jarak

Estimasi OLS mengukur pengaruh *outlier* yang berasal dari kelompok peubah bebas dilakukan dengan menghitung jarak suatu pengamatan ke pusat atau rata-rata data, yaitu dengan menggunakan persamaan *Mahalanobis Distance* (MD), yaitu

$$MD_i = \sqrt{(x_i - T(X))C^{-1}(X)(x_i - T(X))^T}, \; i = 1, 2, \ldots, n \qquad (2.7)$$

dengan

$$T(X) = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (2.8)$$

$$C(X) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - T(X))^T(x_i - T(X)) \qquad (2.9)$$

Metode tersebut menggunakan nilai rataan data T(X) dan matriks kovarian C(X) yang menggunakan seluruh data, sehingga estimasi yang dihasilkan tidak robust atau berarti dapat menghasilkan estimasi yang jauh berbeda saat data menggandung sejumlah *outlier*.

Kriteria *observasi* menggunakan *Mahalanobis Distance* (MD) dalam data regresi yaitu :

1. *Observasi regular*

   Jika suatu *observasi* memenuhi $|MD_i| \leq 2,5$ dan $\left|\frac{e_i}{s}\right| \leq 2,5$ serta $|MD_i| > 2,5$ dan $\left|\frac{e_i}{s}\right| \leq 2,5$ maka *observasi* tersebut dinamakan *Observasi regular*.

2. *Outlier*

   Jika suatu *observasi* memenuhi $|MD_i| \leq 2,5$ dan $\left|\frac{e_i}{s}\right| > 2,5$ serta $|MD_i| > 2,5$ dan $\left|\frac{e_i}{s}\right| > 2,5$ maka *observasi* tersebut dinamakan *outlier*.

(Rousseeuw dan Hubert, 1997)

2.3 Metode PRESS

PRESS singkatan dari *Prediction of Sum Squares*, pertamakali diperkenalkan oleh **D.M. Allen (1971)**, yang digunakan sebagai langkah awal dalam mendeteksi adanya *outlier* dalam data.

Prosedur metode PRESS sebagai berikut :

1. Membuang amatan ke-i pada peubah respon maupun peubah bebasnya, dengan i = 1, 2, ..., n.

2. Menduga persamaan regresi terhadap n-1 titik data yang tersisa dan dinyatakan sebagai $\hat{y}_{(i)}$.

3. Menggunakan setiap persamaan regresi yang diperoleh untuk meramalkan y_i oleh $\hat{y}_{(i)}$ sehingga (y_i - $\hat{y}_{(i)}$) untuk semua kemungkinan persamaan regresi.

4. Menghitung PRESS = $\sum_{i=1}^{n}(y_i - \hat{y}_{(i)})^2$.

Menurut **Jobson, (1991)** untuk menunjukkan adanya *outlier* dalam data, dapat digunakan $\dfrac{\text{PRESS}}{\sum_{i=1}^{n} e_i^2} > 1$. (2.10)

2.4 Estimasi S

Estimasi S dikenalkan oleh **Rousseeuw** dan **Yohai (1984),** misalkan ρ adalah sebuah fungsi sedemikian sehingga berlaku

(i) ρ adalah simetris dan kontinu, serta $\rho(0) = 0$

(ii) Jika c> 0 sedemikian sehingga ρ fungsi naik pada [0, c] dan konstan pada [c, ∞].

(iii) $\dfrac{K}{\rho(c)} = \dfrac{1}{2}$

(Rousseeuw dan Leroy, 2003)

Estimasi S didefinisikan sebagai berikut

$$\hat{\theta}_s = \arg \min_{\hat{\theta}} S(\theta) \tag{2.11}$$

dimana $S(\theta)$ penyelesaian dari

$$\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{e_i}{s}\right) = K \quad (2.12)$$

atau

$$\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{Y_i - X_i^T\theta}{s}\right) = K$$

dengan X_i^T adalah baris ke-i dari matriks X^T dan $K = E_{\Phi}[\rho]$.

(Colin Chen, 2002)

Dalam menentukan nilai ρ dapat digunakan persamaan sebagai berikut

$$\rho(t)=\begin{cases}\dfrac{t^2}{2}-\dfrac{t^4}{2c^2}+\dfrac{t^6}{6c^4} & \text{jika } |t| \le c \\ \dfrac{c^2}{6} & \text{jika } |t| > c\end{cases} \quad (2.13)$$

menurut Rousseeuw dan Leroy (2003), jika dipilih c = 1.547, maka didapat nilai breakdown point yang dilambangkan dengan ε_n^* sebesar 50%. *Breakdown value* untuk estimasi robust S didefinisikan sebagai berikut

$$\varepsilon_n^* = ([n/2] - p + 2)/n \quad (2.14)$$

Dalam metode ini nilai dugaan awal $\hat{\theta}$ yang digunakan adalah estimasi yang diperoleh dengan menggunakan metode kuadrat terkecil dinotasikan oleh $\hat{\theta}_{LS}$.

BAB III

METODE PENELITIAN

Dalam penulisan skripsi ini, langkah – langkah penyelesaian masalah :

1. Membuat algoritma metode estimasi robust S untuk mengestimasi parameter regresi linier berganda dengan langkah – langkah sebagai berikut :
   i. Mengimput data berpasangan $z_i = (\mathbf{x}_i, y_i)$, dengan $\mathbf{x}_i = (x_{i1}, x_{i2}, ..., x_{ip})$ untuk i = 1, 2, ..., n yang memenuhi model regresi linier berganda berbentuk $y_i = \theta_1 x_{i1} + \theta_2 x_{i2} + \ldots + \theta_p x_{ip} + \varepsilon_i$ dimana i = 1, 2, ..., n. Dipilih $x_{i1} = 1$, sehingga didapat model regresi linier berganda dengan menggunakan intersep.
   ii. Menguji data dengan merode PRESS untuk melihat adanya pencilan dalam data.
   iii. Menentukan nilai dugaan awal $\hat{\theta}$ dengan menggunakan metode OLS.
   iv. Menentukan nilai K = $E_{\Phi}\,\rho(c)$.
   v. Menentukan s yang memenuhi persamaan

   $$\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{e_i}{s}\right)=K$$

   dengan menggunakan Metode Newton Rapshon, sehingga diperoleh s minimum dengan melakukan iterasi $\hat{\theta}=\hat{\theta}_{ols}+0.001h$.
2. Membuat program sesuai dengan algoritma metode estimasi S dengan bahasa S – Plus 2000.

3. Membandingkan kemampuan mendeteksi pencilan antara estimasi robust S dengan estimasi OLS pada data *Biaya Pendidikan* , (Rousseeuw dan Leroy) tahun 1986, data ini terdiri dari 49 pengamatan dengan 3 variabel bebas.

BAB IV

PEMBAHASAN

Pada bab ini akan dibahas mengenai estimasi robust S yang dapat mengatasi sisaan yang tidak harus berdistribusi normal.

4.1 Estimasi S

Estimasi S dikenalkan oleh Rousseuw dan Yohai (1984). Estimasi ini bertujuan mengestimasi θ dengan cara minimumkan

$$s(e_1(\theta),\ldots, e_n(\theta)) \tag{4.1}$$

dengan $e_i = y_i - \hat{y}_i$

dimana $s(e_1(\theta), \ldots, e_n(\theta))$ penyelesaian dari:

$$\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{e_i}{s}\right) = K$$

atau

$$\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{Y_i - X_i^T\theta}{s}\right) = K \tag{4.2}$$

dengan X_i^T adalah baris ke-i dari matrik x transpose, dan $K = E_\Phi \rho(c)$

(Colin Chen, 2002)

Pada estimasi S digunakan fungsi sebagai berikut:

$$\rho(t) = \begin{cases} \dfrac{t^2}{2} - \dfrac{t^4}{2c^2} + \dfrac{t^6}{6c^4} & \text{jika } |t| \leq c \\ \dfrac{c^2}{6} & \text{jika } |t| > c \end{cases} \qquad (4.3)$$

Menurut Rousseeuw dan Leroy (2003), jika dipilih c = 1.547 maka didapat nilai breakdown point yang dilambangkan dengan ε_n^* sebesar 50% dan efisiensi asimptotis estimator S adalah 28.7%.

Pada saat *outlier* terdeteksi maka tidak seharusnya dilakukan estimasi ulang dengan tidak mengikutsertakan data *outlier* tersebut sebelum dilakukan kajian darimana sumber *outlier* terjadi. *Outlier* terjadi akibat banyak alasan dan setiap alasan membutuhkan perlakuan yang berbeda. Jika *outlier* timbul karena kesalahan pencatatan atau proses pengukuran, maka membuang data jenis tersebut adalah penyelesaian yang disarankan. Tetapi jika *outlier* merupakan suatu pengamatan yang cukup valid, maka perilaku pengamatan ini harus dipelajari pengaruhnya terhadap model yang digunakan. Disisi lain, adanya *outlier* yang terdeteksi sangat banyak jumlahnya mungkin merupakan petunjuk bahwa model yang dipilih belum cocok dengan data, sehingga masih perlu diperbaiki. Dengan demikian, analisa yang diperlukan untuk menanggani adanya *outlier* membutuhkan pemahaman tentang mengapa *outlier* tersebut muncul dan sekaligus pengetahuan khusus dalam hal apa data tersebut diambil dan kebutuhan proses untuk analisa selanjutnya.

4.2 Deteksi Titik *Leverage* dengan *Minimum Covariance Determinant* (MCD)

Metode *Robust Distance* (RD) untuk mengatasi kelemahan dari pengukuran jarak MD yang didasarkan nilai rataan peubah bebas T(X) dan matrix kovariansi antar peubah bebas C(X), yaitu dengan mengunakan metode perhitungan *Minimum Covariansi Determinant* (MCD). Estimator robust mempunyai kemampuan mengukur jarak sekaligus mempunyai kemampuan mendeteksi kehadiran titik *Leverage* melalui *Robust Distance* (RD) :

$$RD_i = \sqrt{(x_i - T(X))C^{-1}(X)(x_i - T(X))^T}$$

Menurut **Rousseuw dan Van Zomeren, (1990),** estimator MCD adalah pasangan (T,C), dimana T(X) adalah vektor berdimensi-pohon dan C(X) adalah matriks berdimensi (pxp) yang meminimumkan nilai deterniman c pada subset yang berisikan h anggota dari n pengamatan, dimana

$$h = [(n + p + 1)/2]$$

Kriteria *Observasi* mengunakan *Robust Distance* (RD) dalam data regresi yaitu:

1. *Observasi regular*

   Jika suatu *observasi* memenuhi $|RD_i| \leq 2{,}5$ dan $\left|\frac{e_i}{s}\right| \leq 2{,}5$ maka *observasi* tersebut dinamakan *observasi regular*.

2. *Outlier vertikal*

   Jika suatu *observasi* memenuhi $|RD_i| \leq 2{,}5$ dan $\left|\frac{e_i}{s}\right| > 2{,}5$ maka *observasi* tersebut dinamakan *Outlier vertical*.

3. *Leverage point* baik

Jika suatu *observasi* memenuhi $|RD_i| > 2{,}5$ dan $\left|\frac{e_i}{s}\right| \leq 2{,}5$ maka *observasi* tersebut dinamakan *Leverage point* baik.

3. *Leverage point* buruk

Jika suatu *observasi* memenuhi $|RD_i| > 2{,}5$ dan $\left|\frac{e_i}{s}\right| > 2{,}5$ maka *obsevasi* tersebut dinamakan *Leverage point* buruk.

(Rousseeuw dan Hubert, 1997)

4.3 Algoritma Estimasi S

Langkah 1:

Menginput data berpasangan $z_i = (\mathbf{x}_i, y_i)$, dengan $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})$ untuk i = 1,2, …, n yang memenuhi model regresi linier berganda berbentuk :

$$y_i = \theta_1 x_{i1} + \theta_2 x_{i2} + \ldots + \theta_p x_{ip} + \varepsilon_i, \text{ dimana } i = 1,2, \ldots, n \quad (4.4)$$

Persamaan (4.5) dapat ditulis menjadi

$$y = \mathbf{x}\theta + \varepsilon \quad (4.5)$$

dengan

$\mathbf{x} = (x_{i1}, \ldots, x_{ip})$; i = 1,2, ..., n

$\theta = (\theta_1, \theta_2, \ldots, \theta_p)^T$

Dari (4.5) dipilih $x_{i1} = 1$, sehingga didapat model regresi linier berganda dengan menggunakan intersep.

Langkah 2:

Menguji data dengan metode PRESS untuk melihat adanya pencilan dalam data.

Langkah 3:

Mengestimasi model (4.5) dengan mengunakan metode OLS sehingga didapat

$$y = \mathbf{x}\,\theta \tag{4.6}$$

Langkah 4:

Menghitung sisaan $e_i = y_i - \hat{y}_i$ (4.7)

Langkah 5:

Menghitung simpangan baku dari e_i yang dinyatakan sebagai

$$s_{ols} = \sqrt{\frac{n\sum_{i=1}^{n} e_i^2 - \left(\sum_{i=1}^{n} e_i\right)^2}{n(n-1)}} \tag{4.8}$$

Langkah 6:

Menentukan nilai K

$$K = \int_{-\infty}^{-c} \frac{c^2}{6}.\phi(t)dt + \int_{-c}^{c}\left(\frac{t^2}{2} - \frac{t^4}{2c^2} + \frac{t^6}{6c^4}\right)\phi(t)dt + \int_{c}^{\infty} \frac{c^2}{6}.\phi(t)dt \tag{4.9}$$

dengan $\phi(t)$ adalah fungsi densitas probabilitas normal baku dan memilih $c = 1.547$

Langkah 7:

Menghitung s yang memenuhi persamaan

$$\frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{e_i}{s}\right) = K \tag{4.10}$$

dengan menggunakan Metode Newton Rapshon, sehingga diperoleh s minimum dengan melakukan iterasi $\hat{\theta} = \hat{\theta}_{OLS} + 0.001h$

dimana h = -1500, ..., 1500

Langkah 8.

Menetapkan pengamatan yang tergolong outlier dengan analisa standarisasi residual, yaitu pengamatan yang memenuhi kiteria:

$$\left|\frac{r(\hat{\theta}_s)}{s_s}\right| > 2{,}5 \qquad \text{dan} \qquad |RD_i| > 2{,}5 \tag{4.12}$$

Langkah 9.

Menghitung nilai koefisien determinasi (R^2), yang mengukur berapa besar variasi peubah respon yang dapat dijelaskan oleh model linier yang dibentuk.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(\left(y_i - x\hat{\theta}_s\right)^2\right)_{i:n}}{\sum_{i=1}^{n}\left((y_i - \bar{y})^2\right)_{i:n}} \tag{4.13}$$

4.4 Algoritma *Minimum Covariance Determinant* (MCD)

Jenis *outlier* yang disebabkan oleh peubah bebas, yang disebut sebagai titik *leverage*, sangat sukar untuk diketahui sejak awal jika jumlah peubah bebas berbentuk multivariate (p > 2) karena penggunaan visualisasi seperti diagram sebar tidak mampu menggambarkan secara utuh dalam satu gambar dan beberapa *outlier* yang terkandung dalam data. Estimator robust mempunyai kemampuan mendeteksi kehadiran titik *leverage* melalui *Robust Distance* (RD) :

$$RD_i = \sqrt{(x_i - T(X))C^{-1}(X)(x_i - T(X))^T} \tag{4.14}$$

Metode MCD merupakan generalisasi dari estimator S, ditentukan dengan menggunakan cara menentukan rata-rata sebanyak h titik dari **X** dengan nilai deterniman matriks kovariansi yang paling kecil.

Algoritma yang digunakan dalam menghitung MCD sesuai yang diberikan oleh **Rousseeuw dan Van Driessen, (1999)** dalam FAST-MCD, yang dijelaskan secara rinci sebagai berikut :

Langkah 1.

a. Menginput data $\mathbf{X} = \{x_{i1}, x_{i2}, \ldots, x_{ip}\}$ untuk i = 1, 2, …, n.

b. Mendefinisikan nilai h didalam penelitian ini digunakan :

$$h = [(n + p + 1)/2] \tag{4.15}$$

Langkah 2.

Melakukan berulang-ulang langkah berikut ini hingga diperoleh suatu keadaan $\det(\mathbf{C_1(X)}) = \det(\mathbf{C_2(X)})$, dimana **C** adalah matriks kovariansi.

a. Mengambil sebanyak h titik pengamatan secara random dari sejumlah n data berpasangan yang diberi nama H_1.

b. Menghitung estimasi rata-rata T_1, matriks kovariansi $\mathbf{C_1(X)}$ dan $\det(\mathbf{C_1(X)})$.

$$T_1 = \frac{1}{h}\sum_{i\in H_1} x_i$$

$$\mathbf{C_1(X)} = \frac{1}{h}\sum_{i\in H_1}(x_i - T_1)^T (x_i - T_1) \tag{4.16}$$

det $(\mathbf{C_1(X)})$ = perkal;ian komponen nilai eigen matriks kovariansi $(\mathbf{C_1(X)})$

c. Menentukan sebanyak h titik pengamatan yang dinyatakan dalam H_2 dengan kriteria jarak yang paling minimum berdasarkan hasil estimasi rata-rata T_1, matriks kovariansi $\mathbf{C_1(X)}$, melalui :

1. Menghitung jarak tiap titik pengamatan ke pusat data.

$$d_{1(i)} = \sqrt{(x_i - T_1)^T C(X)_1^{-1} (x_i - T_1)} \quad (4.17)$$

2. Mengurutkan nilai-nilai jarak yang diperoleh.

$$d_1(p_1) \leq d_1(p_2) \leq \ldots \leq d_1(p_h) \leq d_1(p_n)$$

3. Menetapkan h titik pengamatan dengan jarak yang paling minimum.

$$H_2 = \{ p_1, p_2, \ldots, p_h \}$$

4. Menghitung estimasi rata-rata T_2, kovariansi $\mathbf{C_2(X)}$ dan det($\mathbf{C_2(X)}$)

$$T_2 = \frac{1}{h}\sum_{i \in H_2} x_i$$

$$\mathbf{C_2(X)} = \frac{1}{h}\sum_{i \in H_2} (x_i - T_2)^T (x_i - T_2) \quad (4.18)$$

det ($\mathbf{C_2(X)}$) = perkalian komponen nilai eigen matriks kovariansi ($\mathbf{C_2(X)}$)

d. Memeriksa hasil, apakah nilai det ($\mathbf{C_1(X)}$) ≠ det ($\mathbf{C_2(X)}$).

Langkah 3.

Untuk mendapatkan hasil determinan matriks kovariansi yang paling minimum maka det($\mathbf{C_1(X)}$) = det ($\mathbf{C_2(X)}$)

$$T(X) = T_1 = T_2$$

$$\mathbf{C(X)} = \frac{med_i d^2_{(T(X), C_1(X))}(i)}{\chi^2_{p,0.5}} C_1(X) \quad (4.19)$$

dengan $med_i d^2$ merupakan median dari kuadrat jarak tiap titik pengamatan

Langkah 4.

Menghitung ulang berdasarkan nilai T(X) dan C(X) yang telah dihitung pada langkah (3) diatas untuk diperoleh pengamatan yang tergolong titik *leverage* atau **x** – *outlier(s)* dan memberikan bobot nol bagi pengamatan jenis tersebut.

$$w_i = \begin{cases} 1 \; jika (x_i - T(X))C^{-1}(X)(x_i - T(X))^T \leq \chi^2_{p,0.975} \\ 0 \qquad untuk\, yang\, lain \end{cases} \tag{4.20}$$

Hasil akhir estimasi nilai T(**X**) dan C(**X**) adalah mengunakan semua pengamatan tanpa pengamatan yang tergolong sebagai titik *leverage* atau **x** – *outlier(s)*.

$$T(\mathbf{X}) = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} \tag{4.21}$$

$$\mathbf{C(X)} = \frac{\sum_{i=1}^{n} w_i (x_i - T(X))^T (x_i - T(X))}{\sum_{i=1}^{n} w_i - 1} \tag{4.22}$$

4.5 Penerapan Terhadap Data Biaya Pendidikan

Pada penerapan metode robust S digunakan data sekunder yang diambil dari hasil penelitian Chatterjee dan Price tahun 1977, tentang berapa besar biaya pendidikan untuk 50 negara bagian di Amerika. **(Rousseeuw dan Leroy, 1986)**

Jumlah pengamatan adalah sebanyak n = 49, dimana peubah respon, yaitu biaya pendidikan dan menggunakan peubah bebas sebanyak p = 3, yaitu jumlah penduduk (x_1), penghasilan perkapita setiap keluarga (x_2), jumlah penduduk yang berusia dibawah 18 tahun (x_3).

Dari data tersebut akan ditentukan estimasi OLS dan diperoleh persamaan sebagai berikut :

$\hat{y}$ = -278.457 + 0.0651x_1 + 0.048x_2 + 0.889x_3

kemudian mengukur pengaruh *outlier* yang berasal dari kelompok peubah bebas dengan menghitung jarak suatu pengamatan ke pusat atau rata-rata data, yaitu menggunakan persamaan MD.

Hasil uji dengan minitab terlihat bahwa peubah bebas jumlah penduduk (x_1), tidak signifikan sehingga hasil parameternya 0 sesuai pada lampiran (5). Selanjutnya dilakukan hasil estimasi parameter OLS ulang tanpa memuat peubah bebas jumlah penduduk (x_1), sehingga diperoleh persamaan regresi sebagai berikut :

$$\hat{y} = -299.515 + 0.059x_2 + 0.934x_3$$

Selanjutnya menentukan estimasi S pada data, dapat dilihat bahwa untuk nilai h = -305 mempunyai nilai s yang paling minimum, sehingga diperoleh nilai estimasi $\theta_0 = -229.518$, $\theta_2 = 0.056$, $\theta_3 = 0.931$ sesuai pada lampiran (7), dengan persaman regresi yang diperoleh adalah :

$$\hat{y} = -229.518 + 0.056x_2 + 0.931x_3$$

dan secara ringkas terlihat pada tabel di bawah ini :

Tabel 4.1. Hasil Estimasi Parameter Regresi Data Biaya Pendidikan

Peubah Bebas	Koefisien Regresi Estimasi OLS	Koefisien Regresi Estimasi S
x_2	0.059	0.056
x_3	0.934	0.931
Intersep	-229.515	-229.518
	$R^2 = 0.476$	$R^2 = 0.527$

(Hasil pada tabel di atas sesuai dengan lampiran 7)

Perbedaan hasil estimasi adalah estimator OLS menghitung MD dengan menggunakan seluruh data (termasuk titik *leverage*), sedangkan estimator S menggunakan metode MCD yang mempertimbangkan pengamatan umum.

Dari data tersebut selanjutnya akan digunakan metode MD untuk estimasi OLS dan hasilnya memperlihatkan bahwa tidak adanya *outlier* yang terdeteksi sesuai pada lampiran (4), sedangkan estimasi S mengunakan metode MC terdeteksi *outlier* sebanyak 2 pengamatan sesuai pada lampiran (11) dan menggunakan metode RD terlebih dahulu ditentukan nilai T(X) dan C(X) dengan menggunakan program MCD, menunjukkan kemampuan mendeteksi adanya *outlier* sebanyak 12 pengamatan, terdiri dari 2 pengamatan tergolong *outlier* vertikal, 10 pengamatan tergolong titik *leverage* baik sesuai pada lampiran (10).

Untuk dapat melihat perbedaan plot hasil estimasi antara robust dengan OLS pada dua dimensi maka diregresikan antara y dengan x_2 dan y dengan x_3, sedangkan regresi antara y dengan x_2 terdeteksi *outlier* sebanyak 12 pengamatan, terdiri dari 2 pengamatan tergolong *outlier* vertikal, 9 pengamatan tergolong titik *leverage* baik dan 1 pengamatan tergolong titik leverage buruk sesuai pada lampiran (13). Regresi antara y dengan x_3 terdeteksi *outlier* sebanyak 10 pengamatan yang tergolong titik *leverage* baik sesuai pada lampiran (14).

Kelebihan dari estimator robust S adalah pada kemampuan untuk membedakan pengamatan yang mempunyai pola umum dengan *outlier* berdasarkan sumber penyebabnya melalui kegiatan *residual* dan pengukuran jarak, sehingga dapat diketahui bagaimana pengaruh suatu titik pengamatan terhadap estimasi yang dihasilkan.

BAB V

KESIMPULAN DAN SARAN

5.1. Kesimpulan

1. Regresi robust adalah sebuah alat penting untuk menganalisa data yang terkontaminasi *outlier* dan memberikan hasil-hasil yang lebih fleksibel jika pada data terdapat *outlier*.
2. Regresi robust memiliki kemampuan mengukur jarak sekaligus mempunyai kemampuan mendeteksi kehadiran titik *leverage* melalui *Robust Distance* (RD).
3. Estimator S adalah metode estimasi parameter regresi yang meminimumkan kriteria :

$$s(e_1(\theta),\ldots, e_n(\theta))$$

dimana penyelesaian dari $s(e_1(\theta), \ldots, e_n(\theta))$ adalah :

$$\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{e_i}{s}\right) = K$$

2. Hasil estimasi pada data Biaya Pendidikan dengan model regresi linier berganda dengan estimasi OLS dan S adalah :

OLS : $\hat{y}$ = -299.515 + 0.059x_2 + 0.934x_3 $R^2 = 0.476$

S : $\hat{y}$ = -229.518 + 0.056x_2 + 0.931x_3 $R^2 = 0.527$

Dalam hal mendeteksi kehadiran *outlier*, estimator OLS dengan menggunakan metode *Malahanobis Distance* (MD) tidak terdeteksi adanya *outlier*. Sedangkan estimator S dengan mengunakan metode *Malahanobis Distance* (MD) menghasilkan *outlier* sebanyak 2 pengamatan dan jika menggunakan metode *Robust Distance* (RD) menunjukkan kemampuan mendeteksi adanya *outlier* sebanyak 12 pengamatan, terdiri dari 2 pengamatan tergolong *outlier* vertikal, 10 pengamatan tergolong titik *leverage* baik.

5.2 Saran

1. Melakukan penelitian lebih dalam mengenai sifat-sifat teoritis estimator robust S.
2. Mempelajari penggunaan estimator robust dalam model regresi nonparametrik.

DAFTAR PUSTAKA

Chen, C., 2002, **Robust Regression and Outlier Detection with the ROBUSTREG,** Institut SAS Inc., Cary, NC.

Draper, N. R., and Smith, H. (1992), **Analisis Regresi Terapan,** John Wiley and Sons, New York.

Wuryanto Eto, 1997, **Petunjuk Praktis S-Plus for Windows,** Laboratorium Komputer Jurusan Matematika FMIPA Universitas Airlangga.

Jobson, J.D., 1991,Applied Multivariate Data Analysis, Vol I : **Regression and Experimental design,** Spinger-Verlag.

Marazzi, A., 1993, **Algorithm, Routines and S Functions for Robust Statistics,** Wassaorth and Brooks/Cole, Pacific Grove, CA.

Rousseeuw dan Hubert, M. (1997), " **Recent Development in PROGRESS** ", dalam LI-Statistical Prosedure & Related Topics, diedit oleh Y. Dodge, The LMS Lecture Notes-Monograph Series, Volume 31, 201-214.

Rousseeuw dan Leroy, A. M. (2003), **Robust Regression and Outlier Detection,** John Wiley and Sons, New York, USA.

Rousseeuw dan Van Driessen, K. (1999), "**A Fast Algoritma for the Minimum Covariance Determinant Estimator** ", Technometrics, Volume 41, 212-223

Rousseeuw dan Van Zomeren, B. C. (1990), " **Unmasking Multivariate Outliers and Leverage Points** ", Journal ofthe American Statistical Association, Volume 85, 633-639.

Sembiring, R. K., 1995, **Analisis Regresi Terapan**, Penerbit ITB, Bandung.

Tsung, C. C., (2000), " **High Breakdown Estimation of Multivariate Location and Covariance with Missing Observations** ", Faculty of Psychologi and Educational Sciences University of Geneva.

LAMPIRAN 1

Data Sekunder Penelitian Biaya Pendidikan

No	Jumlah Penduduk (x_1)	Penghasilan Keluarga (x_2)	Jumlah Penduduk Yang Berusia di Bawah 18 Tahun (x_3)	Biaya Pendidikan (y)
1	508	3944	325	235
2	564	4578	323	231
3	322	4011	328	270
4	846	5233	305	261
5	871	4780	303	300
6	774	5889	307	317
7	856	5663	301	387
8	889	5759	310	285
9	715	4894	300	300
10	753	5012	325	221
11	649	4908	329	264
12	830	5753	320	308
13	738	5439	337	379
14	659	4634	328	342
15	664	4921	330	378
16	572	4869	318	232
17	701	4672	309	231
18	443	4782	333	246
19	446	4296	330	230
20	615	4827	318	268
21	661	5057	304	337
22	722	5540	328	344
23	766	5331	323	330
24	631	4715	317	261
25	390	3828	310	214
26	450	4120	321	245
27	476	3817	342	233
28	603	4243	339	250
29	805	4647	287	243
30	523	3967	325	216
31	588	3946	315	212
32	584	3724	332	208
33	445	3448	358	215
34	500	3680	320	221

No	Jumlah Penduduk (x_1)	Penghasilan Keluarga (x_2)	Jumlah Penduduk Yang Berusia di Bawah 18 Tahun (x_3)	Biaya Pendidikan (y)
35	661	3825	355	244
36	680	4189	306	234
37	797	4336	335	269
38	534	4418	335	302
39	541	4323	344	268
40	605	4813	331	323
41	785	5046	324	304
42	698	3764	366	317
43	796	4504	340	332
44	804	4005	378	315
45	869	5560	330	291
46	726	4989	313	312
47	671	4697	305	316
48	909	5438	307	332
49	831	5309	333	311

(Sumber : Rousseeuw dan Leroy, 1986)

LAMPIRAN 2

Hasil Program PRESS Data Biaya Pendidikan

```
>PRESS(datapressnew)
$ep:
 [1]  -0.1336578 -38.0872672  48.2417467 -45.6679276
 [5]  23.3188100 -15.5854614  73.4819004 -54.5673294
 [9]  29.4039551 -83.5706566 -31.7672556 -34.5550035
[13]  50.7353854  62.2006276  83.7092256 -48.4571094
[17] -39.5669683 -38.1457889 -24.4762486 -11.1007667
[21]  59.8032550  16.2933069  13.1446990 -12.7944912
[25]   6.4092438   9.3322472  -9.6490671 -18.9743990
[29] -14.6571314 -22.3721225 -21.2177557 -29.8799418
[33] -24.4156097   4.0306035 -24.8597500  -8.1667141
[37] -14.1542451  34.8610357  -4.8535481  35.5221157
[41]  -1.8416719  51.2403500  40.8117593  15.0459865
[45] -51.0057962  23.3490158  54.2864143  15.6374858
[49] -19.9690394

$er:
 [1]  -0.126320 -36.773251  41.065655 -42.100041
 [5]  20.214552 -13.864578  67.103055 -49.710447
 [9]  27.507157 -81.039952 -30.665076 -31.470301
[13]  45.646008  60.874911  80.837658 -45.974740
[17] -37.807691 -32.690433 -22.623973 -10.735995
[21]  56.553241  14.788688  12.516312 -12.451348
[25]   5.529443   8.663788  -8.995591 -18.279052
[29] -11.801379 -21.219724 -19.539831 -27.620543
[33] -21.292283   3.657710 -21.992764  -7.323953
[37] -12.870186  33.275940  -4.571927  34.033575
[41]  -1.774440  41.776461  37.592388  10.501778
[45] -46.626492  22.617553  51.489607  14.238695
[49] -18.541868

$epkur:
[1] 66694.61

$ekur:
[1] 57798.62

$ratio:
[1] 1.153914
```

Keterangan :

- ep adalah estimasi model semua kemungkinan regresi terhadap n titik data yang dinyatakan sebagai $\hat{y}$.
- er adalah estimasi model semua kemungkinan regresi terhadap n-1 titik data sisanya dan dinyatakan sebagai $\hat{y}_{(i)}$.
- epkur adalah menggunakan setiap persamaan regresi yang diperoleh untuk meramalkan y_i oleh $\hat{y}$ sehingga diperoleh simpangan ramalan (y_i - $\hat{y}$) untuk semua kemungkinan model regresi.
- ekur adalah mengunakan setiap persamaan regresi yang diperoleh untuk meramalkan y_i oleh $\hat{y}_{(i)}$ sehingga diperoleh simpangan ramalan (y_i - $\hat{y}_{(i)}$) untuk semua kemungkinan model regresi.
- ratio Data Biaya Pendidikan yaitu pembagian (y_i - $\hat{y}_{(i)}$) dengan (y_i - $\hat{y}$) dan diperoleh 1.154, berarti dalam data terdapat *outlier*.

LAMPIRAN 3

Program Deteksi *Outlier* Dengan Estimasi OLS

```
function(xy)
{
	n <- nrow(xy)
	x <- matrix(0, n, 4)
	x0 <- rep(1, n)
	x <- cbind(x0, xy[, -4])
	x1 <- cbind(xy[, -4])
	y <- xy[, 4]
	tetaOLS <- solve(t(x) %*% x) %*% t(x) %*% y
	p <- length(theta)
	yhat <- x %*% theta
	g <- (y - yhat)
	h <- g^2
	jkg <- sum(h)
	s <- sqrt(jkg/(n - 2))
	jkt <- sum((y - mean(y))^2)
	R2 <- 1 - (jkg/jkt)
	MSE <- jkg/(n - p)
	r <- abs(g/s)
	wi <- rep(1, n)
	TX <- apply(x1, 2, mean)
	c <- var(x1)
	vc <- solve(c)
	y1 <- sweep(x1, 2, Tx)
	MD <- sqrt(diag(y1 %*% solve(c) %*% t(y1)))
	for(i in 1:n)
		if(r[i] > 2.5) wi[i] <- 0
	return(tetaOLS, r, wi, R2, MSE, TX,  MD)
}
```

LAMPIRAN 4

Hasil estimasi Deteksi *Outlier* dengan Estimasi OLS Data Biaya Pendidikan

```
$tetaOLS:
            [,1]
   -278.45732291
x1    0.06512159
x2    0.04854698
x3    0.88933105

$MSE:
[1] 1284.414

$r:
          [,1]
 1 0.003602153
 2 1.048629855
 3 1.171032483
 4 1.200529133
 5 0.576440249
 6 0.395363740
 7 1.913517687
 8 1.417548255
 9 0.784396950
10 2.310943678
11 0.874448487
12 0.897410381
13 1.301646313
14 1.735915291
15 2.305175057
16 1.311020450
17 1.078128031
18 0.932203785
19 0.645147565
20 0.306148749
21 1.612678100
22 0.421715774
23 0.356916460
24 0.355063927
25 0.157678169
26 0.247057484
27 0.256519195
28 0.521247355
29 0.336529335
30 0.605103861
```

```
31 0.557199839
32 0.787630266
33 0.607172949
34 0.104303649
35 0.627147949
36 0.208850615
37 0.367007553
38 0.948900166
39 0.130373554
40 0.970504955
41 0.050600117
42 1.191301899
43 1.071988442
44 0.299469791
45 1.329605873
46 0.644964492
47 1.468282970
48 0.406032101
49 0.528741834

$R2:
[1] 0.4958406

$wi:
 [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[26] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1


$TX:
      x1      x2       x3
 661.7143 4655.98 324.5102

$MD:
 [1] 1.2867096 0.8224442 2.4820434 1.6644796 2.3260027
 [6] 2.0785509 1.7852761 1.8145941 1.4549242 0.6884436
[11] 0.8281246 1.8180583 1.9584159 0.2084672 0.8167029
[16] 1.2162898 1.0745446 2.4259116 1.6287656 0.7731043
[21] 1.2763122 1.8582267 1.1467626 0.5547522 2.3684166
[26] 1.5679915 1.5070419 0.8828610 2.8935426 1.2218469
[31] 1.6781849 1.6278734 2.2717239 1.8604164 2.1344966
[36] 1.9934263 1.8370959 1.0967766 1.3437058 1.0157903
[41] 0.8790275 2.8081697 1.6753544 3.6766059 1.7724667
[46] 0.7239618 1.2220241 1.8204568 1.5655446
```

Keterangan :

- Theta merupakan hasil estimasi parameter dengan metode OLS
- r merupakan residual
- wi menunjukkan bahwa pengamatan tergolong *outlier* dengan simbol 0 atau pengamatan umum dengan simbol 1.
- R2 merupakan nilai koefisien determinasi.
- MSE *merupakan Mean Square Error*.
- TX merupakan nilai rataan pengamatan.
- C merupakan matriks kovarian.
- MD merupakan hasil *Mahalanobis Distance.*

LAMPIRAN 5

Estimasi OLS Dengan Program Minitab

- ***Regression Analysis*** dengan semua peubah bebas.

```
The regression equation is
y = - 278 + 0.0651 x1 + 0.0485 x2 + 0.889 x3

Predictor        Coef     SE Coef         T        P
Constant       -278.5       132.5     -2.10    0.041
x1            0.06512     0.04919      1.32    0.192
x2            0.04855     0.01215      4.00    0.000
x3             0.8893      0.3313      2.68    0.010

S = 35.84       R-Sq = 49.6%     R-Sq(adj) = 46.2%

Analysis of Variance

Source            DF         SS         MS         F        P
Regression         3      56845      18948     14.75    0.000
Residual Error    45      57799       1284
Total             48     114644
```

- ***Regression Analysis*** tanpa menggunakan peubah bebas jumlah penduduk (x_1).

```
The regression equation is
y = - 300 + 0.0592 x2 + 0.934 x3

Predictor        Coef     SE Coef         T        P
Constant       -299.5       132.6     -2.26    0.029
x2           0.059221    0.009158      6.47    0.000
x3             0.9339      0.3323      2.81    0.007

S = 36.13       R-Sq = 47.6%     R-Sq(adj) = 45.3%

Analysis of Variance

Source            DF         SS         MS         F        P
Regression         2      54594      27297     20.91    0.000
Residual Error    46      60049       1305
Total             48     114644
```

LAMPIRAN 6

Program Estimasi S

```
nilaiK<-function()
{
cc <- 1.547
integrand1 <- function(re)
    {
        cc <- 1.547
        (cc^2/6) * (1/sqrt(2 * pi)) * exp(-1/2 * re^2)
    }
integrand2 <- function(re)
    {
        cc <- 1.547
        ((re^2/2) - (re^4/(2 * cc^2)) + (re^6/(6 * cc^4)
            )) * (1/sqrt(2 * pi)) * exp(-1/2 * re^2)
    }
integral1 <- integrate(integrand1, lower =  - Inf, upper
        =  - cc)$integral
integral2 <- integrate(integrand2, lower =  - cc, upper
        = cc)$integral
integral3 <- integrate(integrand1, lower = cc, upper =
        Inf)$integral
K <- integral1 + integral2 + integral3
return(K)
    }

rho<-function(r,s)
{
n<-length(r)
cc<-1.547
rho<-rep(0,n)
for(i in 1:n){
        if(abs(r[i]/s) <= cc)
            rho[i] <-(1/2 * (r[i]/s)^2) - (
                (r[i]/s)^4/(2 * cc^2)) + (
                (r[i]/s)^6/(6 * cc^4))
        else rho[i] <- 1/6 * cc^2
    }
return(rho)
}

fsigma<-function(r,s)
{
n<-length(r)
cc<-1.547
rho<-rep(0,n)
for(i in 1:n){
        if(abs(r[i]/s) <= cc)
            rho[i] <-(1/2 * (r[i]/s)^2) - (
                (r[i]/s)^4/(2 * cc^2)) + (
                (r[i]/s)^6/(6 * cc^4))

        else rho[i] <- 1/6 * cc^2
```

```
}
K<-nilaiK()
fs <- (1/n) *sum(rho) - K
return(fs)
}

dfsigma<-function(r,s)
{
n<-length(r)
cc<-1.547
drho<-rep(0,n)
for(i in 1:n) {
      if(abs(r[i]/s) <= cc)
          drho[i] <- (-r[i]^2/s^3)+((2*r[i]^4/cc^2)/s^5)-
    ((r[i]^6/cc^4)/s^7)
      else drho[i] <- 0
    }
dfs <- (1/n) *sum(drho)
return(dfs)
}

estimasi<-function(xy,h)
{
   xy <- as.matrix(xy)
   m <- ncol(xy)
   n <- nrow(xy)
   cc <- 1.547
   x0 <- rep(1, n)
   x <- cbind(x0, xy[,  - m])
   y <- xy[, m]
   bhat <- solve(t(x) %*% x) %*% (t(x) %*% y)
   yhat <- x %*% bhat
   rr <- y - yhat
   s0 <- sqrt(var(rr))
   K<-nilaiK()
   cat("nilai K=",K,"\n")
   bhat1 <- bhat + 0.00001 * h
   ytop <- x %*% bhat1
   r <- y - ytop
   s<-rep(0,5)
   ss<-rep(0,5)
   for(i in 1:5){
       s[i] <- s0 - fsigma(r,s0)/dfsigma(r,s0)
       ss[i]<-(1/n)*sum(rho(r,s[i]))
       cat("[ss",i,"]",ss[i],"  [s",i,"]",s[i],"\n")
                 s0<-s[i]
                 }

   sisa <- abs(ss - K)
   opt <- min(sisa)
   for(i in 1:5) {
       if(opt == sisa[i]) {
           hasil <- ss[i]
           smin<-s[i]
       }
   }
```

```
    p<-length(bhat1)
    jkg <- sum(r^2)
    jkt <- sum((y - mean(y))^2)
    R2 <- 1 - (jkg/jkt)
    MSE <- jkg/(n - p)
    SR<-r/smin

    wi <- rep(1, n)
    for(i in 1:n)
        if(abs(SR[i]) > 2.5) wi[i] <- 0


    ep <- rep(0, n)
    for(i in 1:n) {
        z <- x[ - i,  ]
        btopi <- solve(t(z) %*% z) %*% (t(z) %*%
            y[ - i])
        ep[i] <- y[i] - (x[i,  ] %*% btopi)
    }
    ratio.press <- sum(ep^2)/sum(r^2)

cat("Pada h=",h,"Nilai yang mendekati K adalah ss =",hasil,"dengan
    s=",smin,"\n")
    return(bhat,R2S,bhat1,MSE,R2,SR,wi)
    }
```

LAMPIRAN 7

Hasil Estimasi Robust S Data Biaya Pendidikan Tanpa Variabel Jumlah Penduduk

> Estimasi(datakunew2,-305)

```
nilai K= 0.199498812491075
[ss 1 ] 0.199509416460841   [s 1 ] 34.8033194819638
[ss 2 ] 0.19949881272982   [s 2 ] 34.8060856648421
[ss 3 ] 0.199498812491076   [s 3 ] 34.8060857271244
[ss 4 ] 0.199498812491076   [s 4 ] 34.8060857271244
[ss 5 ] 0.199498812491075   [s 5 ] 34.8060857271244
Pada h= -305 Nilai yang mendekati K adalah ss = 0.19949
8812491075 dengan s= 34.8060857271244
$bhat:
          [,1]
x0 -299.51532124
x2    0.05922092
x3    0.93386661

$bhat1:
          [,1]
x0 -299.51837124
x2    0.05617092
x3    0.93081661

$R2S:
[1] 0.527046

$SR:
          [,1]
 1  0.300662211
 2 -0.783939170
 3  1.117878457
 4 -0.497703372
 5  1.407338738
 6 -0.000944131
 7  2.535380342
 8 -0.790755163
 9  1.303591513
10 -1.798388057
11 -0.528848794
12 -0.387697504
13  1.704286825
```

```
14  2.181069489
15  2.698717854
16 -1.091116890
17 -0.561237708
18 -0.949629542
19 -0.544771726
20  0.010965626
21  1.996598884
22  0.776405606
23  0.845181075
24  0.017342655
25  0.285667005
26  0.410906866
27 -0.006473036
28 -0.125313206
29  0.412219593
30 -0.282337304
31 -0.095940150
32 -0.307223274
33 -0.356009525
34  0.458198033
35 -0.051004995
36  0.384660184
37  0.377454362
38  1.193230803
39  0.129017132
40  1.266084554
41  0.531382358
42  1.850600484
43  1.782645488
44  1.083292414
45 -0.832078956
46  1.147387419
47  1.947490052
48  1.157849887
49  0.067374398

$wi:
 [1] 1 1 1 1 1 1 0 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1
[26] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

$MSEOLS:
[1] 1305.423

$R2OLS:
[1] 0.4762074
```

LAMPIRAN 8

Program MCD

```
mcd<-function(x, h){

    x <- as.matrix(x)

    n <- nrow(x)

    p <- ncol(x)
    int <- c(1:n)
    pilih <- sample(int, h)
    H0 <- x[pilih, ]
    repeat {
        T0 <- apply(H0, 2, mean)
        c0 <- var(H0)
        det0 <- prod(eigen(c0)$values)
        y <- sweep(x, 2, T0)
        d <- sqrt(diag(y %*% solve(c0) %*% t(y)))
        k <- matrix(0, n, 2)
        k[, 1] <- c(1:n)
        k[, 2] <- d
        m <- sort(k[, 2])
        k1 <- k[k[, 2] <= m[h], ]
        H1 <- x[sort(k1[, 1]), ]
        T1 <- apply(H1, 2, mean)
        c1 <- var(H1)
        det1 <- prod(eigen(c1)$values)
        seldet <- abs(det1 - det0)
        if(abs(det1 - det0) < 0.002)
            break
        H0 <- H1
    }
    Tx <- T1
    y <- sweep(x, 2, Tx)
    d2 <- diag(y %*% solve(c1) %*% t(y))
    cx <- median(sort(d2))/qchisq(0.5, p) * c1
    d2w <- diag(y %*% solve(cx) %*% t(y))
    wi <- rep(0, n)
    for(i in 1:n) {
        if(d2w[i] <= qchisq(0.975, p))
            wi[i] <- 1
    }
    w <- matrix(0, n, 2)
    w[, 1] <- c(1:n)
    w[, 2] <- wi
    wm <- w[w[, 2] == 1, ]
    xb <- x[wm[, 1], ]
    TXA <- apply(xb, 2, mean)
    ya <- sweep(x, 2, TXA)
    CXA <- var(xb)
    RD <- sqrt(diag(ya %*% solve(CXA) %*% t(ya)))
    return(TXA, CXA, RD)
}
```

LAMPIRAN 9

Hasil MCD Data Biaya Pendidikan

```
> mcd(datax,26)
$TXA:
       x1       x2       x3
 651.3409 4717.409 322.2727

$CXA:
           x1          x2         x3
x1  21894.649   74930.997  -822.5370
x2  74930.997  401814.480 -3107.8816
x3   -822.537   -3107.882   175.1332

$RD:
 [1] 1.2495091 0.7470725 2.5309086 1.7412054 2.5578745
 [6] 2.2711287 1.8760190 1.7308884 1.7238280 0.8351745
[11] 0.7852064 1.7001458 2.0140423 0.5648556 0.8306865
[16] 1.3432335 1.1864947 2.5061487 1.5724535 0.7891056
[21] 1.6529787 1.7842032 1.0705537 0.5407649 2.5522281
[26] 1.5589838 1.7659375 1.4780939 3.1758664 1.2124164
[31] 1.7625996 2.1057545 3.0236143 1.8619425 3.5614367
[36] 2.1501417 2.8589023 1.0984580 1.6437168 0.9664137
[41] 1.1060113 4.6515606 2.7743808 5.9805230 1.7840154
[46] 0.7380310 1.4193758 1.8586904 1.8954966
```

Keterangan :

- TXA merupakan nilai rataan pengamatan
- CXA merupakan matriks kovarian
- RD merupakan hasil *Robust Distance*

LAMPIRAN 10

Pengelompokkan *Outlier* Pada Data Biaya Pendidikan

i	Estimator OLS			Estimator S		
	MD_i	r_i/s	Jenis *Observasi*	RD_i	r_i/s	Jenis *Observasi*
1	1.2867	0.0036	RO	1.2495	0.4221	RO
2	0.8224	1.0487	RO	0.7471	0.6393	RO
3	2.4820	1.1710	RO	2.5309	1.6737	GLP
4	1.6645	1.2005	RO	1.7412	0.7189	RO
5	2.3260	0.5764	RO	2.5579	1.1502	RO
6	2.0786	0.3954	RO	2.2711	0.1967	RO
7	1.7853	1.9135	RO	1.8760	2.6639	VO
8	1.8146	1.4175	RO	1.7309	0.9026	RO
9	1.4549	0.7843	RO	1.7239	1.3696	RO
10	0.6884	2.3109	RO	0.8352	1.9398	RO
11	0.8281	0.8744	RO	0.7852	0.4146	RO
12	1.8181	0.8974	RO	1.7001	0.3486	RO
13	1.9584	1.3016	RO	2.0140	1.9793	RO
14	0.2085	1.7359	RO	0.5649	2.3663	RO
15	0.8167	2.3051	RO	0.8307	3.0041	VO
16	1.2163	1.3110	RO	1.3432	0.8950	RO
17	1.0745	1.0781	RO	1.1865	0.6517	RO
18	204259	0.9322	RO	2.5061	0.5059	GLP
19	1.6288	0.6451	RO	1.5724	0.2408	RO
20	0.7731	0.3061	RO	0.7891	0.1847	RO
21	1.2763	1.6127	RO	1.6529	2.2696	RO
22	1.8582	0.4217	RO	1.7842	1.0407	RO
23	1.1468	0.3569	RO	1.0705	0.9559	RO
24	0.5548	0.3551	RO	0.5407	0.1234	RO
25	2.3684	0.1577	RO	2.5522	0.5732	GLP
26	1.5679	0.2471	RO	1.5589	0.7016	RO
27	1.5070	0.2565	RO	1.7659	0.1377	RO
28	0.8829	0.5212	RO	1.4781	0.0977	RO
29	2.8935	0.3365	RO	3.1759	0.1502	GLP
30	1.2218	0.6051	RO	1.2124	0.2207	RO
31	1.6782	0.5572	RO	1.7626	0.1662	RO
32	1.6279	0.7876	RO	2.1058	0.4325	RO
33	2.2717	0.6071	RO	3.0236	0.2733	GLP

i	Estimator OLS			Estimator S		
	MD_i	r_i/s	Jenis *Observasi*	RD_i	r_i/s	Jenis *Observasi*
34	1.8604	0.1043	RO	1.8619	0.5133	RO
35	2.1345	0.6271	RO	3.5614	0.2421	GLP
36	1.9934	0.2088	RO	2.1501	0.2371	RO
37	1.8371	0.3670	RO	2.8589	0.0933	GLP
38	1.0968	0.9489	RO	1.0984	1.4909	RO
39	1.3437	0.1304	RO	1.6437	0.3243	RO
40	0.0158	0.9705	RO	0.9664	1.5553	RO
41	0.8790	0.0506	RO	1.1060	0.4945	RO
42	2.8082	1.1913	RO	4.6562	1.7104	GLP
43	1.6754	1.0719	RO	2.7744	1.6547	GLP
44	3.6766	0.2995	RO	5.9805	0.7843	GLP
45	1.7725	1.3296	RO	1.7840	0.8312	RO
46	0.7239	0.6449	RO	0.7380	1.2305	RO
47	1.2220	1.4683	RO	1.4194	2.0834	RO
48	1.8205	0.4060	RO	1.8587	1.0296	RO
49	1.5655	0.5287	RO	1.8955	0.0091	RO
Pola Sebaran Data :				**Pola Sebaran Data :**		
Pengamatan umum (%)			100	Pengamatan Umum (%)		75.5
Outlier			0	Titik Leverage Baik (%)		20.5
				Titik Leverage Buruk(%)		4
Jumlah *Outlier*			0			24.5

Keterangan :

RO = Pengamatan Umum (*regular observation*)
O = *Outlier*
GLP = Titik *Leverage* Baik (*good leverage point*)
BLP = Titik *Leverage* Buruk (*bad leverage point*)

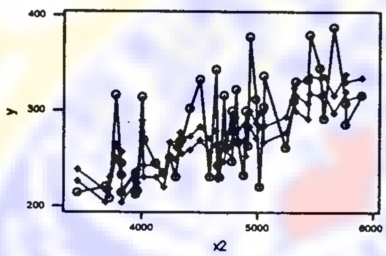
LAMPIRAN 11

i	Estimator S		
	MD_i	r_i/s	Jenis *Observasi*
1	1.2867	0.4221	RO
2	0.8224	0.6393	RO
3	2.4820	1.6737	RO
4	1.6645	0.7189	RO
5	2.3260	1.1502	RO
6	2.0786	0.1967	RO
7	1.7853	2.6639	O
8	1.8146	0.9026	RO
9	1.4549	1.3696	RO
10	0.6884	1.9398	RO
11	0.8281	0.4146	RO
12	1.8181	0.3486	RO
13	1.9584	1.9793	RO
14	0.2085	2.3663	RO
15	0.8167	3.0041	O
16	1.2163	0.8950	RO
17	1.0745	0.6517	RO
18	204259	0.5059	RO
19	1.6288	0.2408	RO
20	0.7731	0.1847	RO
21	1.2763	2.2696	RO
22	1.8582	1.0407	RO
23	1.1468	0.9559	RO
24	0.5548	0.1234	RO
25	2.3684	0.5732	RO
26	1.5679	0.7016	RO
27	1.5070	0.1377	RO
28	0.8829	0.0977	RO
29	2.8935	0.1502	RO
30	1.2218	0.2207	RO
31	1.6782	0.1662	RO
32	1.6279	0.4325	RO
33	2.2717	0.2733	RO
34	1.8604	0.5133	RO
35	2.1345	0.2421	RO
36	1.9934	0.2371	RO
37	1.8371	0.0933	RO
38	1.0968	1.4909	RO
39	1.3437	0.3243	RO
40	0.0158	1.5553	RO

i	Estimator S		
	MD_i	r_i/s	Jenis *Observasi*
41	0.8790	0.4945	RO
42	2.8082	1.7104	RO
43	1.6754	1.6547	RO
44	3.6766	0.7843	RO
45	1.7725	0.8312	RO
46	0.7239	1.2305	RO
47	1.2220	2.0834	RO
48	1.8205	1.0296	RO
49	1.5655	0.0091	RO
Pola Sebaran Data :			
Pengamatan Umum (%)			47(96)
Outlier (%)			2(4)
Jumlah *Outlier*			2
Keterangan : RO = Pengamatan Umum (*regular observation*); O = *Outlier*			

LAMPIRAN 12

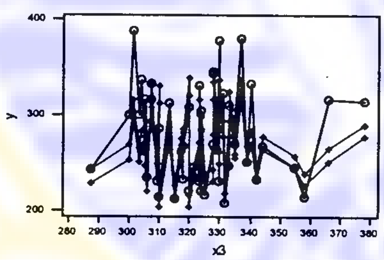
Kebaikan dari estimasi S dalam regresi linier berganda dapat dilihat pada gambar di bawah ini :



Keterangan :
1. Garis hitam menunjukkan garis regresi antara x_2 dengan y.
2. Garis merah menunjukkan garis regresi antara x_2 dengan yOLS.
3. Garis biru menunjukkan garis regresi antara x_2 dengan yS.

Gambar 4.1

(Plot antara x2 dengan y, x2 dengan yOLS dan x2 dengan yS)



Keterangan:
1. Garis hitam menunjukkan garis regresi antara x_3 dengan y.
2. Garis merah menunjukkan garis regresi antara x_3 dengan yOLS.
3. Garis biru menunjukkan garis regresi antara x_3 dengan yS.

Gambar 4.2

(Plot antara x3 dengan y , x3 dengan yOLS dan x3 dengan yS)

LAMPIRAN 13

i	Estimasi antara x_2 dengan y		
	MD_i	r_i/s	Jenis *Observasi*
1	1.2495	0.122	RO
2	0.7471	-0.756	GLP
3	2.5309	1.004	RO
4	1.7412	-0.723	RO
5	2.5579	0.899	RO
6	2.2711	0.024	RO
7	1.8760	2.225	RO
8	1.7309	-0.699	RO
9	1.7239	0.761	RO
10	0.8352	-1.557	RO
11	0.7852	-0.247	RO
12	1.7001	-0.059	RO
13	2.0140	2.280	RO
14	0.5649	2.231	RO
15	0.8307	2.875	VO
16	1.3432	-1.081	RO
17	1.1865	-0.869	RO
18	2.5061	-0.590	GLP
19	1.5724	-0.442	RO
20	0.7891	1.582	RO
21	1.6529	1.584	RO
22	1.7842	1.189	RO
23	1.0705	1.056	RO
24	0.5407	-0.096	RO
25	2.5522	-0.316	GLP
26	1.5589	0.184	RO
27	1.7659	0.220	RO
28	1.4781	0.172	RO
29	3.1759	-0.509	GLP
30	1.2124	-0.429	RO
31	1.7626	-0.514	RO
32	2.1058	-0.355	RO
33	3.0236	0.171	GLP
34	1.8619	0.056	RO
35	3.5614	0.513	GLP
36	2.1501	-0.203	RO
37	2.8589	0.583	GLP
38	1.0984	1.392	RO
39	1.6437	0.571	RO
40	0.9664	1.492	RO

i	Estimasi antara x_2 dengan y		
	MD_i	r_i/s	Jenis *Observasi*
41	1.1060	0.687	RO
42	4.6562	2.597	BLP
43	2.7744	2.114	GLP
44	5.9805	2.249	GLP
45	1.7840	-0.293	RO
46	0.7380	0.976	RO
47	1.4194	1.439	RO
48	1.8587	0.983	RO
49	1.8955	0.561	RO
Pola Sebaran Data :			
Pengamatan umum			37
Pencilan Vertikal			2
Titik *Leverage* Baik			9
Titik *Leverage* Buruk			1
Jumlah *Outlier*			12

Keterangan :
RO = Pengamatan Umum (*regular observation*)
O = *Outlier*
VO = *Outlier* Vertikal (*Vertical outlier*)
GLP = Titik *Leverage* Baik (*good leverage point*)
BLP = Titik *Leverage* Buruk (*bad leverage point*)

LAMPIRAN 14

i	Estimasi antara x_3 dengan y		
	MD_i	r_i/s	Jenis *Observasi*
1	1.2495	-0.356	RO
2	0.7471	-0.431	RO
3	2.5309	0.274	GLP
4	1.7412	0.071	RO
5	2.5579	0.763	RO
6	2.2711	1.074	RO
7	1.8760	2.312	RO
8	1.7309	0.509	RO
9	1.7239	0.758	RO
10	0.8352	-0.608	RO
11	0.7852	0.168	RO
12	1.7001	0.937	RO
13	2.0140	2.234	RO
14	0.5649	1.558	RO
15	0.8307	2.204	RO
16	1.3432	-0.422	RO
17	1.1865	-0.467	RO
18	2.5061	-0.145	GLP
19	1.5724	-0.436	RO
20	0.7891	0.219	RO
21	1.6529	1.425	RO
22	1.7842	1.594	RO
23	1.0705	1.335	RO
24	0.5407	0.093	RO
25	2.5522	-0.758	GLP
26	1.5589	-0.185	RO
27	1.7659	-0.361	RO
28	1.4781	-0.063	RO
29	3.1759	-0.282	GLP
30	1.2124	-0.695	RO
31	1.7626	-0.785	RO
32	2.1058	-0.825	RO
33	3.0236	-0.653	GLP
34	1.8619	-0.615	RO
35	3.5614	-0.141	GLP
36	2.1501	-0.408	RO
37	2.8589	0.269	GLP
38	1.0984	0.857	RO
39	1.6437	0.267	RO
40	0.9664	1.225	RO