

LAPORAN PENELITIAN  
KERJASAMA ANTAR PERGURUAN TINGGI  
TAHUN 2008



ck  
ck  
LP. 62/10  
Pen

PENGEMBANGAN *SOFTWARE* STATISTIKA R BERBAHASA INDONESIA  
DALAM PEMODELAN REGRESI NONPARAMETRIK UNTUK DATA  
DENGAN ERROR LOGNORMAL  
DAN TERDAPAT *OUTLIER*

BIDANG : STATISTIKA

Peneliti :

Nama Tim Pengusul (TPP)  
Nur Chamidah, S.Si.,M.Si.  
Toha Saifudin, S.Si.,M.Si  
**Universitas Airlangga**

Nama Tim Peneliti Mitra (TPM)  
Drs. I Made Tirta, Dip.Sc.,M.Sc.,PhD.  
Drs. Budi Lestari, PGD.Sc.,M.Si.  
**Universitas Jember**

DIBIYAI OLEH DIREKTORAT JENDERAL PENDIDIKAN TINGGI DEPARTEMEN PENDIDIKAN NASIONAL  
Sesuai Dengan Surat Perjanjian Pelaksanaan Penelitian dan Pengabdian Kepada Masyarakat  
Nomor : 016/SP2H/PP/DP2M/2007 Tanggal 29 Maret 2007

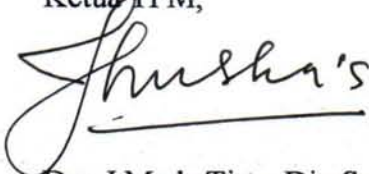
LEMBAGA PENELITIAN DAN PENGABDIAN KEPADA MASYARAKAT  
UNIVERSITAS AIRLANGGA  
TAHUN 2008

**HALAMAN PENGESAHAN**

- 1. Judul Penelitian : Pengembangan *Software* Statistika R Berbahasa Indonesia Dalam Pemodelan Regresi Nonparametrik Untuk Data Dengan Error Lognormal Dan Terdapat *Outlier*.
- 2. Ketua TPP
  - a. Nama Lengkap : Nur Chamidah, S.Si., M.Si.
  - b. Jenis Kelamin : Perempuan
  - c. NIP : 132 205 653
  - d. Jabatan Fungsional : Lektor
  - e. Jabatan Struktural : -
  - f. Bidang Keahlian : Pemodelan Statistika.
  - g. Program Studi/Jur : Matematika/Matematika
  - h. Perguruan Tinggi : Universitas Airlangga
- 3. Anggota Peneliti : 1. Toha Saifudin, S.Si., M.Si. (TPP)  
2. Drs. Budi Lestari, PGD.Sc.,M.Si. (TPM)
- 4. Ketua TPM
  - a. Nama Lengkap : Drs. I Made Tirta, Dip.Sc., M.Sc., Ph.D.
  - b. Jenis Kelamin : Laki-Laki
  - c. NIP : 131 474 500
  - d. Jabatan Fungsional : Lektor Kepala
  - e. Jabatan Struktural : Kepala P3AI-LP3 Universitas Jember.
  - f. Bidang Keahlian : Statistika Komputasi
  - g. Program Studi/Jur : Matematika / Matematika
  - h. Perguruan Tinggi : Universitas Jember.
- 5. Jangka waktu dan pendanaan penelitian
  - a. Jangka waktu yang diusulkan : 2 tahun;
  - b. Jangka waktu yang sudah dijalani : 2 tahun.
  - c. Biaya yang disetujui tahun II (2008) : Rp. 30.000.000,00

Surabaya,  
Ketua TPP,

Menyetujui  
Ketua TPM,



Drs. I Made Tirta, Dip.Sc., M.Sc., Ph.D.  
NIP. 131 474 500



Nur Chamidah, S.Si., M.Si.  
NIP. 132 205 653

Mengetahui  
Dekan FSAINTEK UNAIR,



Drs. Salamun, M.Kes.  
NIP. 131 696 506

Menyetujui  
Ketua LPPM UNAIR,



Dr. Bambang Sektiari L.,DEA.,Drh  
NIP. 131 837 004

## RINGKASAN

*Software R* merupakan salah satu yang termasuk dalam kelompok *Open Source Software* (OSS) yang bersifat legal, murah dan terbuka untuk dikembangkan serta dimodifikasi. Tujuan penelitian ini adalah meningkatkan kemampuan TPP dalam hal pemrograman OSS-R dengan menu Bahasa Indonesia yang terkait dengan pendeteksian pendeteksian *outlier* dalam pemodelan regresi nonparametrik dengan error lognormal berdasarkan metode robust LTS dan robust LMS serta mengimplementasikan pada data riil. Penelitian ini dilakukan dengan langkah-langkah : membuat algoritma program, mendefinisikan fungsi-fungsi R terkait, menggabungkan fungsi dalam satu skrip program, menguji coba program dengan data simulasi dan data real, menghubungkan program tersebut dengan suatu interface grafis (RGUI) dalam program baru yang dinamakan menu StatDemo, menambah item sesuai kebutuhan dalam bahasa Indonesia, mengemas program yang dihasilkan ke bentuk pustaka (library) yang mudah dipakai orang lain. Hasil yang diperoleh : Implementasi OSS-R pada data pohon *Gmelina Arborea Roxb* memberikan hasil bahwa metode robust LTS memiliki kemampuan lebih banyak dalam mendeteksi outlier dibandingkan metode robust LMS, paket demo analisis statistik dengan R dengan menu bahasa Indonesia yang terhubung dengan interface grafis dalam suatu library StatDemo yang bersifat interaktif. Demo analisis statistika dengan program R interaktif yang telah dihasilkan, sangat bermanfaat dalam pembelajaran untuk meningkatkan pemahaman mahasiswa khususnya terhadap teori pemodelan regresi nonparametrik.

## SUMMARY

Software R is one of Open Source Softwares (OSS) groups which is legal, cheap and opened for developing and modifying. The goal of the research are to improve TPP achievement in programming Software R in Indonesian menus which related to detection outlier in modeling of nonparametric regression with lognormally error distributed based on Robust LTS and Robust LMS methods. Both of those methods imply to real data. The research was done through steps: write program algorithm, define related function R, combine the functions into one script program, test the program by using both real and simulation data, connect the program to a graphical interface (RGUI) in new program called StatDemo menu, add item needed in Indonesian language term, package the obtained program in library term which is easily used by other users. The results obtained are : Implementation of OSS-R on *Gmelina Arborea Roxb* tree give result that robust LTS method has more ability than robust LMS method, statistical analysis demo package R in Indonesian menus which is related to an interactively graphical interface called StatDemo library. The statistical analysis demo by using interactive R-program which has been obtained is very useful in teaching process for improving students understanding of nonparametric regression modeling theory.

## CAPAIAN INDIKATOR KINERJA

Keberhasilan penelitian ini dapat dilihat dari hasil pencapaian indikator kinerja yang meliputi publikasi hasil penelitian melalui seminar nasional, *proceeding*, jurnal nasional terakreditasi serta penggunaan OSS-R pada tugas akhir mahasiswa yang dibimbing oleh ketua TPP, pada pembelajaran Rancangan Percobaan untuk pengolahan data hasil rancangan percobaan dan juga dalam workshop dan pelatihan analisis data menggunakan OSS-R. Secara rinci pencapaian indikator kinerja tersebut adalah sebagai berikut :

1. Seminar Nasional Seminar Nasional Matematika dan Statistika 2008, di FSaintek Universitas Airlangga Surabaya, tanggal 20 Desember 2008 dengan judul “Implementasi OSS Statistika R Pada Model Regresi Nonparametrik Dengan Error Lognormal Berdasarkan Estimator Robust *Least Trimmed Squares* (LTS)”.
2. Seminar Nasional Matematika 2009, di FMIPA Universitas Jember, tanggal 28 Pebruari 2009 dengan judul “Implementasi OSS-R Pada Model Regresi Nonparametrik Dengan Error Lognormal Berdasarkan Estimator Robust *Least Median Squares* (LMS)”
3. Prosiding Seminar Nasional Matematika dan Statistika 2008, ISBN : 978-979-19096-0-0, hal.55-64, dengan judul “Implementasi OSS Statistika R Pada Model Regresi Nonparametrik Dengan Error Lognormal Berdasarkan Estimator Robust *Least Trimmed Squares* (LTS)”
4. Jurnal Ilmu Dasar (Terakreditasi) FMIPA Universitas Jember dengan Judul “Perbandingan Metode Robust LTS dan Robust LMS Pada Model Regresi Nonparametrik Dengan Error Lognormal Berdasarkan Estimator P-Spline” (sedang dalam proses penerbitan).
5. Publikasi melalui *website* dengan alamat :<http://r.unej.ac.id> dan <http://elesys.fsaintek.unair.ac.id>
6. Pada semester genap 2008/2009, TPP akan memanfaatkan dan menggunakan *Software R* yang dalam pembelajaran mata kuliah rancangan percobaan untuk pengolahan datanya.

7. Sosialisasi program R dalam bentuk workshop dan pelatihan analisis data menggunakan OSS statistika R pada mahasiswa dan dosen KBK Statistika dan perwakilan KBK lain di Departemen Matematika FMIPA Unair dengan narasumber dan instruktur masing-masing adalah dari TPM ( I Made Tirta dan Budi Lestari) dan dari TPP ( Nur Chamidah dan Toha Saifudin), pada tanggal 21 Desember 2008 di LabKom 5 Departemen Matematika FMIPA Unair.
8. Penggunaan OSS-R pada skripsi mahasiswa yang dibimbing oleh ketua TPP yaitu
  - a. Marisa Rifada (NIM. 080513360) dengan judul skripsi “Estimasi Model Regresi Logistik Nonparametrik Aditif Berdasarkan Estimator Kernel”.
  - b. Iin Khoirotul Umroh (NIM 080513234) dengan judul skripsi : “Estimasi Model Regresi Logistik Nonparametrik Aditif Berdasarkan Estimator Lokal Linier”.



## DAFTAR ISI



	Halaman
HALAMAN PENGESAHAN	i
RINGKASAN DAN SUMMARY	ii
CAPAIAN INDIKATOR KINERJA	iii
PRAKATA	v
DAFTAR ISI	vi
DAFTAR GAMBAR	vii
DAFTAR LAMPIRAN	viii
DAFTAR SINGKATAN	ix
I. PENDAHULUAN	1
II. TUJUAN DAN MANFAAT PENELITIAN TAHUN KE-2	2
2.1. Tujuan Penelitian	2
2.2. Manfaat	3
III. TINJAUAN PUSTAKA	3
3.1. Estimator Polinomial lokal	3
3.2. Pohon Gmelina Arborea Roxb	5
IV. METODE PENELITIAN	6
V. HASIL DAN PEMBAHASAN	6
5.1. Estimator Robust LMS	7
5.2. Estimator Robust LTS	12
VI. KESIMPULAN DAN SARAN	20
6.1. Kesimpulan	20
6.2. Saran	20
DAFTAR PUSTAKA	21
LAMPIRAN	

**DAFTAR GAMBAR**

	<b>Halaman</b>
Gambar 1. Demo Analisis Statistika dengan R untuk Robust LMS	16
Gambar 2. Hasil Estimasi Berdasarkan Robust LMS dengan $K=16$ , Posisi $X_0$ data ke-15	16
Gambar 3. Hasil Estimasi Berdasarkan Robust LMS dengan $K=15$ , Posisi $X_0$ data ke-15	17
Gambar 4. Demo Analisis Statistika dengan R untuk Robust LTS	18
Gambar 5. Hasil Estimasi Berdasarkan Robust LTS dengan $K=16$ , Posisi $X_0$ data ke-15	18
Gambar 6. Hasil Estimasi Berdasarkan Robust LTS dengan $K=15$ , Posisi $X_0$ data ke-15	19

## DAFTAR SINGKATAN

1. OSS : *Open Source Software*
2. LMS : *Least Median Square*
3. LTS : *Least Trimmed Square*
4. GCV : *Generalized Cross Validation*
5. RGUI : *R-Graphical User Interface*
6. HTI : *Hutan Tanaman Industri*
7. K : *Konstanta Kerobustan*
8. MSE : *Mean Square Error*
9. FT : *Fungsi Tujuan*
10. PROGRESS : *Program for Robust Regression*





## I. PENDAHULUAN

Distribusi lognormal merupakan distribusi probabilitas dari setiap variabel random bilamana log dari variabel randomnya berdistribusi normal. Fenomena kelognormalan banyak ditemukan di berbagai bidang keilmuan, antara lain data ekologi seperti konsentrasi gizi, kepadatan penduduk dan biomassa (Anonim, 2006). Limpert, *et al.*, (2001) menyatakan bahwa fenomena kelognormalan juga tampak pada bidang fisis genetika, periode laten dari infeksi penyakit, phytomedicine dan mikrobiologi yang meliputi macam spesies, tipe bakteri, populasi bakteri, di bidang psikologi tanaman meliputi permeabilitas dan pergerakan, hubungan antara pengaruh dosis racun dan patogen terhadap unsur kimiawinya (hormon dan mineral), di bidang teknologi makanan yaitu pemrosesan makanan dengan proses dispersi dan filtering. Hubungan antara variabel respon yang berdistribusi lognormal dengan satu variabel prediktor dapat dinyatakan dengan model regresi multiplikatif error yang berbentuk :

$$y_i^* = m(x_i)\varepsilon_i^* \quad , i = 1, 2, \dots, n \quad (1.1)$$

dengan  $\varepsilon_i \sim LN(0, \sigma^2)$  dan  $m(x_i)$  merupakan fungsi regresi. Dengan melakukan transformasi logaritma natural pada kedua ruas persamaan (1) sehingga diperoleh:

$$y_i = m^*(x_i) + \varepsilon_i \quad (1.2)$$

dengan  $y_i = \ln y_i^*$ ,  $\varepsilon_i = \ln \varepsilon_i^*$ ,  $m^*(x_i) = \ln m(x_i)$  dan  $\varepsilon_i \sim N(0, \sigma^2)$

Untuk mengestimasi fungsi regresi  $f$  dapat dilakukan dengan menggunakan pendekatan parametrik dan nonparametrik.

Teknik regresi nonparametrik merupakan cara yang efektif dalam mengidentifikasi dan menguji struktur pada data regresi tanpa menggunakan asumsi yang terbatas pada bentuk hubungan antara variabel respon dan variabel prediktor. Pendekatan-pendekatan standar regresi nonparametrik seperti estimator polinomial yang terboboti fungsi kernel dan estimator spline sangat sensitif terhadap *unusual observation (outlier)*. Estimasi polinomial tersebut kurang dapat mendeteksi *outlier* dalam data dan tidak mempunyai kemampuan untuk membedakan pola pengamatan reguler dan *outlier* (Simonoff, 2001).

Regresi robust merupakan alat penting untuk menganalisa data yang terkontaminasi *outlier* dan memberikan hasil-hasil yang lebih fleksibel jika dalam data terdapat *outlier*. Metode yang umum digunakan dalam regresi robust adalah

estimasi Huber M, estimasi *high breakdown value* dan kombinasi dari kedua metode ini. Estimasi *high breakdown value* meliputi: estimasi *Least Median of Squares* (LMS), estimasi *Least Trimmed Squares* (LTS), dan estimasi S.(Rousseeuw dan Hubert, 1997).

Pada penelitian tahun kedua ini TPP membahas tentang pemodelan regresi nonparametrik dengan error lognormal yang sudah dicapai, untuk kasus data yang terkontaminasi *outlier* (pencilan) dengan menggunakan estimasi *Robust*. Estimator robust yang dibahas adalah Robust LMS yang meminimalkan median dari error kuadrat dan Robust LTS yang meminimalkan jumlah terpotong (*trimmed*) dari residual kuadrat berdasarkan estimator polinomial yang terboboti oleh fungsi Kernel. Selanjutnya, model tersebut diimplementasikan pada *Software* Statistika R yang berbasis OSS. Penggunaan OSS-R untuk pendeteksian *outlier* berdasarkan metode robust LTS yang diimplementasikan pada data Stackloss (Eubank,1988) telah dibahas pada Chamidah,*et.al.*(2008). Sedangkan Chamidah,*et.al.*(2009) telah membahas penggunaan OSS-R untuk metode robust LMS yang diimplementasikan pada data Cloud Point (Rousseeuw dan Leroy, 2003). Pada laporan ini akan dibahas tentang kedua metode robust tersebut yaitu robust LMS dan robust LTS yang diimplementasikan pada data pohon *Gmelina Arborea Roxb* untuk membandingkan kemampuan kedua metode robust tersebut dalam mendeteksi outlier.

## II. TUJUAN DAN MANFAAT PENELITIAN TAHUN KE-2

### 2.1. Tujuan Penelitian

- a. meningkatkan kemampuan TPP dalam hal pemrograman *Software* R yang berbasis OSS yang terkait dengan pemodelan regresi nonparametrik untuk kasus data yang terkontaminasi *outlier* (pencilan) dengan menggunakan estimator robust LTS dan LMS
- b. mengembangkan *Software* Statistika R dengan mempelajari *interface* (*R-Console*) dengan menu Bahasa Indonesia yaitu Rcmdr,
- c. mensosialisasikan hasil penelitian TPP yaitu pendeteksian outlier pada pemodelan regresi nonparametrik dengan metode robust LTS dan robust LMS berdasarkan estimator polinomial lokal yang terboboti oleh fungsi

$$W = \begin{pmatrix} K_{h(x_i)}(X_1 - x_0) & 0 & \dots & 0 \\ 0 & K_{h(x_i)}(X_2 - x_0) & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \dots & 0 & K_{h(x_i)}(X_n - x_0) \end{pmatrix} \quad (3.5)$$

W adalah matrik yang berisi fungsi bobot, dimana  $K_h(\cdot)$  adalah fungsi kernel  $K$ .

Untuk mendapatkan nilai  $\hat{\beta}$  digunakan metode kuadrat terkecil terboboti sebagai berikut :

$$\begin{aligned} S &= (y - X\beta)^T W (y - X\beta) \\ &= y^T W y - y^T W X \beta - \beta^T X^T W y + \beta^T X^T W X \beta \\ &= y^T W y - 2\beta^T X^T W Y + \beta^T X^T W X \beta \end{aligned} \quad (3.6)$$

Nilai dugaan kuadrat terkecil terboboti untuk  $\beta$  adalah  $\hat{\beta}$  yang bila disubstitusikan ke dalam persamaan (3.6) akan meminimumkan  $S$  diperoleh :

$$\begin{aligned} \frac{\partial S}{\partial \beta} &= -2\hat{\beta}^T X^T W y + 2\hat{\beta}^T X^T W X \hat{\beta} = 0 \\ \hat{\beta} &= (X^T W X)^{-1} X^T W Y \end{aligned} \quad (3.7)$$

Setelah mendapatkan  $\hat{\beta}$ , kita akan mendapatkan  $\hat{m}^*(x)$  dengan memasukkan nilai  $\hat{\beta}$  ke dalam persamaan berikut :

$$\hat{m}^*(x) = \hat{\beta}_0 + \hat{\beta}_1(x - x_0) + \dots + \hat{\beta}_p(x - x_0)^p \quad (3.8)$$

sehingga estimasi untuk model (3.1) diperoleh

$$\begin{aligned} \hat{m}(x_i) &= e^{\hat{m}^*(x_i)} \\ &= e^{\hat{\beta}_0 + \hat{\beta}_1(x_i - x_0) + \dots + \hat{\beta}_p(x_i - x_0)^p} \end{aligned} \quad (3.9)$$

### 3.2. Pohon *Gmelina Arborea Roxb.*

Salah satu hasil hutan yang ada di Indonesia adalah *Gmelina Arborea Roxb.*, tumbuhan ini berasal dari famili Verbenaceae. Pohon *Gmelina Arborea Roxb.* tumbuh sangat cepat, tingginya bisa mencapai 40 meter. Batangnya kurus dan permukaannya halus, dan berwarna abu-abu gelap, yang semakin lama akan berwarna coklat. Daunnya berbentuk seperti hati dan bunganya berwarna oranye dan kuning, dan menghasilkan madu. Kayu dari pohon ini memiliki banyak kegunaan, di

dunia industri kayu *Gmelina Arborea Roxb* digunakan untuk furniture, bahan untuk pengepakan, chipboard, kano, alat musik dan lain-lain. Dan jika dibandingkan dengan jenis kayu yang lain *Gmelina Arborea Roxb* sangat baik untuk industri kertas. Para pemeluk agama hindu juga biasa menggunakan akar, kulit batang dan buahnya untuk obat-obatan. Oleh karena itu peneliti tertarik untuk menggunakan data pohon *Gmelina Arborea Robx* di HTI Trans-Wanakasita Jambi (Lampiran 1) dengan diameter pohon (cm) sebagai variabel prediktor (X) dan volume pohon ( $m^3$ ) sebagai variabel respon (Y).

#### IV. METODE PENELITIAN

Penelitian yang diajukan terdiri atas dua kegiatan utama yaitu yang pertama membuat algoritma untuk estimasi robust LMS dan LTS berdasarkan estimator polynomial lokal yang terboboti fungsi kernel dan yang kedua mengimplementasikan algoritma program tersebut ke dalam *Software* Statistika R dalam menu Bahasa Indonesia, kemudian melanjutkan dengan menguji coba program tersebut pada data simulasi dan data real di laboratorium TPP. Langkah-langkah yang harus dilakukan dalam kegiatan tersebut adalah sebagai berikut :

1. membuat algoritma program untuk Robust LTS dan LMS
2. mendefinisikan fungsi-fungsi R terkait.
3. menggabungkan fungsi dalam satu skrip program
4. menguji coba program dengan data simulasi dan data real
5. menghubungkan program tersebut dengan suatu interface grafis (RGUI) dengan menambah program baru pada menu StatDemo.
6. menambah item sesuai kebutuhan dalam bahasa Indonesia.
7. mengemas program yang dihasilkan ke bentuk pustaka (*library*) yang mudah dipakai orang lain.

#### V. HASIL DAN PEMBAHASAN

Pada penelitian tahun ke-2 ini, TPP membahas pemodelan regresi nonparametrik dengan error lognormal pada kasus data yang terkontaminasi *outlier*

Pada saat *outlier* terdeteksi maka tidak seharusnya dilakukan estimasi ulang dengan tidak mengikutsertakan data *outlier* tersebut sebelum dilakukan kajian dari mana 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. Di sisi 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 menangani 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.

#### 5.1.1. Algoritma untuk menentukan nilai *bandwidth* optimal berdasarkan kriteria GCV

Secara rinci algoritma kriteria GCV untuk menentukan nilai *bandwidth* optimal diuraikan sebagai berikut :

##### Langkah 1.

Menginput data berpasangan  $\{x_i, y_i\}$  untuk  $i=1, 2, \dots, n$  dan mengurutkan data tersebut dari yang terkecil berdasarkan  $x_i$ .

##### Langkah 2.

Menentukan  $x_0 = x_a$ ;  $a = 2, 3, \dots, n-1$  yang berada pada range data.

##### Langkah 3.

Menentukan orde polinomial lokal  $p$  untuk mengestimasi fungsi regresi  $m(x_i)$

dengan estimator polinomial lokal yaitu  $\hat{m}(x_i) = \sum_{j=0}^p \beta_j (x_i - x_0)^j$ .

##### Langkah 4.

Mendefinisikan matriks  $y$  dan  $X$  sebagai berikut :

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \text{ dan } X = \begin{pmatrix} 1 & (x_1 - x_0) & \cdots & (x_1 - x_0)^p \\ \vdots & \vdots & & \vdots \\ 1 & (x_n - x_0) & \cdots & (x_n - x_0)^p \end{pmatrix} \quad (5.5)$$

**Langkah 5.**

Menentukan nilai  $h(x_0)$ ,  $h(x_0) = (x_{a+b} - x_{a-b})/2$  (5.6)

dengan  $x_{a\pm b}$  adalah nilai  $x$  disekitar  $x_0$

$$b = 1, 2, \dots, [(n-1)/2]$$

dan menentukan nilai GCV untuk setiap  $h$  dengan langkah-langkah sebagai berikut :

- i. Menghitung fungsi Kernel Gaussian

$$K_{h(x_0)}(x_i - x_0) = \frac{1}{h(x_0)} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{x_i - x_0}{h(x_0)} \right)^2}, \quad -\infty < (x_i - x_0) < \infty \quad (5.7)$$

- ii. Mendapatkan nilai matriks  $\mathbf{W}$  sesuai dengan persamaan berikut

$$\mathbf{W} = \begin{pmatrix} K_h(x_1 - x_0) & 0 & \dots & 0 \\ 0 & K_h(x_2 - x_0) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & K_h(x_n - x_0) \end{pmatrix} \quad (5.8)$$

- iii. Menentukan matriks  $\mathbf{A}(h(x_0))$  yang dinyatakan sebagai

$$\mathbf{A}(h(x_0)) = \mathbf{X}(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \quad (5.9)$$

- iv. Mendapatkan nilai  $MSE(h(x_0)) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{m}_{h(x_0)}(x_i))^2$  dengan

$$MSE(h(x_0)) = n^{-1} (\mathbf{y} - \mathbf{A}(h(x_0)) \mathbf{y})^T (\mathbf{y} - \mathbf{A}(h(x_0)) \mathbf{y}) \quad (5.10)$$

- v. Menghitung nilai GCV.

$$GCV(h(x_0)) = \frac{MSE(h(x_0))}{\left[ n^{-1} \text{tr}(\mathbf{I} - \mathbf{A}(h(x_0))) \right]^2} \quad (5.11)$$

**Langkah 6.**

Mendapatkan nilai  $h$  optimal yang didapatkan dari nilai GCV yang minimal.

**5.1.2. Algoritma Estimator Robust LMS**

Secara rinci algoritma estimator robust polinomial lokal LMS diuraikan sebagai berikut :

**Langkah 1.**

Mendefinisikan matriks  $\mathbf{y}$  dan  $\mathbf{X}$  sesuai dengan persamaan (5.5).

$$\text{dengan } S_{LMS} = c_{k,n} \left( 1 + \frac{5}{n-p} \right) (\hat{r}(\beta_{LMS})) \quad (5.20)$$

dengan konstanta  $c_{k,n}$  diperoleh dari persamaan

$$c_{k,n} = 1 / \Phi^{-1} \left( \frac{k+n}{2n} \right) \quad (5.21)$$

### Langkah 11.

Menghitung nilai koefisien determinasi ( $R^2$ ), yang mengukur berapa besar varians variabel respon dapat dijelaskan oleh model regresi yang dibentuk.

$$R^2_{LMS} = 1 - \frac{\left( (y_i - \hat{m}_{LMS}(x_i))^2 \right)_{k:n}}{\left( (y_i - \bar{y})^2 \right)_{k:n}}$$

#### 5.2.1. Estimator Robust LTS

Estimator LTS adalah metode estimasi kurva regresi yang meminimumkan kriteria :

$$\hat{\beta}_{LTS} = \min_{\beta} \sum_{i=1}^k (r^2(\beta))_{i:n} \quad (5.22)$$

dengan  $r^2(\beta)_{i:n}$  adalah kuadrat *residual* ke- $i$  yang diurutkan dari nilai paling kecil hingga paling besar  $(r^2)_{1:n} \leq \dots \leq (r^2)_{n:n}$  dan  $r_i^2 = (y_i - \hat{m}_h(x_i))^2 K_h(x_i - x_0)$ , untuk  $i = 1, 2, \dots, n$ .  $k$  adalah konstanta *trimming* yang memenuhi  $n/2 < k < n$ , menyatakan ke-robust-an estimator LTS.

Nilai  $k$  sekaligus menunjukkan jumlah data yang digunakan untuk mengestimasi kurva regresi, dimana nilai  $k$  berada pada interval dan dapat dipilih untuk menghasilkan tingkat kerobust-an tinggi dan disebut nilai *breakdown*  $\varepsilon_n^*$ , yaitu

$$\left\lfloor \frac{n}{2} \right\rfloor + 1 \leq k \leq \left\lfloor \frac{(n + (p+1) + 1)}{2} \right\rfloor, \text{ dengan } p \geq 1 \quad (5.23)$$

Sedangkan untuk menghasilkan nilai *breakdown* menggunakan persamaan :

$$\varepsilon_n^* = \frac{(n - k + 1)}{n} \quad (5.24)$$

Beberapa persyaratan dalam penentuan nilai  $k$  tersebut adalah :





- a. Nilainya terletak diantara :  $\left[ \frac{n}{2} \right] \leq k \leq n$
- b. Nilai  $[x]$  adalah bilangan bulat terbesar yang lebih kecil atau sama dengan  $x$ .

### 5.2.2. Algoritma Estimator Robust LTS

Proses perhitungan untuk mendapatkan hasil estimasi dari estimator robust LTS tidak mempunyai bentuk baku (*close-form*), sehingga didekati dengan suatu algoritma. Algoritma yang digunakan dalam penelitian ini disebut sebagai *basic resampling algorithm*, seperti pertama kali diperkenalkan **Rousseeuw dan Leroy, (1987)** dalam *PROGRESS (Program for Robust Regression)* dan diperbaiki kecepatan untuk perhitungan sampel berukuran besar oleh **Rousseeuw dan Van Driessen, (1999)** dalam *FAST-LTS*. Secara rinci algoritma estimator robust polinomial lokal LTS diuraikan sebagai berikut :

#### Langkah 1.

Mendefinisikan matriks  $y$  dan  $X$  sesuai dengan persamaan (5.5).

#### Langkah 2.

Mengambil sebanyak  $m$  subset berdasarkan kombinasi  $(p+1)$  dari  $n$ .

$$m = \binom{n}{p+1} = \frac{n!}{(p+1)!(n-p-1)!} \quad (5.25)$$

#### Langkah 3.

Menentukan  $x_{0c}$  yang berada pada range subset  $J_c$ ;  $c = 1, 2, \dots, m$ .

#### Langkah 4.

Mendefinisikan matriks  $y_c$  dan  $X_c$  sesuai dengan persamaan (5.13)

#### Langkah 5.

Memasukkan nilai *bandwidth* yang optimal ke dalam fungsi bobot  $W_c$  untuk setiap subset  $J_c$ .

#### Langkah 6.

Mendefinisikan nilai  $k$  yang berada pada interval untuk menghasilkan tingkat ke-robust-an tinggi, dalam penelitian ini digunakan

$$\left[ \frac{n}{2} \right] + 1 \leq k \leq \left[ \frac{(n+(p+1)+1)}{2} \right] \quad (5.26)$$

**Langkah 7.**

Mengevaluasi model regresi yang menggunakan intersep melalui proses *adjustment* untuk setiap subset  $J_c$  dengan langkah berikut :

- iv. Menghitung  $\hat{\beta}_c$  untuk setiap subset  $J_c$  yang dinyatakan oleh

$$\hat{\beta}_c = (\mathbf{X}_c^T \mathbf{W}_c \mathbf{X}_c)^{-1} \mathbf{X}_c^T \mathbf{W}_c \mathbf{y}_c \quad (5.27)$$

- v. Menghitung nilai  $t_i$  untuk setiap subset  $J_c$  tanpa menggunakan nilai estimasi intersep.

$$\begin{aligned} t_i &= y_i - \hat{y}_i \\ &= y_i - \hat{\beta}_1(x_i - x_0) - \dots - \hat{\beta}_p(x_i - x_0)^p, \text{ untuk } i=1, 2, \dots, n \end{aligned}$$

- vi. Mengurutkan nilai  $t_i$  dari kecil ke besar sebagai berikut

$$t_1 \leq t_2 \leq \dots \leq t_n$$

- iv. Membentuk kelas interval dari nilai-nilai  $t_i$  yang telah diurutkan yang masing-masing berisi  $k$  pengamatan sehingga diperoleh sebanyak  $(n-k+1)$  kelas interval.

- v. Menghitung nilai intersep baru  $\hat{\beta}_0 = \bar{t}$  yang diperoleh dari nilai rata-rata kelas interval dengan jumlah kuadrat simpangan terkecil.

$$\min \sum_{i=1}^k (t_i - \bar{t})^2 \quad (5.28)$$

- vii. Menyatakan estimasi kurva regresi dengan nilai intersep baru sehingga didapat model regresi

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1(x_i - x_0) + \dots + \hat{\beta}_p(x_i - x_0)^p \quad (5.29)$$

**Langkah 8.**

Menghitung fungsi tujuan (FT) untuk setiap subset  $J_c$  yang dinyatakan sebagai

$$FT_{LTS} = \sqrt{\frac{1}{k} \sum_{i=1}^k r^2(\hat{\beta})_{i:n}} \quad (5.30)$$

dengan

$$r(\hat{\beta})_i = y_i - \hat{y}_i$$

$r^2(\hat{\beta})_{i:n}$  adalah nilai kuadrat *residual* yang telah diurutkan untuk  $i = 1, 2, \dots, n$ .

**Langkah 9.**

Menentukan estimator LTS ( $\hat{\beta}_{LTS}$ ) berdasarkan nilai FT minimum untuk setiap subset  $J_c$ ;  $c = 1, 2, \dots, m$ .

**Langkah 10.**

Menghitung nilai *preliminary scale estimate* ( $s_{LTS}$ ), yaitu nilai awal estimasi melalui persamaan berikut :

$$s_{LTS} = d_{k,n} \sqrt{\frac{1}{k} \sum_{i=1}^k (r^2(\hat{\beta}_{LTS}))_{in}} \quad (5.31)$$

dengan konstanta  $c_{k,n}$  dan  $d_{k,n}$  diperoleh dari persamaan berikut :

$$c_{k,n} = 1/\Phi^{-1}\left(\frac{k+n}{2n}\right)$$

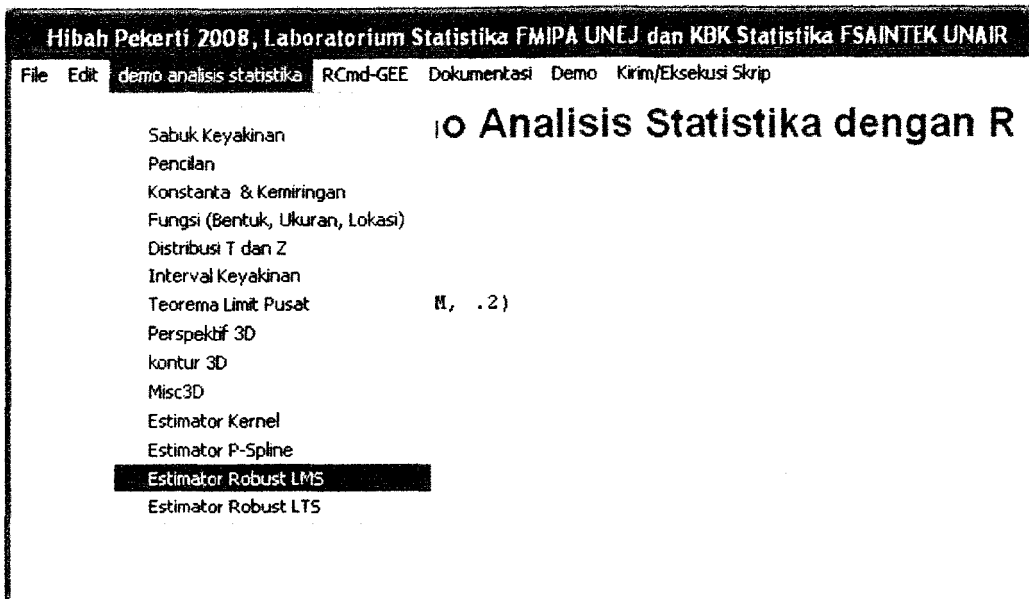
$$d_{k,n} = 1/\sqrt{1 - \frac{2n}{kc_{k,n}} \phi(1/c_{k,n})} \quad (5.32)$$

**Langkah 11.**

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

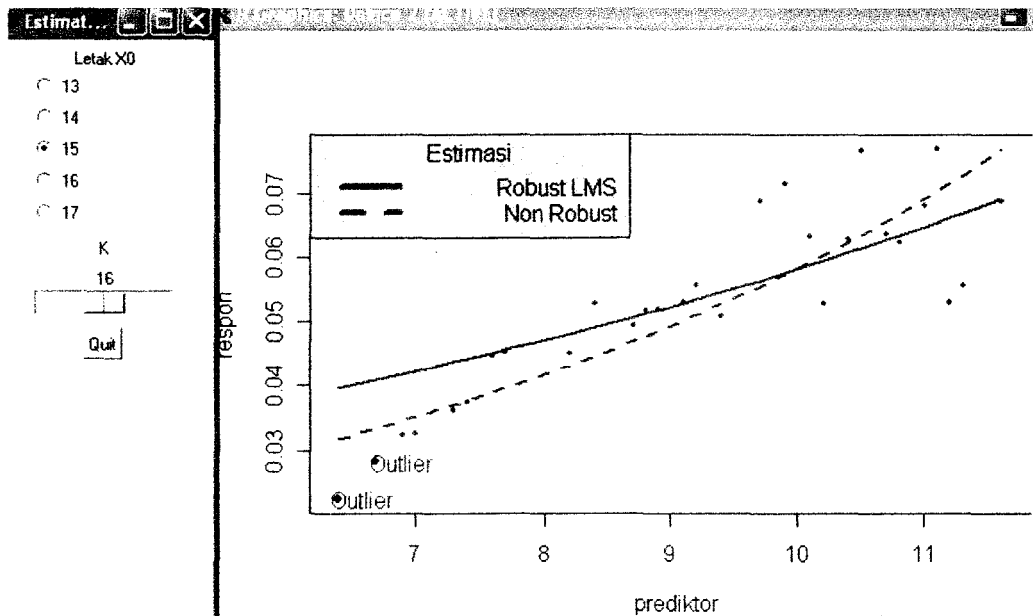
$$\left| \frac{r(\hat{\beta}_{LTS})}{s_{LTS}} \right| > 2,5 \quad (5.33)$$

Berdasarkan algoritma program yang telah dibuat, TPP mendefinisikan fungsi-fungsi R yang terkait dan menggabungkan fungsi tersebut dalam satu skrip program. Hal tersebut dilakukan di laboratorium TPP. Setelah berhasil menguji program R dengan data simulasi maupun data riil, kemudian menghubungkan program tersebut dengan suatu *interface* grafis (RGUI) dengan program baru dengan menu StatDemo. Pada program menu StatDemo, data yang digunakan adalah data riil yaitu data Pohon *Gmelina Arborea Roxb* di areal HTI-Trans PT Wanakasita Nusantara Jambi. Cara menjalankan program tersebut adalah dari jendela R-Console, dipanggil terlebih dahulu library (StatDemo), setelah itu ketik StatDemo(), hasil dari *running* StatDemo seperti terlihat pada Gambar 1 berikut ini :



**Gambar 1. Demo Analisis Statistika dengan R untuk Robust LMS**

Dari jendela demo analisis statistika R seperti tampak pada Gambar 1, jika di klik menu demo analisis statistika kemudian di klik Estimator Robust LMS akan mendapatkan hasil sebagai berikut:



**Gambar 2. Hasil Estimasi berdasarkan Robust LMS dengan  $K=16$ , posisi  $X_0$  data ke-15**

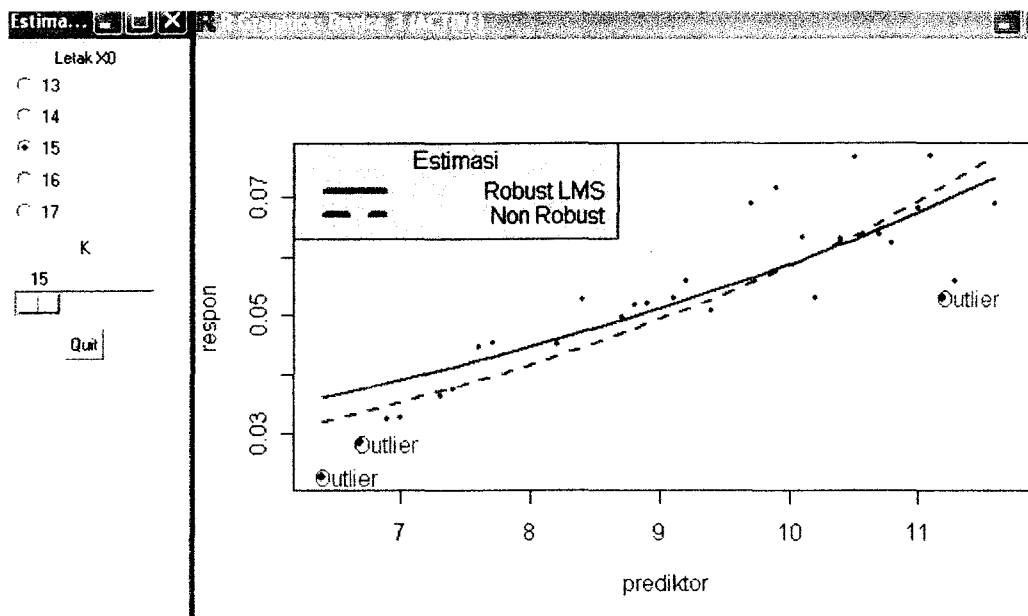
Berdasarkan Gambar 2, terlihat hasil estimasi robust LMS pada data pohon *Gmelina Arborea Roxb* dapat mendeteksi dua outlier. Pada estimator Robust LMS

terlihat tidak terpengaruh adanya dua *outlier* tersebut jika dibandingkan dengan estimator yang bersifat nonrobust. Hasil estimasi model berdasarkan robust LMS dengan  $K=16$  *bandwidth* optimal  $h = 2.6$  dan  $x_0 = 9,2$  yaitu :

$$m(x) = \exp(-2,927 + 0,107(x - 9.2)) \quad (5.34)$$

dengan nilai MSE adalah 0,031.

Jika scroll pada konstanta K digerakkan pada angka 15 Gambar 2 akan berubah plot estimasinya seperti tampak pada Gambar 3 berikut ini :

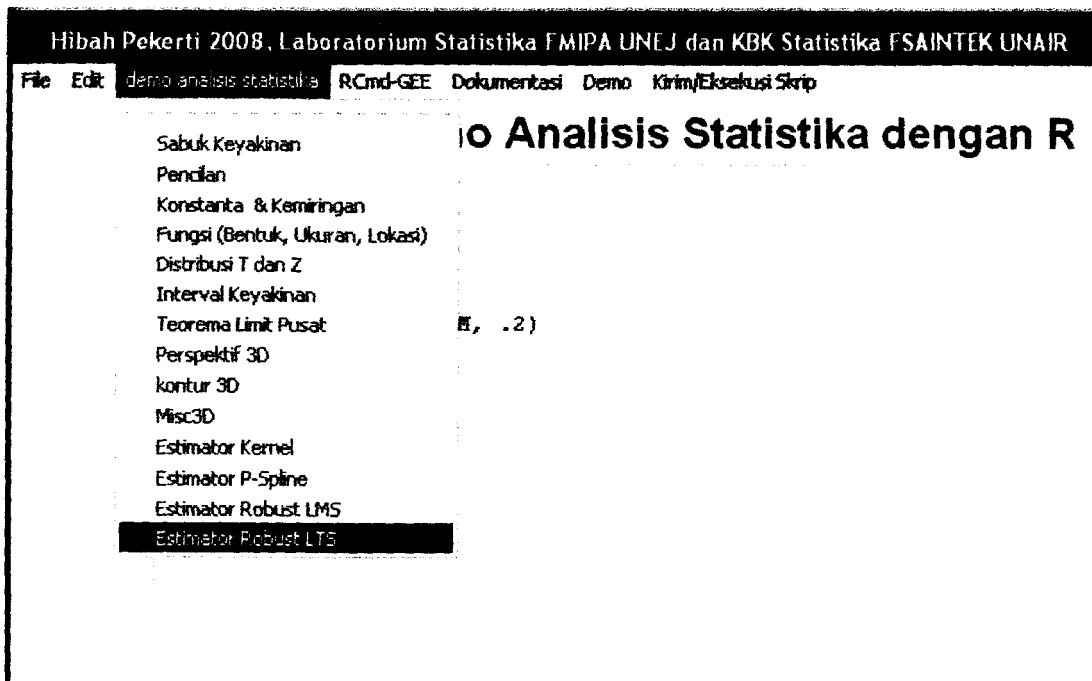


**Gambar 3.** Hasil Estimasi berdasarkan Robust LMS dengan  $K=15$ , posisi  $X_0$  data ke-15 Pada Gambar 3, dengan  $K$  sama dengan 15 terdeteksi 3 *outlier* pada posisi  $X_0$  dan *bandwidth* yang sama. Hasil estimasi model berdasarkan robust LMS dengan  $K=15$  *bandwidth* optimal  $h = 2,6$  dan  $x_0 = 9,2$  yaitu :

$$m(x) = \exp(-2,944 + 0,136(x - 9.2)) \quad (5.35)$$

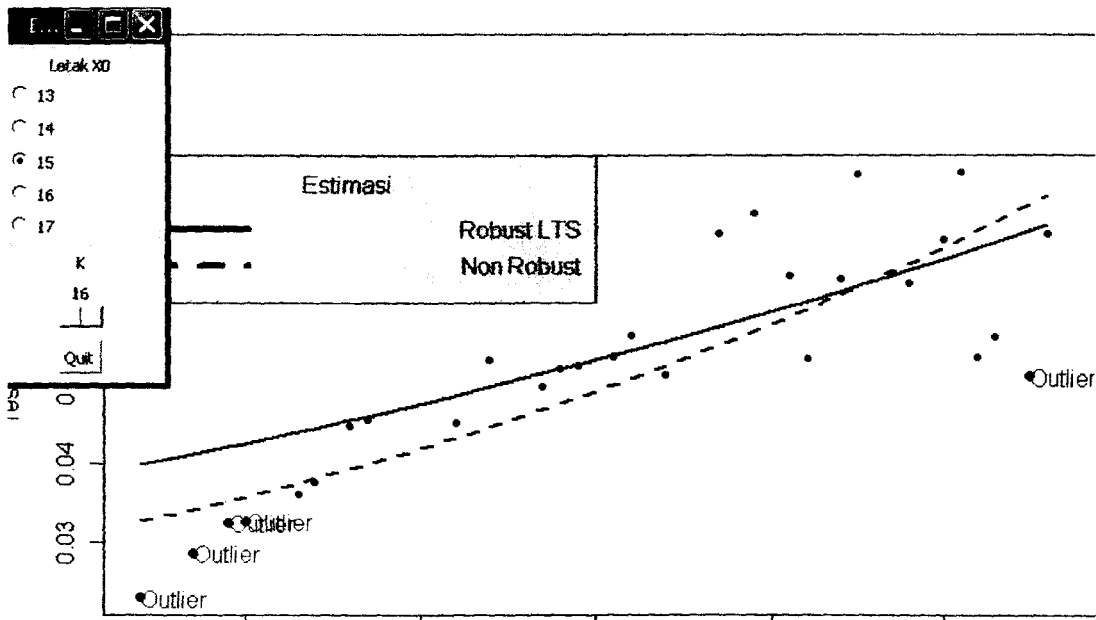
dengan nilai MSE adalah 0,023. Sesuai dengan persamaan (5.15) semakin kecil nilai  $K$  maka semakin besar nilai *breakdown value* sehingga semakin banyak *outlier* yang terdeteksi.

Implementasi OSS-R untuk estimator Robust LTS pada data pohon Gmelina Arborea Roxb ditunjukkan sebagai berikut : dari jendela demo analisis statistika R seperti tampak pada Gambar 4, jika di klik menu demo analisis statistika kemudian di klik Estimator Robust LTS akan mendapatkan hasil seperti Gambar 5 berikut ini:



**Gambar 4. Demo Analisis Statistika dengan R untuk Robust LTS**

Dari jendela demo analisis statistika R seperti tampak pada Gambar 4, jika di klik menu demo analisis statistika kemudian di klik Estimator Robust LTS akan mendapatkan hasil sebagai berikut:



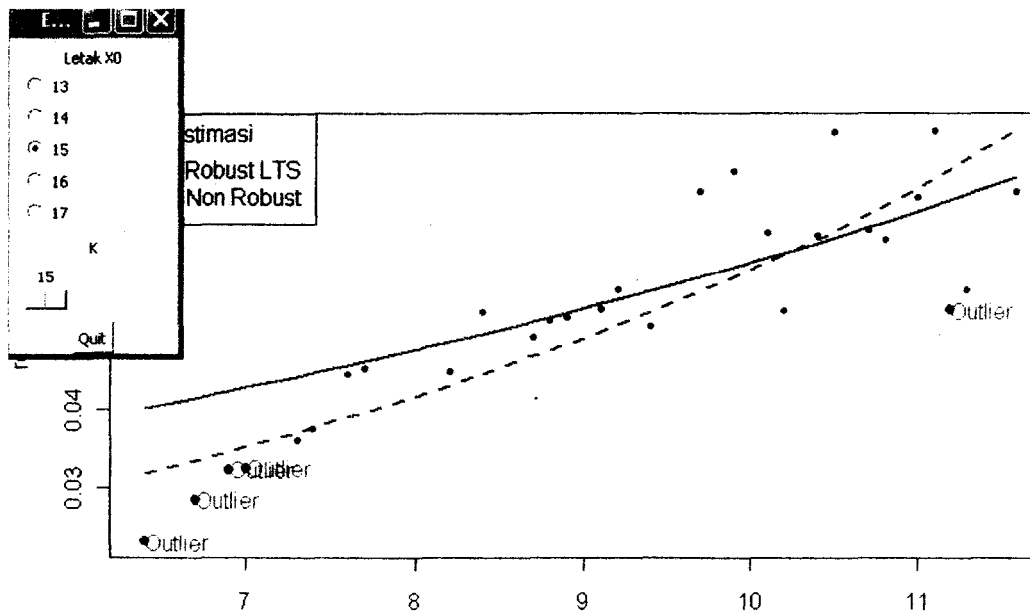
**Gambar 5. Hasil Estimasi berdasarkan Robust LTS dengan  $K=16$ , posisi  $X_0$  data ke-15**

Berdasarkan Gambar 5, terlihat bahwa dengan estimator Robust LTS untuk data yang sama yaitu data pohon *Gmelina Arborea Roxb* dapat mendeteksi 5 outlier. Hasil estimasi model berdasarkan estimator robust LTS dengan  $K=16$  bandwidth optimal  $h = 2,6$  dan  $x_0 = 9,2$  yaitu :

$$m(x) = \exp(-2,907 + 0,113(x - 9,2)) \tag{5.36}$$

dengan nilai MSE adalah 0,032

Apabila scroll pada konstanta K digerakkan pada angka 15 Gambar 5 akan berubah plot estimasinya seperti tampak pada Gambar 6 berikut ini :



**Gambar 6. Hasil Estimasi berdasarkan Estimator Robust LTS dengan  $K=15$ ,  $X_0=15$**

Pada Gambar 5, dengan K sama dengan 15 terdeteksi 5 outlier pada posisi  $X_0$  dan bandwidth yang sama. Hasil estimasi model berdasarkan estimator robust LTS dengan  $K=15$  bandwidth optimal  $h = 2,6$  dan  $x_0 = 9,2$  adalah:

$$m(x) = \exp(-2,910 + 0,108(x - 9.2)) \tag{5.37}$$

dengan nilai MSE adalah 0,033.

Berdasarkan hasil aplikasi estimator robust LMS pada data pohon *Gmelina Arborea Roxb* diperoleh hasil terbaik dengan pemilihan nilai K sesuai dengan persamaan (5.14) yaitu kemampuan mendeteksi outlier sebanyak 3 dengan nilai MSE adalah 0,023 sedangkan untuk estimator robust LTS diperoleh hasil yaitu kemampuan mendeteksi outlier sebanyak 5 outlier dengan nilai MSE adalah 0,032



dan LTS pada data pohon *Gmelina Arborea Roxb* menunjukkan bahwa metode robust LTS memiliki kemampuan mendeteksi *outlier* lebih banyak dari pada metode robust LMS. Dari hasil pengujian asumsi error berdistribusi lognormal dengan uji Kolmogorov-Smirnov diperoleh nilai *p-value* adalah 0,746 sehingga disimpulkan bahwa error berdistribusi lognormal.

## VI. KESIMPULAN DAN SARAN

### 6.1 Kesimpulan

Pada penelitian tahun ke-2 ini menghasilkan peningkatan kemampuan TPP dalam hal pemrograman *Software R* yang berbasis OSS dengan menu Bahasa Indonesia yang terkait dengan pendeteksian *outlier* menggunakan metode robust LMS dan robust LTS berdasarkan estimator polinomial yang terboboti oleh fungsi kernel. Dari hasil implementasi OSS-R untuk pendeteksian *outlier* dengan metode robust LMS dan LTS pada data pohon *Gmelina Arborea Roxb* diperoleh hasil bahwa estimator robust LTS dapat mendeteksi lebih banyak *outlier* daripada estimator robust LMS.

Secara keseluruhan hasil penelitian pada tahun pertama dan kedua adalah telah dikembangkannya *Software Statistika R* yang diimplementasikan pada pemodelan regresi nonparametrik dengan estimator kernel dan estimator spline yang terintegrasi dengan *interface* grafik (RGUI) dengan menu Bahasa Indonesia, dan demo analisis statistika dengan program R interaktif yang telah dihasilkan, sangat bermanfaat dalam pembelajaran untuk meningkatkan pemahaman mahasiswa khususnya terhadap teori pemodelan regresi nonparametrik dan pendeteksian *outlier* dalam data.

### 6.2 Saran

Hasil penelitian yang terkait dengan demo analisis statistika dengan program R dapat digunakan dalam pembelajaran yang bersifat interaktif seperti simulasi dan statistika matematika dengan melakukan modifikasi dari program R yang telah dihasilkan.

**Lampiran 1. Data Pohon *Gmelina Arborea Roxb* (HTI Trans Wanakasita Jambi, 2004)**

No	Diameter (cm)	Volume (m3)
1	6.4	0.007156
2	6.7	0.012281
3	6.9	0.010266
4	7	0.01194
5	7.3	0.015525
6	7.4	0.01762
7	7.6	0.012433
8	7.7	0.01141
9	8.2	0.019855
10	8.4	0.021992
11	8.7	0.034299
12	8.8	0.023748
13	8.9	0.031418
14	9.1	0.022852
15	9.2	0.028466
16	9.4	0.027148
17	9.7	0.032589
18	9.9	0.036139
19	10.1	0.037491
20	10.2	0.044531
21	10.4	0.045295
22	10.5	0.044994
23	10.7	0.037567
24	10.8	0.049525
25	11	0.051771
26	11.1	0.04897
27	11.2	0.053097
28	11.3	0.055808
29	11.6	0.068749

## LAMPIRAN 2. Program Robust LMS

```

inverse<-function(X,eps=1e-006)
{
  X<-as.matrix(X)
  Xsvd<-svd(X)
  diago<-Xsvd$d[Xsvd$d>eps]
  if(length(diago)==1)
  {Xplus<-as.matrix(Xsvd$v[,1])%*%t(as.matrix(Xsvd$u[,1])/diago)}
  else {
Xplus<-svd$v[,1:length(diago)]%*%diag(1/diago)%*%t(Xsvd$u[,1:length(diago)])
  }
  return(Xplus)
}
LMS<-function(data,h,m)
{
  z<-data[,1]
  y<-data[,2]
  n<-length(y)
  expy<-exp(y)
  p<-2
  M<-matrix(0,n,2)
  M[,1]<-y
  M[,2]<-z
  comr<-choose(n,p)
  myhat<-matrix(0,n,comr)
  mbeta<-matrix(0,1,comr)
  vintersep<-rep(0,comr)
  vft<-rep(0,comr)
  me<-matrix(0,n,comr)
  x0<-z[m]
  x<-matrix(0,n,p)
  for(i in 1:n){
    for(j in 1:p){
      x[i,j]<-(z[i]-x0)^(j-1)
    }
  }
  datax<-matrix(0,n,1)
  datax[,1]<-x[,2]
  kh1<-(1/sqrt(2 * pi)) * exp(-(((z - x0)/h)^2)/2)
  w1<-diag(1, n, n) * (kh1/h)
  sn<-t(x)%*%w1%*%x
  tn<-t(x)%*%w1%*%y
  beta <- solve(sn)%*%tn
  ytopi<-x%*%beta
  expytopi<-exp(ytopi)
  bb<-(n/2)+1
}

```

```

ba<-(n+p+1)/2
v<-0
cat("interval k:",bb,"<=k<=",ba,"\n")
k<-as.numeric(readline("pilih nilai k: "))
for(i in 1:(n-1))
  for(j in (i+1):n){v<-v+1
    if((v%%20)==0)cat("iterasi ke-",v,"\n")
    xy<-M[c(i,j),]
    y1<-as.vector(xy[,-2])
    z1<-as.vector(xy[,-1])
    x1<-matrix(0,p,p)
    for(i in 1:p){ for(j in 1:p){
      x1[i,j]<-(z1[i]-x0)^(j-1) }}
    kh<-(1/sqrt(2*pi))*exp(-(((z1-x0)/h)^2)/2)
    w<-diag(1,p,p)*(kh/h)
    f<-t(x1)%*%w%*%x1
    ap<-inverse(f,eps=1e-006)
    at<-ap%*%t(x1)%*%w%*%y1
    as<-matrix(at[2])
    yest<-datax%*%as
    ti<-(y-yest)
    ti<-as.vector(ti)
    a<-matrix(sort(ti),n,1)
    bv<-(n-k+1)/n
    ab<-matrix(rep(0,n-k+1),n-k+1,1)
    for(i in 1:(n-k+1)){
      ab[i]<-(a[i+k-1,1]-a[i]) }
    abmin<-min(ab[1:(n-k+1),1])
    for(i in 1:(n-k+1)){
      if(ab[i]==abmin){ intersep<-0.5*(a[i]+a[i+k-1,1])
        yhat<-yest+matrix(rep(0.5*(a[i]+a[i+k-1,1]),n),n,1)
          e<-(y-yhat)
            g<-e^2
              r<-sort(g) } }
        ft<-sqrt(r[k])
        dl<-(n+k)/(2*n)
        ckn<-1/qnorm(dl)
        sstar<-ckn*(1+(5/(n-p)))*ft
        for(i in 1:n){
          e[i,1]<-e[i,1]/sstar }
        myhat[,v]<-yhat
        vintersep[v]<-intersep
        mbeta[,v]<-as
        vft[v]<-ft
        me[,v]<-e
          }
    minft<-vft[1]
    for(g in 2:v){ if(minft > vft[g]){ minft<-vft[g]}

```

```

mn<-g }}
ftakhir<-minft
yhataakhir<-as.vector(myhat[,mn])
err<-(y-yhataakhir)
mse<-sum(err*err/n)
cat("\n Nilai Mean Square Error (MSE) = ", format(mse),"\n")
cat("\n Nilai Fungsi Tujuan (FT)minimum - ", format(ftakhir),"\n")
expyhataakhir<-exp(yhataakhir)
intersepakhir<-vintersep[mn]
betaakhir<-mbeta[,mn]
print(intersepakhir)
print(betaakhir)
eakhir<-me[,mn]
wi<-rep(1,n)
estdr<-abs(eakhir)
for(i in 1:n){if(estdr[i]>-2.5) wi[i]<-0 }
M<-cbind(wi,y,z)
mout<-as.matrix(M[M[,1]==0,])
mreg<-M[M[,1]==1,]
if(ncol(mout)==1){
outy<-mout[2,]
outz<-mout[3,] }
else {outy<-as.vector(mout[,2])
outz<-as.vector(mout[,3])}
regy<-as.vector(mreg[,2])
expregy<-exp(regy)
regz<-as.vector(mreg[,3])
outy1<-as.vector(mout[,2])
expouty<-exp(outy1)
print(mout)
print(mreg)
win.graph()
plot(z,expyhataakhir,type="l",col="blue",lwd=2,xlim=c(min(z),max(z)+0.1),ylim=c(
min(expy),max(expy)),ylab="respon",xlab="prediktor")
par(new=T)
plot(regz,expregy,type="p",pch=20,col="blue",xlim=c(min(z),max(z)+0.1),ylim=c(
min(expy),max(expy)),ylab="respon",xlab="prediktor") par(new=T)
plot(outz,expouty,type="p",pch=19,lty=20,col="red",xlim=c(min(z),max(z)+0.1),yli
mi=c(min(expy),max(expy)),ylab="respon",xlab="prediktor")
out1<-outz+0.2
text(out1,expouty,"Outlier",col="red")
lines(z,expytopi,lty=2,lwd=2,col="blue",xlim=c(min(z),max(z)+0.1),ylim=c(min(exp
y),max(expy)),ylab="respon",xlab="prediktor")
temp <- legend("topleft", legend = c(" ", " "),text.width = strwidth("Non Robust"),
lty = 1:2, xjust = 1, yjust = 1,col="blue",lwd=4,bg="grey", title = "Estimasi")
text(temp$rect$left + temp$rect$w, temp$text$y,c("Robust LMS", "Non Robust"),
pos=2) }

```

**LAMPIRAN 3. Program Robust LTS**

```

inverse<-function(X,eps=1e-006)
{
  X<-as.matrix(X)
  Xsvd<-svd(X)
  diago<-Xsvd$d[>eps] if(length(diago)==1)
  {Xplus<-as.matrix(Xsvd$v[,1])%*%t(as.matrix(Xsvd$u[,1])/diago)}
else{Xplus<svd$v[,1:length(diago)]%*%diag(1/diago)%*%t(Xsvd$u[,1:length
(diago)]) }return(Xplus)}
LTS<-function(data,h,m)
{
  z<-data[,1]
  y<-data[,2]
  n<-length(y)
  expy<-exp(y)
  p<-2
  M<-matrix(0,n,2)
  M[,1]<-y
  M[,2]<-z
  comr<-choose(n,p)
  myhat<-matrix(0,n,comr)
  mbeta<-matrix(0,1,comr)
  vintersep<-rep(0,comr)
  vft<-rep(0,comr)
  me<-matrix(0,n,comr)
  x0<-z[m]
  x<-matrix(0,n,p)
  for(i in 1:n){ for(j in 1:p){ x[i,j]<-(z[i]-x0)^(j-1) } }
  datax<-matrix(0,n,1)
  datax[,1]<-x[,2]
  kh1<-(1/sqrt(2 * pi)) * exp(-(((z - x0)/h)^2)/2)
  w1<- diag(1, n, n) * (kh1/h)
  sn<-t(x)%*%w1%*%x
  tn<-t(x)%*%w1%*%y
  beta <- solve(sn)%*%tn
  ytopi<-x%*%beta
  expytopi<-exp(ytopi)
  bb<-(n/2)+1
  ba<-(n+p+1)/2
  v<-0
  cat("interval k:",bb,"<=k<=",ba,"n")
  k<-as.numeric(readline("pilih nilai k: "))
  for(i in 1:(n-1)) for(j in (i+1):n){v<-v+1
    if((v%20)==0)cat("iterasi ke-",v,"n")
    xy<-M[c(i,j),]
    y1<-as.vector(xy[-2])
  }
}

```

```

z1<-as.vector(xy[,-1])
x01<-z1[2]
x1<-matrix(0,p,p)
for(i in 1:p){for(j in 1:p){
  x1[i,j]<-(z1[i]-x01)^(j-1) }
}
kh<-(1/sqrt(2 * pi)) * exp(-(((z1 - x01)/h)^2)/2)
w<-diag(1,p,p)*(kh/h)
f<-t(x1)%*%w)%*%x1
ap<-inverse(f,eps=1e-006)
at<-ap %*% (t(x1) %*% w)%*%y1)
as<-matrix(at[2])
yest<-datax)%*%as
ti<-(y-yest)
ti<-as.vector(ti)
a<-matrix(sort(ti),n,1)
bv<-(n-k+1)/n
tbar<-matrix(rep(0,n-k+1),n-k+1,1)
Q<-matrix(rep(0,n-k+1),n-k+1,1)
tbar[1,1]<-mean(a[1:k],1)
for(j in 1:n){Q[1,1]<-Q[1,1]+(a[j,1]-tbar[1,1])^2 }
for(i in 2:(n-k+1)){tbar[i,1]<-(k*tbar[i-1,1]-a[i-1,1]+a[i+k-1,1])/k
Q[i,1]<-Q[i-1,1]+(a[i-1,1])^2+(a[i+k-1,1])^2-k*(tbar[i,1])^2+k*(tbar[i-1,1])^2 }
qmin<-min(Q[1:(n-k+1),1])
for(i in 1:(n-k+1)){if(Q[i,1]==qmin){intersep<-tbar[i,1]
yhat<-yest+matrix(rep(tbar[i,1]),n,1)
e<-(y-yhat)
g<-e^2
r<-sort(g) } }
s<-sum(r[1:k])
ft<-sqrt(s/k)
dl<-(n+k)/(2*n)
ckn<-1/qnorm(dl)
dalam<-(2*n)/(k*ckn)
d<-1/sqrt(1-(dalam*dnorm(1/ckn)))
sd<-d*ft
for(i in 1:n){e[i,1]<-e[i,1]/sd }
myhat[,v]<-yhat
mbeta[,v]<-as
vintersep[v]<-intersep
vft[v]<-ft
me[,v]<-e
minft<-vft[1]
for(g in 2:v){ if(minft > vft[g]){minft<-vft[g] mn<-g } }
yhataakhir<-as.vector(myhat[,mn])
err<-(y-yhataakhir)
mse<-sum((err*err)/n)
print(mse)

```



```

cat("\n Nilai Fungsi Tujuan = ", format(ftakhir), "\n")
expyhataakhir<-exp(yhataakhir)
  eakhir<-me[,mn]
  rr<-sort(eakhir^2)
  atas<-sum(rr[1:k])
  ybar<-mean(y)
  yy<-sort((y-ybar)^2)
  bawah<-sum(yy[1:k])
  error<-abs(eakhir)
  betaakhir<-mbeta[,mn]
  intersepakhir<-vintersep[mn]
  print(betaakhir)
  print(intersepakhir)
  ftakhir<-minft
  wi<-rep(1,n)
  stdr<-abs(eakhir)
  for(i in 1:n){if(stdr[i]>2.5) wi[i]<-0}
  M<-cbind(wi,y,z)
  mout<-as.matrix(M[M[,1]==0,])
  mreg<-M[M[,1]==1,]
  if(ncol(mout)==1){outy<-mout[2,]
  outz<-mout[3,] }
  else {outy<-as.vector(mout[,2])
  outz<-as.vector(mout[,3])}
  regy<-as.vector(mreg[,2])
  expregy<-exp(regy)
  regz<-as.vector(mreg[,3])
  outy1<-as.vector(mout[,2])
  expouty<-exp(outy1)
  print(mout)
  print(mreg)
plot(z,expyhataakhir,type="l",col="blue",lwd=2,xlim=c(min(z),max(z)+0.1),ylim=c(
min(expy),max(expy)),ylab="respon",xlab="prediktor")
par(new=T)
plot(regz,expregy,type="p",pch=20,col="blue",xlim=c(min(z),max(z)+0.1),ylim=c(
min(expy),max(expy)),ylab="respon",xlab="prediktor")
par(new=T)
plot(outz,expouty,type="p",pch=19,lty=20,col="red",xlim=c(min(z),max(z)+0.1),yli
m=c(min(expy),max(expy)),ylab="respon",xlab="prediktor")
  out1<-outz+0.2
text(out1,expouty,"Outlier",col="red")
lines(z,expytopi,lty=2,lwd=2,col="blue",xlim=c(min(z),max(z)+0.1),ylim=c(min(exp
y),max(expy)),ylab="respon",xlab="prediktor")
temp <- legend("topleft", legend = c(" ", " "),text.width = strwidth("Non Robust"),
lty = 1:2, xjust = 1, yjust = 1,col="blue",lwd=4,bg="grey",title = "Estimasi")
text(temp$rect$left + temp$rect$w, temp$text$y,c("Robust LTS", "Non Robust"),
pos=2) }

```

**Lampiran 4. Program R untuk Demo Analisis Statistika R**

```

StatDemo<-function(){
require(tcltk)
ju <- tktoplevel()
teks <- tktext(ju,height=30)
tkwm.title(ju,"Hibah Pekerti 2008, Nur Chamidah(FSAINTEK UNAIR) dan I Made
Tirta (FMIPA UNEJ)")
fontTextLabel<-tkfont.create(family="sans",size=18,weight="bold")
fontFixedWidth <- tkfont.create(family="courier",size=18)
tkpack(kb1 <- tklabel(ju,text="Demo Analisis Statistika dengan R" ,fg="blue",
font=fontTextLabel))
tkinsert(teks, "0.0", "\nrequire(Rgraphviz)\nset.seed(123)\nV <- letters[1:10]\nM <-
1:4\nG1 <- randomGraph(V, M, .2)\nplot(g1)")

source('D:/Penelitian1/Hibah-Pekerti/StatDemo/R/Clt.r')
source('D:/Penelitian1/Hibah-Pekerti/StatDemo/R/SabukAnim.r')
source('D:/Penelitian1/Hibah-Pekerti/StatDemo/R/FsAnim.r')
source('D:/Penelitian1/Hibah-Pekerti/StatDemo/R/kontr.r')
source('D:/Penelitian1/Hibah-Pekerti/StatDemo/R/p3d.r')
source('D:/Penelitian1/Hibah-Pekerti/StatDemo/R/p3dg.r')
source('D:/Penelitian1/Hibah-Pekerti/StatDemo/R/PencarAnim.r')
source('D:/Penelitian1/Hibah-Pekerti/StatDemo/R/ScatAnim.r')
source('D:/Penelitian1/Hibah-Pekerti/StatDemo/R/tik.r')
source('D:/Penelitian1/Hibah-Pekerti/StatDemo/R/tnorm.r')
source('D:/Penelitian1/Hibah-Pekerti/StatDemo/R/estimasi-kernel.R')
source('D:/Penelitian1/Hibah-Pekerti/StatDemo/R/pspline2.r')
source('D:/Penelitian1/Hibah-Pekerti/StatDemo/R/robustLTS')
source('D:/Penelitian1/Hibah-Pekerti/StatDemo/R/robustLMS')

buka.file<-function(){
file <- tclvalue(tkgetOpenFile())
chn <- tkopen(file, "r")
tkinsert(teks, "0.0", tclvalue(tkread(chn)))
wfile <<- file }
simpan.file<-function(){
file <- tclvalue(tkgetSaveFile(
initialfile=tclvalue(tkfile.tail(wfile)),
initialdir=tclvalue(tkfile.dir(wfile))))
if (!length(file)) return()
chn <- tkopen(file, "w")
tkputs(chn, tclvalue(tkget(txt,"0.0", "end")))
tkclose(chn)
wfile <<- file}
topMenu <- tkmenu(ju)
tkconfigure(ju,menu=topMenu)

```

```

Menu1 <- tkmenu(topMenu,tearoff=FALSE)
tkadd(Menu1,"command",label="Buka Skrip",command=buka.file)
tkadd(Menu1,"command",label="Simpan Skrip",command=function()
tkgetSaveFile())
tkpack(teks)
tkadd(topMenu,"cascade",label="File",menu=Menu1)
hlp.txt <- function() library(RcmdrGEE)
eval.txt <- function(){eval(parse(text=tclvalue(tkget(teks, "0.0", "end")))) }
salin <- function(){
  focused <- tkfocus()
  selection <- strsplit(tclvalue(tktag.ranges(focused, "sel")), " ")[[1]]
  if (is.na(selection[1])) return()
  textedt <- tclvalue(tkget(focused, selection[1], selection[2]))
  tkclipboard.clear()
  tkclipboard.append(textedt) }
hapus <- function(){
  focused <- tkfocus()
  selection <- strsplit(tclvalue(tktag.ranges(focused, "sel")), " ")[[1]]
  if (is.na(selection[1])) return()
  tkdelete(focused, selection[1], selection[2]) }
potong <- function(){
  salin()
  hapus() }
tempel <- function(){
  focused <- tkfocus()
  textedt <- tclvalue(Tcl("selection get -selection CLIPBOARD"))
  if (length(textedt) == 0) return()
  tkinsert(focused, "insert", textedt) }

Menu2 <- tkmenu(topMenu,tearoff=FALSE)
tkadd(Menu2,"command",label="Salin",command=salin)
tkadd(Menu2,"command",label="Tempel",command=tempel)
tkadd(Menu2,"command",label="Potong",command=potong)
tkadd(Menu2,"command",label="Hapus",command=hapus)
tkadd(topMenu,"cascade",label="Edit",menu=Menu2)
rgl3d<-function(){require(misc3d)
demo(teapot)
demo(shapes3d)}

miscGUI<-function(){
require(PBSmodelling)
demo(runExamples)}
Menu3 <- tkmenu(topMenu,tearoff=TRUE)
tkadd(Menu3,"command",label="Sabuk Keyakinan",command=SabukAnim)
tkadd(Menu3,"command",label="Pencilan",command=scaterAnim)
tkadd(Menu3,"command",label="Konstanta & Kemiringan",command=PencarAnim)

```

```

tkadd(Menu3, "command", label="Fungsi (Bentuk, Ukuran, lokasi)",
command=fungsiAnim)
tkadd(Menu3, "command", label="Distribusi T dan Z", command=tnorm)
tkadd(Menu3, "command", label="Interval Keyakinan", command=tik)
tkadd(Menu3, "command", label="Teorema Limit Pusat", command=tlp)
tkadd(Menu3, "command", label="Perspektif 3D", command=p3d)
tkadd(Menu3, "command", label="kontur 3D", command=kontr)
tkadd(Menu3, "command", label="Misc3D", command=rgl3d)
tkadd(Menu3, "command", label="Estimator Kernel", command=KerGas)
tkadd(Menu3, "command", label="Estimator P-Spline", command=PSPLINE)
tkadd(Menu3, "command", label="Estimator Robust
LMS", command=ROBUSTLMS)
tkadd(Menu3, "command", label="Estimator Robust LTS", command=ROBUSTLTS)
tkadd(topMenu, "cascade", label="demo analisis statistika", menu=Menu3)
qccDemo <- function() {
  library(qcc)
  example(qcc) }
clustDemo<-function(){
  library(amac)
  example(hclust) }
pcaDemo<-function(){
  require(amac)
  example(plot.acp) }
ccaDemo<-function(){
  require(ade4)
  example(cca)}
discrimDemo<-function(){
  require(ade4)
  example(discrimin) }
edaDemo <- function() {
  library(CoCoAn)
  example(CAIV.plot) }
geoRDemo <- function() {
  require(geoR)
  example(plot.geodata) }
mapDemo <- function() {
  require(maptools)
  example(plot.Map)}
rimageDemo <- function() {demo(image)}
treeDemo <- function() { require(maptree)
  par(mfrow=c(1,1))
  example(draw.tree)
  example(draw.clust)}
grapDemo<- function() { demo(graphics)}
glmDemo<- function() { example(glm)}
geeDemo<- function() { require(gee)
  example(gee)}

```

```

dinDemo<- function() { require(dynamicGraph)
  example(dynamicGraph)}
dealDemo<- function() { #Bayesian Network
  require(deal)
  par(mfrow=c(1,1))
  demo(ks1) }
tsDemo <- function() { require(stats)
  example(ts) }
ismevDemo <- function() { library(ismev)
  demo(exchange.rate) }
Menu4 <- tkmenu(topMenu,tearoff=FALSE)
tkadd(Menu4,"command",label="Demo-QCC",command=qccDemo)
tkadd(Menu4,"command",label="Demo-Kluster",command=clustDemo)
tkadd(Menu4,"command",label="Demo-Diskriminan",command=discrimDemo)
tkadd(Menu4,"command",label="Demo-CART",command=treeDemo)
tkadd(Menu4,"command",label="Demo-
Pencitraan(Imaging)",command=rimageDemo)
tkadd(Menu4,"command",label="Demo-GeoStat",command=geoRDemo)
tkadd(Menu4,"command",label="Demo-Grafik",command=grapDemo)
tkadd(Menu4,"command",label="Demo-GLM",command=glmDemo)
tkadd(Menu4,"command",label="Demo-GEE",command=geeDemo)
tkadd(Menu4,"command",label="Network",command=dealDemo)
tkadd(Menu4,"command",label="Demo-Grafik Dinamik",command=dinDemo)
tkadd(Menu4,"command",label="Demo Olah Peta",command=mapDemo)
tkadd(Menu4,"command",label="MiscGUI",command=miscGUI)

browseManual2 <- function() { browseURL("../doc/RcomManSld.pdf") }
browseManual4 <- function() { browseURL("../doc/pancaran.pdf") }
browseManual5 <- function() { browseURL("../doc/StatDasSld.pdf")}
Menu5 <- tkmenu(topMenu,tearoff=FALSE)
tkadd(Menu5,"command",label="Panduan (SlideE1)",command=browseManual2)
tkadd(Menu5,"command",label="Artikel Jurnal
(BukuE1)",command=browseManual4)
tkadd(Menu5,"command",label="Diktat-ANSAR
(SlideE1)",command=browseManual5)
tkadd(topMenu,"command",label="RCmd-GEE",command=hlp.txt)
tkadd(topMenu,"cascade",label="Dokumentasi",menu=Menu5)
tkadd(topMenu,"cascade",label="Demo",menu=Menu4)

tkpack(kb1 <- tklable(ju,text="Stats-Lab FMIPA UNEJ (IMT)& KBK Statistika
FSaintek UNAIR (Nur Chamidah)",fg="darkred",
font=fontFixedWidth))
tkpack(kb <- tkbutton(ju,text="Keluar (Stop)", command=function() tkdestroy(ju)))

tkadd(Menu1,"command",label="Keluar",command=function() tkdestroy(ju))
tkadd(topMenu,"command",label="Kirim/Eksekusi Skrip",command=eval.txt)
tkfocus(ju)}

```

# **LAMPIRAN 5**

## **SERTIFIKAT SEMINAR NASIONAL**



**UNIVERSITAS JEMBER**  
**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**  
**JURUSAN MATEMATIKA**

**SERTIFIKAT**

Nomor : 7/SEM.NAS/MAT/2009

Diberikan kepada :

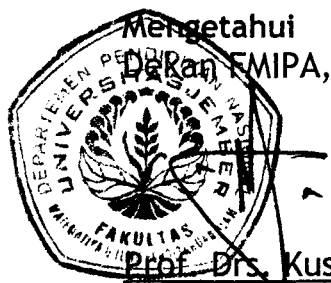
*Nur Chamidah, S.Si., M.Si.*

Atas partisipasinya sebagai

*Pemakalah*

Dalam Seminar Nasional Matematika 2009

dengan Judul : Implementasi Open Source Software (OSS) R Pada Model Regresi Nonparametrik  
Dengan Error Lognormal Berdasarkan Estimator Robust Least Median Squares (LMS)  
pada tanggal 28 Pebruari 2009 di Jurusan Matematika FMIPA - UNEJ



Mengetahui  
Dekan FMIPA,

Prof. Drs. Kusno, DEA, Ph.D  
NIP. 131 592 357

Jember, 28 Pebruari 2009  
Ketua Panitia,

Kosala Dwidja Purnomo, SSi, MSi  
NIP. 132 206 019



# Sertifikat

IR Perputakaan Universitas Airlangga

No. 11/SEMNAS/HIMATIKA-UA/LITBANG-E/XI/2008

Diberikan kepada

**NUR CHAMIDAH**

Universitas Airlangga

Atas partisipasinya sebagai

**PEMAKALAH**

Dengan Judul "Implementasi OSS Statistika R pada Model Regresi Nonparametrik dengan Error Lognormal Berdasarkan Estimator Robust *Least Trimmed Squares* (LTS)"

dalam SEMINAR NASIONAL Matematika dan Statistika 2008

dengan tema "Peranan Matematika dan Statistika dalam Marketing Research dan Industri sebagai Motor Penggerak Roda Perekonomian Nasional"

yang diselenggarakan oleh Himpunan Mahasiswa Departemen Matematika bekerja sama dengan Departemen Matematika Fakultas Sains dan Teknologi Universitas Airlangga

Surabaya, 20 Desember 2008

Dekan Fakultas Sains dan Teknologi  
Universitas Airlangga



**Drs. Salamun, M. Kes.**

Laporan Penelitian

NIP. 131696506

Ketua Departemen Matematika  
Fakultas Sains dan Teknologi  
Universitas Airlangga

**Dr. Eridani**  
Pengembangan Software Statistika .....  
NIP. 132049016

Ketua Pelaksana  
Seminar Nasional Matematika dan  
Statistika 2008

**Napoli Andita**  
NIM. 080710357

Nur Chamidah

# **LAMPIRAN 6**

## **DAFTAR PESERTA WORKSHOP DAN PELATIHAN OSS-R 2008**

# WORKSHOP DAN PELATIHAN ANALISIS DATA DENGAN OSS-R 2008

IR Perpustakaan Universitas Airlangga

NO	NAMA	NAMA LENGKAP	INSTITUSI	NIM / NIP	NO. HP	TTD
1	R. Arif Wibowo	Drs. R. ARIF WIBOWO, M.Si	Fisika Unair	131933021	081330318026	1.
2	Dr. Ni'mah	Dr. Ni'matuzahroh	Biologi Unair	132011657	08121773272	2.
3	Khusnia	Khusnia	Math Unair	080610258	085735082881	3.
4	Linda	HerLinda Agustia Rihandiny	Math Unair	080610190	085646033350	4.
5	Faya Rosida Rosjida	Faya Rosyida	Math Unair	080610125	03171009060	5.
6	Etin Feby P.	ETIN FEBY PAWESTI	Math Unair	080513296	085649457570	6.
7	Sulaiman R.	Sulaiman Rhosyrd	Biologi Unair	080513222	081357699080	7.
8	Z. Ummah	Zahrotul UMMah	Math Unair	080710345	085655357205	8.
9	Himmatul Ulyah	Himmatul Ulyah	Math Unair	080710056	085649261818	9.
10	Arista	Arista Ayu Sapitri	Math Unair	080710344	085655311766	10.
11	Renny Ernawati	Renny Ernawati	Math Unair	080513361	085645321097	11.
12	Adip Widyaroko	Adip Widyaroko	Math Unair	080513271	085648953513	12.
13	Latifatul Fitria	Latifatul Fitria	Biologi Unair	080513297	085648154709	13.
14	Juwinatah	Juwinatah	Biologi Unair			14.
15	Lutfinia Kholida	Lutfinia Kholida	Kimia Unair	080513137	085649942142	15.
16	Amaliatul Choiriah	Amaliatul Choiriah	Kimia Unair	080513299	085655616941	16.
17	Dahlia Puspitasari	Dahlia Puspitasari	Biologi Unair	080513143	08123430979	17.
18	Wayan	Wayan Ari Susanti	Biologi Unair	080513160	085648071574	18.
19	Pramita Laksitarahmi	Pramita Laksitarahmi	Biologi Unair	080513279	085850307033	19.
20	Tri Ruri Handayani	Tri Ruri Handayani	Biologi Unair	080513288	081331062367	20.
21	Iin Khoiratul U	Iin Khoiratul Ummah	Math Unair	080513234	085649489642	21.
22	I Ketut Bambang	I Ketut Bambang Gderawi	Math Unair	080513290	085648939995	22.
23	Winanti Puspita S	Winanti Puspita Sari	Math Unair	080610192	085646527560	23.
24	Marisa Rifada	Marisa Rifada	Math Unair	080513360	085648266260	24.
25						
26						
27						
28						