



**IMPLEMENTASI ALGORITMA MODEL *ROBUST* FAKTOR UNTUK  
TABEL DUA ARAH PADA PROGRAM R**

**SKRIPSI**

oleh

**Jefri Ardian  
NIM 111810101015**

**JURUSAN MATEMATIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS JEMBER  
2015**



**IMPLEMENTASI ALGORITMA MODEL *ROBUST* FAKTOR UNTUK  
TABEL DUA ARAH PADA PROGRAM R**

**SKRIPSI**

diajukan guna melengkapi tugas akhir dan memenuhi salah satu syarat  
untuk menyelesaikan Program Studi Matematika (S1)  
dan mencapai gelar Sarjana Sains

oleh

**Jefri Ardian  
NIM 111810101015**

**JURUSAN MATEMATIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS JEMBER  
2015**

## PERSEMBAHAN

Skripsi ini saya persembahkan untuk :

1. Allah SWT yang telah memberikan kehidupan sempurna ini;
2. Bapak Senam dan Ibu Tukinah yang telah memberikan doa, cinta dan kasih sayangnya dalam perjalanan hidup saya;
3. bapak dan ibu guru sejak taman kanak-kanak sampai perguruan tinggi, yang telah memberikan ilmu dan membimbing dengan penuh kesabaran;
4. Almamater Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam.

## MOTTO

“Barangsiapa bertawakal kepada Allah niscaya Allah menjadikan baginya kemudahan dalam urusannya.”

(QS. At – Thalaq: 65)<sup>\*)</sup>

“Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya”

(QS. Al-Baqarah: 286)<sup>\*\*)</sup>

---

\*) Departemen Agama Republik Indonesia. 2006. *Al-Qur'an Al – Karim dan Terjemah Bahasa Indonesia*. Kudus: MENARA KUDUS.

\*\*\*) Departemen Agama Republik Indonesia. 2006. *Al-Qur'an dan Terjemahannya*. Surabaya: Duta Ilmu Surabaya.

**PERNYATAAN**

Saya yang bertanda tangan di bawah ini :

nama : Jefri Ardian

NIM : 111810101015

menyatakan dengan sesungguhnya bahwa karya ilmiah yang berjudul “Implementasi Algoritma Model *Robust* Faktor untuk Tabel Dua Arah pada Program R” adalah benar-benar hasil karya sendiri, kecuali kutipan yang sudah saya sebutkan sumbernya, belum pernah diajukan pada institusi manapun dan bukan karya jiplakan. Saya bertanggung jawab atas keabsahan dan kebenaran isinya sesuai dengan sikap ilmiah yang harus dijunjung tinggi.

Demikian pernyataan ini saya buat dengan sebenarnya, tanpa ada tekanan dan paksaan dari pihak manapun serta bersedia mendapat sanksi akademik jika ternyata di kemudian hari pernyataan ini tidak benar.

Jember, November 2015

Yang menyatakan,

Jefri Ardian

NIM 111810101015

**SKRIPSI**

**IMPLEMENTASI ALGORITMA MODEL *ROBUST* FAKTOR UNTUK  
TABEL DUA ARAH PADA PROGRAM R**

oleh  
Jefri Ardian  
NIM 111810101015

Pembimbing :

Dosen Pembimbing Utama : Dr. Alfian Futuhul Hadi, S.Si., M.Si

Dosen Pembimbing Anggota : Dian Anggraeni, S.Si., M.Si

**PENGESAHAN**

Skripsi berjudul “Implementasi Algoritma Model *Robust* Faktor untuk Tabel Dua Arah pada Program R” telah diuji dan disahkan pada :

hari, tanggal :

tempat : Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas  
Jember.

TimPenguji :

Ketua,

Sekretaris,

Dr. Alfian Futuhul Hadi, S.Si., M.Si

Dian Anggraeni, S.Si., M.Si

NIP. 19740719 200012 1 001

NIP. 19820216 200604 2 002

Penguji I,

Penguji II,

Prof. Drs. I Made Tirta, M.Sc., Ph.D

M. Ziaul Arif, S.Si., M.Sc

NIP. 19591220 198503 1 002

NIP. 19850111 200812 1 002

Mengesahkan

Dekan,

Prof. Drs. Kusno, DEA., Ph.D

NIP. 19610108 198602 1 001

## RINGKASAN

**Implementasi Algoritma Model *Robust* Faktor untuk Tabel Dua Arah pada Program R**; Jefri Ardian, 111810101015; 2015; 39 halaman; Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.

*Robust* Faktor analitik merupakan salah satu teknik statistika untuk menganalisis tentang saling ketergantungan dari beberapa variabel secara simultan pada data yang mengandung *outliers*. Tujuan *Robust* Faktor analitik adalah untuk menyederhanakan dari bentuk hubungan antara beberapa variabel yang diteliti menjadi sejumlah faktor yang lebih sedikit dari pada variabel yang diteliti.

Metode yang digunakan dalam *robust* faktor analitik pada tabel dua arah yaitu dengan menggunakan estimasi *Robust Alternating Regression* (RAR). RAR merupakan penduga iteratif untuk mendapatkan nilai penduga *loading* dan penduga skor yang dilakukan secara bergantian atau bolak-balik. *Software* statistik yang digunakan untuk analisis data menggunakan S-plus yang merupakan *software* statistik yang komersil dan program R yang merupakan *software* statistik *open-source*. Secara umum sintaks dari bahasa R adalah ekuivalen dengan bahasa pemrograman S, sehingga sebagian besar keperluan analisis statistika dan pemrograman dengan R adalah hampir identik dengan perintah yang dikenal di S-plus. Dengan menggunakan *software* statistik S-plus dan program R, penelitian ini bertujuan untuk mengevaluasi hasil analisis fungsi-fungsi yang digunakan pada *script* yang dibuat oleh Croux *et al.* (2003) dengan menggunakan estimasi *Robust Alternating Regression* (RAR).

Data yang digunakan dalam penelitian ini adalah data penelitian perbenihan padi di beberapa kabupaten di Pulau Jawa menggunakan 10 varietas dan 8 lokasi percobaan. Kemudian data diolah dengan menggunakan *script* yang disesuaikan pada *software* S-plus dan program R. Perlakuan untuk menjalankan *software* S-plus dan program R dilakukan dengan *me-restart* ulang untuk setiap menjalankan

*software* S-plus dan program R. Hasilnya adalah jumlah iterasi S-plus lebih sedikit dari pada program R. Biplot yang dihasilkan dari S-plus dan program R mempunyai kemiripan dengan informasi pengaruh interaksi yang diberikan S-plus sebesar 73,33% sedangkan program R sebesar 70,9%. Berdasarkan MSE, program R memiliki MSE yang lebih kecil daripada S-plus. MSE yang dihasilkan oleh S-plus sebesar 0,2786216 dan program R sebesar 0,001381457. Selain itu, program R lebih sensitif untuk mendeteksi adanya *outliers* pada data dilihat dari bobot yang diberikan pada efek baris dan efek kolom dan boxplot yang direpresentasikan dari efek baris dan efek kolom.

## PRAKATA

Alhamdulillah, puji syukur kehadirat Allah SWT yang telah memberikan rahmat dan hidayah-Nya sehingga tugas akhir yang berjudul “Implementasi Algoritma Model *Robust* Faktor untuk Tabel Dua Arah pada Program R” dapat terselesaikan dengan baik. Tugas akhir ini disusun untuk memenuhi syarat menyelesaikan pendidikan strata satu (S1) di Jurusan Matematika dan Ilmu Pengetahuan Alam Universitas Jember.

Penyusunan skripsi ini tidak terlepas dari bantuan berbagai pihak, baik secara langsung maupun tidak langsung. Oleh karena itu, penulis menyampaikan terima kasih kepada :

1. Dr. Alfian Futuhul Hadi, S.Si., M.Si selaku Dosen Pembimbing Utama dan Dian Anggraeni S.Si., M.Si selaku Dosen Pembimbing Anggota yang telah memberikan bimbingan secara intensif dan bantuan untuk penyempurnaan skripsi ini;
2. Prof. Drs. I Made Tirta, M.Sc., Ph.D dan M. Ziaul Arif, S.Si., M.Sc selaku Dosen Penguji yang telah memberikan kritik dan saran yang membangun untuk penyempurnaan skripsi ini;
3. seluruh dosen dan karyawan Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam yang telah memberikan ilmu serta fasilitas yang membantu selama proses perkuliahan berlangsung;
4. Bapak Senam dan Ibu Tukinah yang selalu memberikan dukungan baik lahir maupun batin;
5. sahabati tercinta Niya Liyani yang telah memberikan dukungan serta motivasi dalam pembuatan skripsi ini;
6. HIMATIKA “Geokompstat” yang telah memberikan berbagai pengalaman berorganisasi selama menjadi mahasiswa di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam

7. UKM KSR PMI Unit Universitas Jember yang memberikan berbagai pengalaman dan *softskill* yang bermanfaat dalam kehidupan;
8. para akademisi (Yulio, Haki, Darul, Rafi', Ivan, Hendri, Emil, Feri, Rian, Zulfi, Saiful) yang senantiasa memberi dukungan dalam proses menyelesaikan tugas akhir;
9. keluarga besar KRAMAT'11 yang tidak dapat saya sebut satu per satu yang selalu menemani dan memberikan dukungan dalam hal apapun.

Penulis menyadari bahwa dalam menyusun skripsi ini masih terdapat kekurangan, baik isi maupun susunannya. Oleh karena itu, penulis mengharapkan saran dan kritik demi penyempurnaan skripsi ini. Akhirnya penulis berharap semoga skripsi ini dapat memberi manfaat dan sumbangan bagi pembaca.

Jember, November 2015

Jefri ardian

DAFTAR ISI

	Halaman
HALAMAN JUDUL .....	i
HALAMAN PERSEMBAHAN .....	ii
HALAMAN MOTTO .....	iii
HALAMAN PERNYATAAN.....	iv
HALAMAN PEMBIMBINGAN.....	v
HALAMAN PENGESAHAN.....	vi
RINGKASAN .....	vii
PRAKATA .....	ix
DAFTAR ISI.....	xi
DAFTAR TABEL .....	xiii
DAFTAR GAMBAR.....	xiv
DAFTAR LAMPIRAN .....	xv
<b>BAB 1. PENDAHULUAN .....</b>	<b>1</b>
<b>1.1 Latar Belakang .....</b>	<b>1</b>
<b>1.2 Rumusan Masalah.....</b>	<b>3</b>
<b>1.3 Tujuan .....</b>	<b>3</b>
<b>1.4 Manfaat .....</b>	<b>3</b>
<b>BAB 2. TINJAUAN PUSTAKA.....</b>	<b>4</b>
<b>2.1 Analisis Faktor .....</b>	<b>4</b>
<b>2.2 Outliers (Pencilan) .....</b>	<b>4</b>
<b>2.3 Robust Principal Component Analysis (RobPCA).....</b>	<b>5</b>
<b>2.4 Robust Alternating Regression (RAR) Estimator .....</b>	<b>7</b>
<b>2.5 Software S-plus .....</b>	<b>9</b>
<b>2.6 Program R.....</b>	<b>10</b>
<b>2.7 Analisis Robust Biplot .....</b>	<b>10</b>
<b>2.8 Mean Square Error (MSE).....</b>	<b>11</b>

2.9 Implementasi Algoritma pada S-plus.....	12
<b>BAB 3. METODE PENELITIAN.....</b>	<b>16</b>
3.1 Data .....	16
3.2 Desain Penelitian .....	16
3.3 Struktur Fungsi Program R.....	19
<b>BAB 4. HASIL DAN PEMBAHASAN.....</b>	<b>21</b>
4.1 Implementasi Algoritma pada Program R .....	21
4.1.1 Algoritma Weight.w11 .....	21
4.1.2 <i>Robust Principal Komponent Analysis</i> (RobPCA).....	23
4.1.3 Algoritma Twoway.rob .....	24
4.2 Fungsi-fungsi yang digunakan pada <i>Script S-plus</i> dan Program R .	31
4.3 Aplikasi <i>Script</i> pada Data Riil dengan S-plus dan Program R .....	32
4.4 Interpretasi Biplot.....	33
4.5 Identifikasi <i>Outliers</i> .....	35
<b>BAB 5. PENUTUP.....</b>	<b>37</b>
5.1 Kesimpulan .....	37
<b>DAFTAR PUSTAKA .....</b>	<b>38</b>
<b>LAMPIRAN.....</b>	<b>40</b>

**DAFTAR TABEL**

	Halaman
4.1 Fungsi-fungsi yang digunakan pada <i>Script</i> S-plus dan Program R.....	31
4.2 Nilai Hasil Model <i>Robust</i> Faktor .....	32
4.3 Pembobot bagi Pengaruh Baris terhadap Pencilan pada S-plus .....	35
4.4 Pembobot bagi Pengaruh Kolom terhadap Pencilan pada S-plus .....	35
4.5 Pembobot bagi Pengaruh Baris terhadap Pencilan pada Program R .....	36
4.6 Pembobot bagi Pengaruh Kolom terhadap Pencilan pada Program R .....	36

**DAFTAR GAMBAR**

	Halaman
2.1 Bagan Alur Model <i>Robust</i> Faktor Analitik pada S-plus .....	13
3.1 Skema alur penelitian .....	18
4.1 Biplot S-plus .....	33
4.2 Biplot Program R .....	34
4.3 (a) Boxplot Pengaruh Baris (Genotip) Data Pembenihan Padi dengan S-plus .....	35
4.3 (b) Boxplot Pengaruh Kolom (Lingkungan) Data Pembenihan Padi dengan S-plus .....	35
4.4 (a) Boxplot Pengaruh Baris (Genotip) Data Pembenihan Padi dengan Program R .....	36
4.4 (b) Boxplot Pengaruh Kolom (Lingkungan) Data Pembenihan Padi dengan Program R .....	36

**DAFTAR LAMPIRAN**

	Halaman
<b>A. Data Perbenihan Padi .....</b>	<b>40</b>
<b>B. Script Data Perbenihan Padi pada S-plus.....</b>	<b>41</b>
B.1 Script Weight.w11 .....	41
B.2 Script Prcomp.rob dan Keluarannya.....	41
B.3 Script Twoway.rob dan Keluarannya .....	42
<b>C. Script Data Perbenihan Padi pada Program R.....</b>	<b>55</b>
C.1 Script Weight.w11 .....	55
C.2 Script PcaProj dan Keluarannya .....	56
C.3 Script Twoway.rob dan Keluarannya .....	56

## BAB 1. PENDAHULUAN

### 1.1 Latar Belakang

Regresi merupakan suatu metode statistika yang digunakan untuk menyelidiki pola hubungan antara dua atau lebih variabel. Apabila pola hubungan yang diselidiki terdiri dari satu variabel respon dan satu variabel bebas maka disebut regresi linier sederhana. Sedangkan regresi linier berganda adalah regresi linier yang terdiri dari satu variabel respon dan lebih dari satu variabel bebas. Regresi adalah metode statistika yang baik untuk memodelkan suatu data yang tidak mengandung pencilan atau *outliers*.

Jika data yang diteliti mengandung suatu *outlier*, maka model regresi tidak memenuhi asumsi dan tidak cocok untuk data yang akan dimodelkan, karena nilai koefisien dari model regresi dipengaruhi oleh adanya *outlier*. Model yang dihasilkan tidak dapat digunakan untuk memprediksi, sebelum *outlier* yang ada dapat diatasi. Diperlukan analisis yang mampu mengakomodir adanya pencilan yaitu dengan pendekatan *robust* analitik. Regresi *robust* merupakan metode regresi yang digunakan ketika residual berdistribusi tidak normal dan mengandung beberapa *outlier* yang berpengaruh pada model (Ryan, 1997). Pada regresi *robust*, taksiran yang *robust* terhadap *outlier* akan dicari sehingga *outlier* yang ada tidak akan diabaikan. Salah satu regresi *robust* yang digunakan adalah faktor analitik.

Faktor analitik merupakan suatu teknik untuk menganalisis tentang saling ketergantungan dari beberapa variabel secara simultan dengan tujuan untuk menyederhanakan dari bentuk hubungan antara beberapa variabel yang diteliti menjadi sejumlah faktor yang lebih sedikit dari pada variabel yang diteliti (Suliyanto, 2005). Pada faktor analitik, variabel bebas dan terikatnya tidak dibedakan. Keuntungan dari penggunaan *robust* faktor analitik adalah tabel data yang mengandung pencilan dapat diperiksa lebih mudah (Hawkins *et al.*, 2001). Analisis data dengan pendekatan *robust* analitik dilakukan pada data yang akan

disajikan dalam bentuk tabel dua arah. Tabel dua arah atau komponen adalah tabel yang menunjukkan dua kategori atau dua karakteristik. Metode yang digunakan dalam *robust* faktor analitik pada tabel dua arah yaitu dengan menggunakan estimasi *Robust Alternating Regression* (RAR). RAR merupakan penduga iteratif untuk mendapatkan nilai penduga *loading* dan penduga skor yang dilakukan secara bergantian atau bolak-balik.

Beberapa penelitian sebelumnya mengenai *Robust Alternating Regression* (RAR) diantaranya Croux *et al.* (2003) dan Hadi (2011). Pada penelitian ini dengan menggunakan estimasi *Robust Alternating Regression* (RAR) akan dilakukan analisis pada tabel dua arah dengan menggunakan data pembenihan padi di beberapa kabupaten di Pulau Jawa. Data yang diperoleh akan diolah dengan menggunakan *software* statistik. Tersedia banyak *software* statistik yang bisa digunakan untuk melakukan analisis data diantaranya dengan menggunakan S-plus yang merupakan *software* statistik yang komersil dan program R yang merupakan *software* statistik *open-source*.

S-plus merupakan salah satu *software* statistik yang menggunakan bahasa pemrograman S. S-plus merupakan salah satu *software* statistik yang masih jarang penggunaannya seperti pada bidang pertanian, perikanan, ekonomi, dan lain-lain. S-plus dapat menangani data, *functions* dan *fitted models* sebagai *objects*, sehingga membuat analisis data menjadi lebih fleksibel (Burn, 1998). Sedangkan program R merupakan suatu sistem analisis data statistik dengan menggunakan bahasa R yang merupakan versi *open-source* dari bahasa pemrograman S. Fungsi dan kemampuan dari R sebagian besar diperoleh melalui *Add-on packages/library*. Suatu *library* adalah kumpulan perintah atau fungsi yang dapat digunakan untuk melakukan analisis tertentu (Suhartono, 2008). Secara umum sintaks dari bahasa R adalah ekuivalen dengan bahasa pemrograman S, sehingga sebagian besar keperluan analisis statistika dan pemrograman dengan R adalah hampir identik dengan perintah yang dikenal di S-plus. Penelitian sebelumnya mengenai *Robust Alternating Regression* (RAR) oleh Croux *et al.* (2003) menggunakan S-plus untuk analisis data. Dengan menggunakan program R, peneliti akan mentransformasi fungsi-fungsi yang digunakan pada *script* yang

dibuat oleh Croux *et al.* (2003) dengan menggunakan estimasi *Robust Alternating Regression* (RAR) dan mengevaluasi hasil penanganan *outliers software* S-plus 8.0 dan program R 2.7.0 dengan menggunakan data perbenihan padi di beberapa kabupaten di Pulau Jawa untuk mengetahui kemampuan *software*.

## 1.2 Rumusan Masalah

Rumusan masalah dari penelitian ini adalah sebagai berikut :

- a. bagaimana mentransformasi fungsi-fungsi yang digunakan pada *script* penanganan *outlier* pada data tabel dua arah menggunakan *robust* faktor analitik ke dalam program R yang *open-source*?
- b. bagaimana evaluasi hasil penanganan *outlier* pada data tabel dua arah menggunakan *robust* faktor analitik dengan menggunakan *software* S-plus 8.0 dan program R 2.7.0 untuk mengetahui kemampuan *software* dengan menggunakan data perbenihan padi di beberapa kabupaten di Pulau Jawa?

## 1.3 Tujuan

Adapun tujuan dari penelitian ini adalah sebagai berikut :

- a. mendapatkan *script* penanganan *outlier* pada data tabel dua arah menggunakan *robust* faktor analitik hasil transformasi dari S-plus ke program R,
- b. memperoleh hasil evaluasi penanganan *outlier* pada data tabel dua arah menggunakan *robust* faktor analitik pada *software* S-plus dan program R.

## 1.4 Manfaat

Adapun manfaat yang dapat diperoleh dari penelitian ini adalah sebagai berikut :

- a. menghindari adanya pembuangan atau pengabaian *outliers* pada *robust* faktor analitik,
- b. menyediakan alternatif analisis statistik yang *open-source* untuk menghindari adanya pembuangan atau pengabaian *outliers* pada *robust* faktor analitik.

## BAB 2. TINJAUAN PUSTAKA

### 2.1 Analisis Faktor

Analisis faktor merupakan suatu teknik untuk menganalisis tentang ketergantungan dari beberapa variabel secara simultan dengan tujuan untuk menyederhanakan bentuk hubungan antara beberapa variabel yang diteliti menjadi sejumlah faktor yang lebih sedikit dari variabel yang diteliti. Hal ini berarti, analisis faktor dapat juga menggambarkan tentang struktur data dari suatu penelitian (Suliyanto, 2005).

Analisis faktor adalah suatu teknik interdependensi, sehingga tidak ada pembagian variabel menjadi variabel bebas dan variabel tergantung dengan tujuan utama yaitu mendefinisikan struktur yang terletak diantara variabel-variabel dalam analisis. Tujuan umum dari teknik analisis faktor adalah menemukan suatu cara untuk mereduksi informasi yang terkandung didalam sejumlah variabel-variabel original ke dalam set variabel yang lebih kecil dari dimensi-dimensi gabungan dan baru. Pada prinsipnya analisis faktor digunakan untuk mengelompokkan beberapa variabel yang memiliki kemiripan untuk dijadikan satu faktor, sehingga dimungkinkan dari beberapa atribut yang mempengaruhi satu komponen variabel dapat diringkas menjadi beberapa faktor utama yang jumlahnya lebih sedikit (Hair *et al.*, 2010).

### 2.2 Outlier (Pencilan)

Definisi *outlier* adalah satu atau lebih pengamatan yang nilainya menyimpang cukup jauh dari pengamatan lain. Data *outlier* adalah pengamatan yang dapat diidentifikasi secara jelas yang berbeda dari pengamatan yang lain, namun data *outlier* dapat menunjukkan karakteristik dari populasi sehingga tidak harus diabaikan atau dibuang. Menurut Kriegel (2010) suatu nilai dianggap sebagai *outliers*, apabila menyimpang lebih dari tiga kali standar deviasi. Apabila data *outlier* bermasalah, tidak mewakili populasi dan bertentangan dari tujuan

analisis, maka secara serius dapat memberikan hasil uji statistik yang berbeda. Pemeriksaan atau identifikasi *outlier* pada data harus dilakukan karena *outlier* memberikan pengaruh pada ragam dan setelah *outlier* teridentifikasi maka dapat diputuskan untuk mempertahankan atau menghapus *outlier* tersebut (Hair *et al.*, 1992).

Identifikasi *outlier* umumnya didasarkan pada *Mahalanobis Distance* (MD). Metode MD adalah suatu metode yang digunakan untuk mendapatkan suatu data dengan jarak tertentu terhadap mean data tersebut, sehingga diperoleh suatu penyebaran data yang memiliki pola terhadap nilai mean. Metode ini didasarkan pada korelasi antara variabel dengan pola yang berbeda yang dapat diidentifikasi dan dianalisis. Metode ini memiliki nilai penguat yang fleksibel sehingga mudah disesuaikan dengan perubahan kondisi. Penguat dari metode ini ada pada nilai kovariannya. Secara umum MD dapat didefinisikan sebagai berikut :

$$MD_i = \sqrt{(x_i - \bar{x})^T S^{-1} (x_i - \bar{x})} \quad (2.1)$$

dimana,

$x_i$  = data objek ke- $i$

$\bar{x}$  = vektor rata-rata

S = matriks *variance-covariance*

untuk  $i = 1, \dots, n$ ,  $x_i$  merupakan *outlier* jika  $MD > X_{\alpha;p}^2$ . Prinsip MD adalah menghitung jarak di ruang multidimensional antara sebuah pengamatan dengan pusat dari semua pengamatan (Hair *et al.*, 1998).

### 2.3 Robust Principal Component Analysis (RobPCA)

*Principal Component Analysis* (PCA) atau analisis komponen utama adalah metode analisis peubah ganda yang bertujuan memperkecil dimensi peubah asal sehingga diperoleh peubah baru (komponen utama) yang tidak saling berkorelasi, tetapi menyimpan sebagian besar informasi yang terkandung pada peubah asal (Jolliffe, 2002). Terdapat tiga keunggulan analisis komponen utama yaitu memiliki kombinasi linear dengan *Mean Square Error* (MSE) optimal dalam meringkas rangkaian vektor berdimensi tinggi menjadi rangkaian vektor berdimensi rendah dan kemudian membentuknya kembali, parameter model dapat secara langsung

dihitung dari data dan peringkasan mudah dioperasikan untuk menunjukkan parameter model (Chen, 2002).

Meskipun mempunyai beberapa keunggulan, analisis komponen utama akan sangat dipengaruhi dengan adanya *outliers* pada data. Kelemahan analisis komponen utama dengan adanya *outliers* menyebabkan diperlukannya metode analisis komponen utama yang *robust* terhadap adanya *outliers*. Analisis komponen utama yang *robust* didapatkan dengan mengganti matriks varian kovarian dengan suatu estimator yang *robust*. Estimator *robust* yang diinginkan adalah memiliki *breakdown point* (tingkat *robust*) yang tinggi. *Breakdown point* merupakan proporsi terkecil dari pencemar yang mampu mempengaruhi estimator, yaitu nilainya mengalami pergeseran yang cukup jauh dari  $t(X)$ . Sifat lainnya yang diinginkan adalah *affine equivariant*. Suatu estimator bagi ukuran pemusatan, yaitu  $t(X)$  dan ukuran simpangan, yaitu  $C(X)$  adalah *affine equivariant* jika dan hanya jika untuk setiap vektor baris  $b \in \mathfrak{R}^p$  (konstanta dalam ruang berdimensi  $p$ ) dan setiap matriks non-singular  $A_{p \times p}$  berlaku,

$$\begin{aligned} t(AX + b) &= At(X) + b, \\ C(AX + b) &= AC(X)A^t \end{aligned} \quad (2.2)$$

Estimator *robust Minimum Volume Ellipsoid* (MVE) merupakan pasangan  $t(X)$  dan  $C(X)$ , dimana  $t(X)$  merupakan vektor rata-rata dan  $C(X)$  merupakan matriks  $p \times p$  simetris definit-positif, dari suatu sub-sampel berukuran  $h$  pengamatan dimana *volume ellipsoid* dari sub-sampel tersebut adalah yang minimal. Dengan  $h_0 \leq h \leq n$  dan  $h_0$  merupakan nilai integer terkecil dari  $\left(\frac{(n+p+1)}{2}\right)$ ,

$$\begin{aligned} MVE \approx \min\{m_j^{2p} \det(C(X)_j)\}^{1/2}, \\ j = 1, \dots, \binom{n}{h} \end{aligned} \quad (2.3)$$

(Rousseeuw, 1985).

Analisis komponen utama yang *robust* juga menggunakan metode *projection-pursuit* (PP). Tujuan digunakannya metode *projection-pursuit* untuk mendapatkan struktur pada data peubah ganda dengan memproyeksikannya pada subruang berdimensi lebih rendah. PP tepat digunakan untuk menganalisis data

dengan jumlah peubah yang besar. Sub-ruang berdimensi rendah dipilih dengan memaksimalkan indeks proyeksi tertentu (Huber, 1985).

#### 2.4 Robust Alternating Regression (RAR) Estimator

Matriks  $Y$  berukuran  $n \times p$  digunakan untuk melambangkan catatan yang berisi peubah pengamatan atau karakteristik (pada kolom) atas beberapa individu (objek/case) pada barisnya. Suatu skor faktor dicatat sebagai  $f_{ik}$ , dengan vektor skor ke- $i$  diberikan sebagai  $f_i = (f_{i1}, \dots, f_{ik})^T$ , sedangkan *loading* vektor ke- $j$  adalah  $\lambda_j = (\lambda_{j1}, \dots, \lambda_{jk})^T$  keduanya tidak diketahui. Vektor dari seluruh *loading* dan skor faktor dilambangkan dengan  $\theta = (f_1^T, \dots, f_n^T, \lambda_1^T, \dots, \lambda_p^T)$ , dan

$$\hat{y}_{ij}(\theta) = \sum_{k=1}^k f_{ik} \lambda_{kj} = f_i^T \lambda_j = \lambda_j^T f_i \quad (2.4)$$

adalah nilai dugaan (*fitted value*) bagi  $y_{ij}$ , dengan memilih  $\theta$  sedemikian sehingga nilai dugaan dan nilai aktual dari matriks sangat dekat, didefinisikan nilai dugaan vektor skor  $\hat{f}_i$  dan  $\hat{\lambda}_j$  sebagai nilai dugaan vektor *loading*. Nilai dugaan matriks data  $\hat{Y}$  dapat didekomposisi menjadi  $\hat{Y} = \hat{F} \hat{\Lambda}^T$  dengan baris-baris dari  $\hat{F}$  adalah nilai dugaan skor dan baris-baris dari  $\hat{\Lambda}^T$  adalah nilai dugaan *loading*.

Rank dari  $\hat{Y}$  sebesar-besarnya adalah  $k < p$ , sedangkan rank dari  $Y$  adalah sebesar  $p$ . Pendekatan metode kuadrat terkecil adalah meminimumkan kuadrat sisaan :

$$\hat{\theta}_{LS} = \operatorname{argmin}_{\theta} \sum_{i=1}^n \sum_{j=1}^p (y_{ij} - \hat{y}_{ij}(\theta))^2 \quad (2.5)$$

dan menghasilkan nilai dugaan  $\hat{Y}$  yang dapat dilihat sebagai pendekatan “terbaik” (dalam kacamata kuadrat terkecil) terhadap matriks  $Y$  oleh matriks ber-*rank*  $k$ . Teorema Eckart-Young (Gower dan Hand, 1996) menyebutkan bahwa aproksimasi terbaik dapat diperoleh melalui dekomposisi nilai singular (*singular value decomposition/ SVD*) terhadap matriks  $Y = UDV^T$ , dengan mengganti semua nilai singular dalam matriks  $D$  dengan nol kecuali untuk  $k$  buah nilai singular terbesar pertama. Diperoleh  $D_k$  dan akhirnya  $\hat{Y} = U D_k V^T$ , dengan menyatakan  $\hat{F} = \sqrt{n}U$  dan  $\hat{\Lambda} = V D_k / \sqrt{n}$  diperoleh apa yang disebut solusi komponen utama untuk masalah faktor analisis (Johnson dan Wichern, 1998).

Matriks peragam contoh dari vektor dugaan adalah  $\hat{F}^T \hat{F} = I_k$  dimana hal ini konsisten dengan asumsi awal  $\text{Cov}(F) = I_k$ . Nilai dugaan  $\hat{F}$  dan  $\hat{\Lambda}^T$  hanya dinyatakan secara khusus untuk suatu transformasi linier, karena  $\hat{Y} = (\hat{F}^T T^T)(\hat{\Lambda} T^{-1})^T$  untuk sembarang matriks T non-singular berukuran  $k \times k$  matriks, maka  $\hat{F}^T T^T$  dan  $\hat{\Lambda} T^{-1}$  mendapat nilai yang sama bagi fungsi tujuan  $\hat{\theta}_{LS}$ . Meskipun demikian nilai dugaan  $\hat{Y}$  yang diperoleh bersifat unik, jika ditambahkan suatu kendala bahwa dugaan matriks peragam bagi vektor skor adalah matriks satuan, maka dugaan  $\hat{F}$  dan  $\hat{\Lambda}^T$  menjadi suatu transformasi orthogonal sehingga matriks  $\hat{\Lambda} \hat{\Lambda}^T$  juga unik. Karena kriteria kuadrat terkecil memberikan pembobot yang berlebih kepada residu yang bernilai besar, gagasan awal mengatasinya adalah menggunakan kriteria simpangan mutlak terkecil (*Least Absolute Deviations*) atau disebut kriteria L1 yang diketahui memberikan penduga aditif bagi tabel dua arah yang robust (Terbeck dan Davies, 1998).

Kriteria L1 menghasilkan penduga yang meminimumkan simpangan mutlak sebagai berikut :

$$\hat{\theta}_{L1} = \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^n \sum_{j=1}^p |y_{ij} - \hat{y}_{ij}(\theta)| \quad (2.6)$$

Nilai  $\hat{F}$  dan  $\hat{\Lambda}^T$  dapat diperoleh secara optimal jika dipenuhinilai dugaan  $\hat{f}_i$  meminimumkan  $\sum_{j=1}^p |y_{ij} - f_i^T \hat{\lambda}_j|$  dan  $\hat{y}_j$  meminimumkan  $\sum_{i=1}^n |y_{ij} - \hat{f}_i^T \lambda_j|$ , oleh karena itu, pendekatan ini tidaklah meminimumkan kedua jumlah mutlak pada kriteria L1 secara bersamaan, melainkan dengan menetapkan indeks- $j$  dan skor  $f_i$  yang kemudian mencari  $\lambda_j$  yang meminimumkan  $\sum_{i=1}^n |y_{ij} - \hat{f}_i^T \lambda_j|$ . Meminimumkan  $\sum_{i=1}^n |y_{ij} - \hat{f}_i^T \lambda_j|$  dilakukan satu persatu untuk setiap  $j=1, \dots, p$  berkenaan dengan kriteria L1 untuk skor yang ditetapkan. Analog dengan itu, untuk *loading*  $\lambda_j$  yang ditetapkan, mencari  $f_i$  yang meminimumkan  $\sum_{j=1}^p |y_{ij} - f_i^T \hat{\lambda}_j|$  (untuk setiap  $i = 1, \dots, n$  bergantian) berkenaan dengan meminimumkan kriteria L1 pada nilai *loading* yang diberikan. Bergantian bolak-balik pada pemimuman  $\sum_{i=1}^n |y_{ij} - \hat{f}_i^T \lambda_j|$  dan  $\sum_{j=1}^p |y_{ij} - f_i^T \hat{\lambda}_j|$  mengarahkan pada skema pendugaan iteratif yang disebut *alternating regression* (regresi bolak-balik).

Regresi L1 sensitif terhadap pengamatan pengungkit (*leverage points*), jika terdapat skor atau *loading* yang memencil maka regresi L1 dapat sangat terpengaruh olehnya. Dengan memberikan pembobot yang bernilai lebih rendah (*downweighting*) terhadap pengamatan pengungkit didapatkan regresi L1 terboboti (*weight L1 regression*), dan menghasilkan penduga sebagai berikut :

$$\hat{\theta}_{RAR} = \operatorname{argmin}_{\theta} \sum_{i=1}^n \sum_{j=1}^p w_i(\theta) v_j(\theta) |y_{ij} - \hat{y}_{ij}(\theta)| \quad (2.7)$$

Satu fungsi objektif tunggal menduga  $\hat{F}$  dan  $\hat{\Lambda}^T$  secara simultan dari baris-baris dan kolom-kolom matriks Y, untuk mendapatkan nilai dugaan  $\hat{\theta}_{RAR}$  digunakan regresi bolak-balik *robust (Robust Alternating Regression)*. Penduga  $\hat{\theta}_{RAR}$  tidak akan salah arah oleh adanya pengamatan *outliers*. Pembobot bagi baris,  $w_i(\theta)$  didefinisikan oleh  $w_i(\theta) = \min\left(1, \frac{x_k^2, 0.95}{RD_i^2}\right)$  untuk  $i = 1, \dots, n$  dimana  $x_k^2, 0.95$  adalah batas atas bagi nilai kritis 5% sebaran khi-kuadrat dengan derajat bebas k.

## 2.5 Software S-plus

S-plus merupakan salah satu *software* statistik yang masih jarang penggunaannya oleh *stakeholder*, seperti pada bidang pertanian, perikanan, ekonomi, dan lain-lain. S-plus memiliki keunggulan pada beberapa point analisis, yaitu model *tree*, metode resampling, dan regresi nonparametrik (*smoothing*). S-plus dirancang oleh AT & T Bells Labs dengan menggunakan bahasa S. S-plus dapat menangani data, *functions* dan *fitted models* sebagai *objects*, sehingga membuat analisis data menjadi lebih fleksibel (Burn, 1998).

Penggunaan Microsoft Office yang kompatibel dengan S-plus, memberikan akses yang mudah untuk manipulasi data, grafik, dan statistik. Beberapa hal yang dapat dilakukan dengan menggunakan S-plus antara lain dapat memasukkan data dari hampir semua sumber, mengedit data, membuat plot dengan klik tombol, mengatur setiap detail dari grafis dan hasil yang menakjubkan, keluaran tampak profesional untuk keluaran ke dalam bentuk laporan dokumen, melakukan analisis statistik dari sistem menu, menjalankan fungsi analisis satu per satu pada baris

perintah atau dalam sejumlah fungsi menggunakan *script window*, membuat fungsi sendiri (Insighful corp., 2007).

## 2.6 Program R

Program R merupakan suatu sistem analisis data statistik yang komplit sebagai hasil dari kolaborasi penelitian berbagai ahli statistik di seluruh dunia. Versi awal dari R dibuat pada tahun 1992 di Universitas Auckland, New Zealand oleh Ross Ihaka dan Robert Gentleman. Pada saat ini, *source code* kernel R dikembangkan terutama oleh R Core Team yang beranggotakan 17 orang statistisi dari berbagai penjuru dunia (lihat <http://www.r-project.org/contributors.html>). Selain itu, para statistisi lain pengguna R di seluruh dunia juga memberikan kontribusi berupa kode, melaporkan bug dan membuat dokumentasi untuk R. Paket statistik R bersifat *multiplatforms*, dengan *file* instalasi *binary/file rar* yang tersedia untuk sistem operasi windows, Mac OS, Mac OS X, Linux, Free BSD, Net BSD, irix, Solaris, AIX dan HPUX. Secara umum sintaks dari bahasa R adalah ekuivalen dengan paket statistik S-plus, sehingga sebagian besar keperluan analisis statistika dan pemrograman dengan R adalah hampir identik dengan perintah yang dikenal di S-plus (Suhartono, 2008).

Fungsi dan kemampuan dari R sebagian besar diperoleh melalui *Add-on packages/library*. Suatu *library* adalah kumpulan perintah atau fungsi yang dapat digunakan untuk melakukan analisis tertentu (Suhartono, 2008). Keunggulan dari R diantaranya adalah efektif dalam pengelolaan data, terdiri dari koleksi *tools* statistik yang terintegrasi untuk analisis data dan dapat dikembangkan sesuai keperluan dan kebutuhan.

## 2.7 Analisis Robust Biplot

Analisis biplot adalah peragaan secara grafik dari baris dan kolom sebuah matriks data, dengan baris mewakili objek dan kolom mewakili peubah. Adapun yang diperoleh gambaran dari analisis biplot adalah kedekatan antar objek, keragaman peubah, korelasi antar peubah, dan keterkaitan peubah dengan objek (Matjjik dan Sumertajaya, 2011). Pada analisis *robust* biplot, analisis biplot dapat dibangkitkan dengan menggunakan matrik *covariance* yang *robust*. Analisis

biplot dengan menggunakan matriks *covariance* yang *robust* dilakukan dengan menduga *eigen value* dan *eigen vektor* kiri ( $U$ ) dan kanan ( $A$ ) sehingga hasil duga tersebut tahan terhadap data *outlier* (Hawkins *et al.*, 2001).

Gabriel (1971) mengemukakan bahwa ukuran pendekatan matriks  $X$  dengan biplot dalam bentuk :

$$\rho^2 = \frac{(\lambda_1 + \lambda_2)}{\sum_{k=1}^r \lambda_k}$$

Dengan  $\lambda_1$  adalah nilai eigen terbesar ke-1,  $\lambda_2$  adalah nilai eigen terbesar ke-2 dan  $\lambda_k, k = 1, 2, \dots, r$  adalah nilai eigen ke- $k$ . Apabila  $\rho^2$  mendekati nilai satu, maka biplot memberikan penyajian yang semakin baik mengenai informasi data yang sebenarnya. Biplot merupakan upaya membuat gambar di ruang berdimensi banyak menjadi gambar di ruang berdimensi dua. Pereduksian dimensi ini mengakibatkan menurunnya informasi yang terkandung dalam biplot. Biplot yang dapat memberikan informasi sebesar 70% dari seluruh informasi dianggap cukup (Mattjik dan Sumertajaya, 2011).

### 2.8 Mean Square Error (MSE)

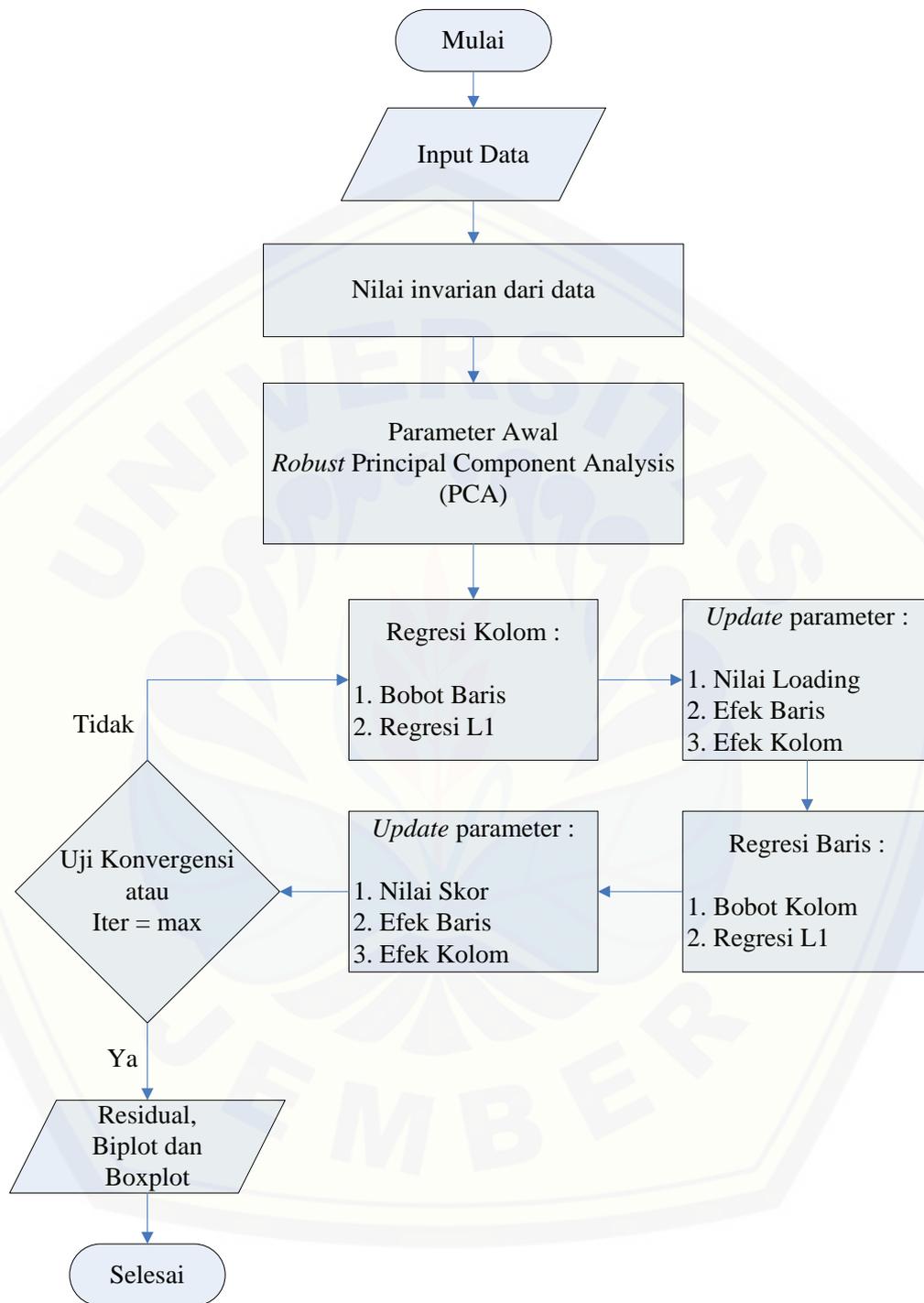
Terdapat beberapa metode yang biasa digunakan untuk menguji model manakah yang lebih baik. Namun terkadang dua metode yang berbeda dapat memberikan hasil yang berbeda, sehingga tidak ada ketetapan khusus metode manakah yang lebih baik dalam mengukur kecocokan model. MSE adalah rata-rata kuadrat *error* yang merupakan selisih antara sesuatu yang diduga dan penduganya. Masing-masing *error* atau sisaan dikuadratkan, kemudian dijumlahkan dan dibagi dengan jumlah observasi. MSE dapat digunakan untuk menentukan model yang paling baik dengan hasil dari MSE memiliki nilai yang cenderung kecil. Secara matematis MSE dapat dituliskan sebagai berikut :

$$MSE = \frac{\sum (y - \hat{y})^2}{n}$$

dengan  $y$  merupakan variabel respon,  $\hat{y}$  merupakan prediktor dan  $n$  adalah banyaknya data. Model dengan nilai MSE yang mendekati nol mengindikasikan bahwa model tersebut merupakan model yang lebih baik.

## 2.9 Implementasi Algoritma pada S-plus

*Script robust* faktor analitik yang digunakan untuk analisis data pada penelitian Croux *et al.* (2003) dengan menggunakan *software* S-plus adalah *weight.w11*, *prcomp.rob* dan *twoway.rob*. Ketiga *script* tersebut dijalankan secara bersamaan dikarenakan *script* *weight.w11* dan *prcomp.rob* merupakan bagian dari *script* *twoway.rob*. *Script* *weight.w11* berfungsi untuk menghitung nilai bobot dari baris maupun kolom data sebelum iterasi dilakukan. *Script* *prcomp.rob* berfungsi untuk menghitung nilai awal skor yang dipergunakan pada proses regresi kolom. Sedangkan *script* *twoway.rob* merupakan *script* yang digunakan untuk memodelkan data ke dalam bentuk tabel dua arah yang *robust* terhadap *outlier*. Berikut ini merupakan bagan alur model *robust* faktor analitik untuk penanganan *outliers* pada tabel dua arah menggunakan *software* S-plus :



Gambar 2.1 Bagan Alur Model *Robust* Faktor Analitik pada S-plus

Keterangan :

-  : sebagai awal dan akhir dari suatu program
-  : menunjukkan bagian arah intruksi yang dijalankan
-  : data yang akan dimasukkan ke dalam proses algoritma atau data yang dihasilkan dari proses algoritma
-  : satu atau beberapa himpunan perintah yang akan dilaksanakan secara beruntun
-  : memutuskan arah atau percabangan yang diambil sesuai dengan kondisi yang dipenuhi, yaitu benar atau salah

Bagan diatas menunjukkan bahwa alur model *robust* faktor analitik untuk penanganan *outliers* pada tabel dua arah menggunakan *software* S-plus dimulai dari menginputkan data. Dari data yang diinputkan dicari nilai *invarian* data atau nilai awal efek baris dan nilai tengah dari data berdasarkan barisnya. Kemudian dicari nilai awal dari skor menggunakan *Robust Principal Component Analysis* (RobPCA) atau *Robust Analisis Kumponen Utama* (AKU) dengan fungsi `f <- prcomp.rob(x, k, sca = "A", scores = T)$scores[, 1:k]`. Pada fungsi tersebut nilai skor yang diambil sepanjang kolom 1 sampai *k*. Selain itu juga didefinisikan nilai objektif sebagai nilai awal yang digunakan untuk uji kekonvergenan. Nilai awal yang diperoleh digunakan untuk mencari nilai regresi dari kolom. Pada tahap ini terlebih dahulu dihitung bobot baris dari data menggunakan fungsi `weightc1 <- weight.w11(f, n, k)` yang *downweight outlier*. Nilai bobot yang diperoleh merupakan salah satu syarat untuk melakukan regresi L1 sehingga menjadi regresi L1 yang terboboti.

Hasil dari regresi L1 yang terboboti merupakan nilai dari penduga *loading*. Setelah regresi L1 yang terboboti berdasarkan kolom selesai, dilakukan pembaruan parameter baru yang digunakan untuk regresi pada baris. Parameter yang diperbarui diantaranya adalah nilai penduga *loading*, efek baris, efek kolom, nilai skala dan rata-rata keseluruhan. Regresi pada baris dimulai dengan menghitung bobot dari kolom dengan fungsi `weightr1 <- weight.w11(1, p, k)` yang *downweight outlier*. Bobot dari kolom yang diperoleh digunakan

untuk regresi L1 pada kolom data sehingga menjadi regresi L1 yang terboboti. Fungsi dari regresi L1 yang terboboti didefinisikan sebagai `sigmfreg <- llfit (weightr1 * cbind (vecpl,1) , weightr1 * (x [i , ]- ms - b) , intercept=F)$coef`. Nilai yang dihasilkan merupakan nilai dugaan skor. Sama seperti regresi pada kolom, setelah selesai dilakukan regresi pada baris maka dilakukan pembaruan pada parameternya. Parameter yang diperbarui adalah diantaranya nilai penduga skor, efek baris, efek kolom, nilai dari skala dan nilai rata-rata keseluruhan.

Dinilai satu iterasi apabila sudah dilakukan proses regresi pada kolom dan regresi pada baris. Regresi pada kolom dan pada baris tidak dapat dilakukan secara bersamaan, nilai hasil regresi pada kolom digunakan untuk proses regresi pada baris dan hasil dari regresi baris digunakan untuk proses regresi pada kolom. Proses ini dinamakan *Robust Alternating Regression* (RAR). Setiap akhir dari iterasi dilakukan perhitungan terhadap nilai residual. Nilai residual digunakan untuk menentukan nilai objektif yang baru sebagai nilai yang akan di uji kekonvergenannya. Apabila sudah konvergen iterasi dihentikan dan apabila belum konvergen, iterasi akan berlanjut hingga iterasi maksimal. Keluaran dari *script* yang dihasilkan yaitu nilai rata-rata keseluruhan, efek baris, efek kolom, nilai skala, penduga *loading*, penduga skor, residual, bobot baris, bobot kolom, *Mean Square Error* (MSE), nilai eigen, nilai kekonvergenan, jumlah iterasi, model, metode dan fungsi pemanggil *script*. Keluaran lainnya dari *script* adalah dalam bentuk plot diantaranya biplot dan boxplot. Biplot digambarkan berdasarkan nilai residualnya dan boxplot diperoleh dari nilai efek baris dan efek kolomnya (Croux *et al*, 2003).

### BAB 3. METODE PENELITIAN

#### 3.1 Data

Data yang digunakan dalam penelitian ini adalah data penelitian perbenihan padi di beberapa kabupaten di Pulau Jawa yang digunakan oleh Aini (2015). Pada penelitian ini menggunakan 10 varietas dan 8 lokasi percobaan. Varietas padi yang diteliti didefinisikan dengan G1, G2, G3, G4, G5, G6, G7 dan G8. Sedangkan lokasi percobaannya didefinisikan dengan L1, L2, L3, L4, L5, L6, L7, L8, L9 dan L10.

#### 3.2 Desain Penelitian

Data yang diperoleh diolah dan dianalisis dengan menggunakan S-plus dan program R. Metode analisis uji kebaikan model menggunakan *Mean Square Error* (MSE) yang dibandingkan nilai MSE-nya antara analisis dengan menggunakan S-plus 8.0 dan analisis dengan menggunakan program R 2.7.0. Alur penelitian pada tugas akhir ini dapat dijelaskan sebagai berikut :

a. *Script* Program

*Script* program yang digunakan pada penelitian ini diperoleh dari <http://www.statistik.tuwien.ac.at/public/filz/programs.html> yang dibuat oleh Croux *et al.* (2003). Terdapat tiga *script* yang digunakan yaitu *weight.w11*, *prcomp.rob* dan *twoway.rob* yang ketiganya merupakan *script* pada S-plus. Ketiga *script* tersebut akan diimplementasikan pada program R 2.7.0.

b. Implementasi *Script*

*Script* yang diperoleh dari *weight.w11*, *prcomp.rob* dan *twoway.rob* di implementasikan pada program R. Sintaks maupun fungsi-fungsi yang terdapat di dalam *script* disesuaikan dengan kondisi masing-masing program tetapi tetap mendefinisikan perintah yang sama.

c. Menginputkan Data

d. Analisis Data Menggunakan S-plus

Pada tahap ini data yang ada akan diolah berdasarkan *script* *twoway.rob*. *Script* *twoway.rob* membutuhkan *script* *weight.w11* dan *prcomp.rob* pada saat program dijalankan. Dengan kata lain, *script* *weight.w11* dan *prcomp.rob* merupakan bagian dari *script* *twoway.rob* pada saat di jalankan pada S-plus 8.0. *Script* *twoway.rob* merupakan *script* yang digunakan untuk menentukan model *robust* faktor analitik untuk penanganan *outliers* pada tabel dua arah. *Script* *weight.w11* untuk mencari nilai bobot baris maupun bobot kolom data sedangkan *script* *prcomp.rob* untuk mencari nilai awal skor sebelum iterasi.

e. Analisis Data Menggunakan Program R

Pada tahap ini data diolah dengan *script* *twoway.rob* yang telah disesuaikan dengan program R 2.7.0. Sama halnya dengan *script* S-plus, pada program R memerlukan *script* *weight.w11* dan *script* *PcaProj*. *Script* *twoway.rob* pada program R yang disesuaikan berdasarkan *script* *Croux et al.* (2003). *Script* *twoway.rob* merupakan *script* yang digunakan untuk menentukan model *robust* faktor analitik untuk penanganan *outliers* pada tabel dua arah. *Script* *weight.w11* untuk mencari nilai bobot baris maupun bobot kolom data sedangkan *script* *PcaProj* untuk mencari nilai awal skor sebelum iterasi.

f. Perbandingan *Mean Square Error* (MSE)

Model yang lebih baik dapat dilihat berdasarkan nilai MSE yang dihasilkan. Semakin kecil nilai MSE yang dihasilkan oleh model, menunjukkan bahwa model tersebut semakin baik.

g. Representasi Biplot

Pada tahap ini merepresentasikan hasil dari algoritma ke dalam bentuk biplot untuk menduga pengaruh interaksi dan menginterpretasikan model melalui biplot.

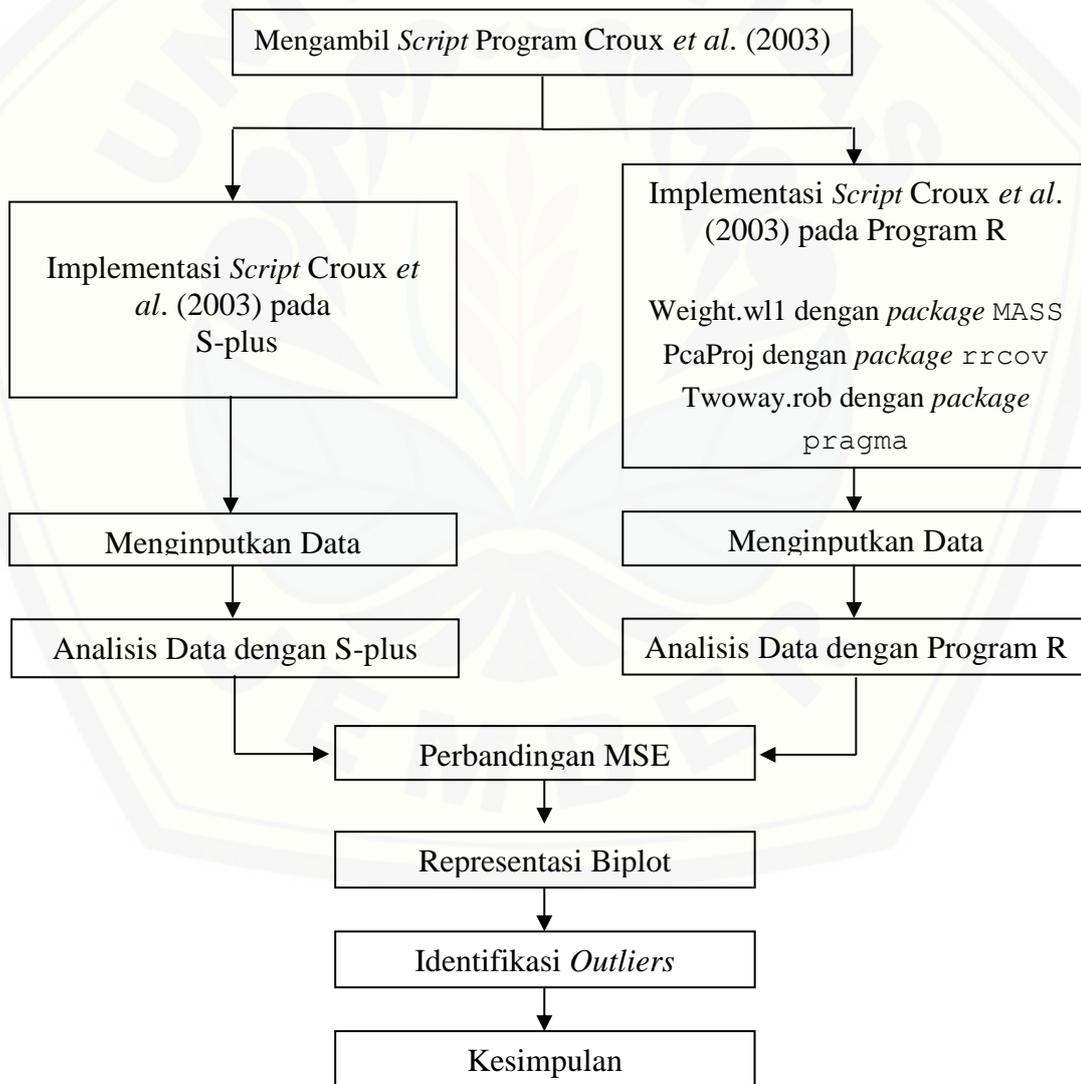
#### h. Identifikasi *Outliers*

Berdasarkan pembobot efek baris dan efek kolom yang diperoleh dari hasil iterasi, akan dibuat boxplot untuk menggambarkan ada tidaknya *outliers* pada data yang diolah.

#### i. Kesimpulan

Berdasarkan dari proses pengolahan data dapat diperoleh hasil evaluasi penanganan *outlier* pada data tabel dua arah menggunakan *robust* faktor analitik pada *software* S-plus dan program R.

Secara skematik, alur penelitian tugas akhir ini digambarkan dalam Gambar 3.1



Gambar 3.1 Skema Alur Penelitian

### 3.3 Struktur Fungsi Program R

Metode yang digunakan dalam penelitian ini adalah estimasi *Robust Alternating Regression* (RAR). Proses perhitungan dipermudah dengan menggunakan bantuan salah satu *software* dalam statistika yaitu Rstudio. Cara penggunaan Rstudio mirip dengan program R. Rstudio adalah *Integrated Development Environment* (IDE) utama bagi pengembang yang menggunakan bahasa R. Uniform Resource Locator (URL) Rstudio dapat diunduh pada link <http://www.rstudio.com/ide/download/desktop>. Fungsi R yang digunakan dalam proses analisis disesuaikan dengan algoritma yang telah dibuat oleh Croux *et al.* (2003). Fungsi R yang digunakan dalam penelitian ini diantaranya ialah :

- a. `as.matrix`, digunakan untuk mengubah data yang dimasukkan ke dalam bentuk matriks. Struktur fungsi `as.matrix` adalah sebagai berikut :  
`as.matrix(data)`
- b. `as.data.frame`, digunakan untuk menjadikan data yang dimasukkan sebagai kerangka data. Kerangka data ini yang menjadi struktur data fundamental yang digunakan pada proses pemodelan data. Struktur fungsi `as.data.frame` adalah sebagai berikut :  
`as.data.frame(data)`
- c. `biplot`, digunakan untuk memvisualisasi hasil analisis dalam bentuk plot dua dimensi. Struktur fungsi `biplot` adalah sebagai berikut :  
`biplot(x,y)`
- d. `boxplot`, digunakan untuk memvisualisasi efek baris dan efek kolom dengan plot berbentuk balok. Struktur fungsi `boxplot` adalah sebagai berikut :  
`boxplot(x)`
- e. `cov.rob`, digunakan untuk menghitung penduga skala dengan *hight breakdown point* menggunakan metode *Minimum Volume Ellipsoid*. Fungsi ini terdapat pada algoritma `weight.w11` dengan menggunakan *package MASS*. Struktur fungsi `cov.rob` adalah sebagai berikut :  
`cov.rob(data)`
- f. `PcaProj`, digunakan untuk analisis komponen utama yang *robust* dengan menggunakan metode *Projection Pursuit* (PP). Fungsi ini menggunakan *package rrcov*. Struktur fungsi `PcaProj` adalah sebagai berikut :

```
PcaProj(data, k = 0, kmax = ncol(data), scale=TRUE,  
na.action=na.fail, trace=FALSE)
```

- g. `l1linreg`, digunakan untuk menghitung nilai dugaan loading dan dugaan skor dengan regresi L1 (*Least Absolute Deviations*). Fungsi ini terdapat pada algoritma `twoway.rob` dengan menggunakan *package* `pragm`. Struktur fungsi `l1linreg` adalah sebagai berikut :

```
L1linreg(A, b, p = 1, tol = 1e-07, maxiter = 200)
```

