

36768/H109



ITS
Institut
Teknologi
Sepuluh Nopember



RTSE
519.536
Bas
i-1
2009

TESIS-SS09 2304

**IMPUTASI BERGANDA MENGGUNAKAN METODE REGRESI
DAN METODE *PREDICTIVE MEAN MATCHING*
UNTUK MENANGANI *MISSING DATA***

RAKHMAT BASUKI
NRP. 1308201029

DOSEN PEMBIMBING
Dr. Suhartono, S.Si, M.Sc
Yos Rusdiansyah, SE, MM

PROGRAM MAGISTER
JURUSAN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2009

PERPUSTAKAAN ITS	
Tgl. Terima	11-8-2009
Terima Dari	H
No. Agenda Prp.	354



TESIS-SS09 2304

**MULTIPLE IMPUTATION USING REGRESSION METHOD
AND PREDICTIVE MEAN MATCHING METHOD
TO HANDLE MISSING DATA**

RAKHMAT BASUKI
NRP. 1308201029

SUPERVISOR
Dr. Suhartono, S.Si, M.Sc
Yos Rusdiansyah, SE, MM

PROGRAM OF MAGISTER
DEPARTMENT OF STATISTICS
FACULTY OF MATHEMATICS AND NATURAL SCIENCES
INSTITUTE OF TECHNOLOGY SEPULUH NOPEMBER
SURABAYA
2009

**IMPUTASI BERGANDA MENGGUNAKAN METODE REGRESI
DAN METODE *PREDICTIVE MEAN MATCHING*
UNTUK MENANGANI *MISSING DATA***

TESIS

**Tesis disusun untuk memenuhi salah satu syarat memperoleh gelar
Magister Sains (M.Si)**

di

Institut Teknologi Sepuluh Nopember

Oleh :

RAKHMAT BASUKI

NRP : 1308 201 029

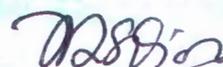
Tanggal Ujian : 3 September 2009

Periode Wisuda : Maret 2010

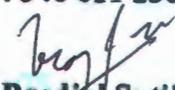
Disetujui Oleh :


1. **Dr. Suhartono, S.Si, M.Sc**
NIP. 132 135 220

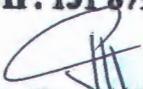
(Pembimbing I)


2. **Yos Rusdiansyah, SE, MM**
NIP. 340 011 236

(Pembimbing II)


3. **Dr. Brodjol Sutijo U, M.Si**
NIP. 131 879 383

(Penguji)


4. **Dr. Purbadi, M.Sc**
NIP. 131 652 051

(Penguji)


5. **Prof. Drs. Nur Iriawan, M.Ikom, Ph.D**
NIP. 131 782 011

(Penguji)


6. **Prof. Dra. Susanti Linuwih, M.Stat, Ph.D**
NIP. 130 368 808

(Penguji)



Direktur Program Pascasarjana,

Prof. Dr. Ir. Suparno, MSIE

NIP. 130 532 035

**IMPUTASI BERGANDA MENGGUNAKAN METODE REGRESI
DAN METODE *PREDICTIVE MEAN MATCHING*
UNTUK MENANGANI *MISSING DATA***

Nama Mahasiswa : Rakhmat Basuki
NRP : 1308201029
Pembimbing : Dr. Suhartono, S.Si, M.Sc
Co-pembimbing : Yos Rusdiansyah, SE, MM

ABSTRAK

Salah satu permasalahan yang sering dihadapi dalam survei adalah data yang hilang (*missing data*). Hal ini disebabkan oleh beberapa hal diantaranya adalah non respon, kelalaian saat melakukan perekaman data atau bahkan oleh hal lain yang tidak diketahui sebabnya. Jika *missing data* dibiarkan begitu saja maka inferensi statistik dengan menggunakan metode standar untuk data yang lengkap tidak dapat dilakukan, terlebih lagi jika jumlahnya banyak. Imputasi, yaitu mengisi *missing data* dengan sebuah nilai tertentu, merupakan salah satu solusi dalam menangani kasus ini. Dengan melakukan imputasi maka data yang lengkap akan bisa diperoleh sehingga analisa standar untuk data lengkap bisa dilakukan. Imputasi tunggal, yaitu mengisi setiap *missing data* dengan sebuah nilai tertentu ternyata tidak merefleksikan adanya ketidakpastian pada prediksi dari *missing data* tersebut dan dalam kasus tertentu menghasilkan estimasi varian dari parameter bias menuju nol. Oleh karena itu, imputasi berganda lebih tepat diterapkan karena mampu merefleksikan adanya ketidakpastian pada prediksi dari *missing data* tersebut. Metode regresi dan *predictive mean matching* merupakan dua metode yang bisa digunakan pada imputasi berganda. Pada penelitian ini kedua metode tersebut diuji melalui simulasi dan diterapkan pada data survei industri BPS kemudian hasilnya diperbandingkan. Pada data yang berdistribusi normal multivariat, kinerja kedua metode adalah sama baik. Namun, pada data yang tidak berdistribusi normal multivariat metode *predictive mean matching* menunjukkan kinerja yang lebih baik dibandingkan metode imputasi regresi.

Kata Kunci : Imputasi, *Missing data*, Metode Regresi, *Predictive Mean Matching*

MULTIPLE IMPUTATION WITH REGRESSION METHOD AND PREDICTIVE MEAN MATCHING METHOD TO HANDLE MISSING DATA

By : Rakhmat Basuki
Student Identity Number : 1308201029
Supervisor : Dr. Suhartono, S.Si, M.Sc
Co-Supervisor : Yos Rusdiansyah, SE, MM

ABSTRACT

Missing data is one of a common problem in survey. Its caused by many things as non response, errors and omissions when data recorded or by unknown reasons. When missing data are left blank, complete data statistics that would have been used in the absence of missing data can no longer be calculated and data analysts can no longer use standard complete data methods to draw inferences; especially when the number of missing is large. Imputation, its that impute the missing value with a real value is one of solution in this case. With imputation, the complete data will be derived so that standard analysis for a complete data is applicable. Single imputation, its that fill every missing data with a single value does not reflect the uncertainty about the predictions of the unknown missing values, and the resulting estimated variances of the parameter estimates will be biased toward zero. Instead of filling in a single value for each missing value, multiple imputation replaces each missing value with a set of plausible values that represent the uncertainty about the right value to impute. Therefore, multiple imputation is more appropriate. Regression method and predictive mean matching method are two methods which can used in multiple imputation. In this research, both of the methods are tested by simulation learning. Practical learning is applied to industrial survey data of BPS and then the output are compared. In multivariate normal condition, both of the method are work properly. When the normality assumption is violated, the predictive mean matching method yield imputed values that are more plausible and more appropriate than the regression method.

Key Words : Imputation, Missing Data, Regression Method, Predictive Mean Matching Method

KATA PENGANTAR

Alhamdulillahirobbil'alamiin, tiada kata yang patut penulis ucapkan kecuali puji syukur kehadiran Allah SWT, yang telah melimpahkan begitu banyak nikmat dan karunia-Nya sehingga penulis dapat menyelesaikan tesis dengan judul "***Imputasi Berganda Menggunakan Metode Regresi dan Metode Predictive Mean Matching untuk Menangani Missing Data***". Shalawat serta salam senantiasa penulis curahkan kepada Nabi Muhammad SAW beserta para keluarga, sahabat, serta orang-orang yang selalu *istiqomah* mengikui jalan-Nya hingga akhir zaman.

Selesainya tesis ini tidak lepas dari bantuan berbagai pihak, oleh karena itu penulis ingin menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Allah SWT, atas segala nikmat, ujian dan cobaan-Nya. Aku sebagai hamba-Mu banyak mendapat hikmah atas segala apa yang menimpaku.
2. Istriku tercinta, Nur Khumaedah atas do'a dan tetesan air mata dalam shalat malamnya yang senantiasa mendoakan bagi kebaikan penulis. Terima kasih juga atas motivasi dan nasihatnya. Ananda Nurul Izzah dan Abdullah Faqih, semoga Allah SWT merahmati dan mencintai kalian semua. Ma'afkan abimu ya nak.....!!!
3. Ibu, Bapak dan keluarga di Banjarnegara, yang senantiasa mendoakan untuk keberhasilan penulis.
4. Bapak Dr. Suhartono, M.Sc selaku pembimbing utama atas keuletan dan motivasinya dalam membimbing. Bagi saya, anda adalah dosen luar biasa yang terkadang juga membuat saya "sungkan" atas semua yang telah bapak berikan.
5. Bapak Yos Rusdiansyah, SE, MM selaku Co-Pembimbing yang telah sudi meluangkan waktu di sela-sela kesibukannya.
6. Bapak Dr.Sony Sunaryo, M.Si., selaku dosen wali sekaligus Ketua Jurusan Statistik, atas bantuannya.
7. Bapak Prof. Dr. Drs. I Nyoman Budiantara, M.Si., selaku ketua program studi Magister Jurusan Statistika ITS Surabaya, yang tidak bosan-bosannya mengingatkan untuk segera menyelesaikan tesisnya.
8. Para dosen pengajar dan staf Jurusan Statistika FMIPA ITS Surabaya, yang dengan tulus ikhlas telah memberikan bekal ilmu selama penulis mengikuti studi.
9. Para dosen penguji, terima kasih atas masukannya, pengalaman menegangkan yang tak terlupakan.

10. Kepala BPS yang telah memberikan kesempatan kepada penulis mengikuti studi di program Magister Jurusan Statistika ITS Surabaya.
11. Kepala BPS Propinsi Kalimantan Tengah dan Kepala BPS Kabupaten Sukamara yang telah memberikan izin dan kesempatan kepada penulis untuk mengikuti studi di ITS Surabaya.
12. Teman-teman seperjuangan S-2 BPS angkatan 3, terima kasih atas bantuan dan dukungannya. BERSAMA KITA BISA.
13. Kang Adam atas idenya; Mba Pre dan Mba Ari sebagai “co-pembimbing”.
14. Teman-teman seperjuangan di Kost Makam Atas; Bagus: sehat-sakit tetap bimbingan bersama; Pak Ripto: Makasih printernya; Sarip-Siswadi: makasih atas “kamar besarnya”, We: sama-sama *missing* ya kita ?? Amir: heebaat, ampuuuuhh.
15. Semua pihak yang membantu yang tidak dapat penulis sebutkan satu-persatu.

Penulis menyadari, tesis ini masih banyak kekurangan, saran dan kritik yang membangun sangat penulis harapkan. Semoga tesis ini dapat memberikan manfaat dan sumbangan untuk menambah wawasan keilmuan bagi kita semua.

Surabaya, September 2009

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN	iii
ABSTRAK	v
KATA PENGANTAR	ix
DAFTAR ISI	xi
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xv
DAFTAR LAMPIRAN	xvii
BAB 1 PENDAHULUAN	
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.3 Tujuan Penelitian	5
1.4 Manfaat Penelitian	6
1.5 Batasan Permasalahan	6
BAB 2 TINJAUAN PUSTAKA	
2.1 Asumsi pada Mekanisme <i>Missing Data</i>	7
2.2 Pola <i>Missing Data</i>	8
2.3 Imputasi Berganda dengan Metode Regresi	10
2.4 Imputasi Berganda dengan Metode <i>Predictive Mean Matching</i>	12
2.5 Mengkombinasikan Inferensi dari Kumpulan Data Hasil Imputasi	13
2.6 Efisiensi dari Imputasi Berganda	14
2.7 Uji Korelasi	15
2.8 Pengujian Koefisien regresi	15
2.9 Uji Kolmogorov Smirnov	16
2.10 Uji t untuk Data Berpasangan	17

BAB 3 METODOLOGI PENELITIAN	
3.1	Sumber Data 19
3.2	Variabel Penelitian 20
3.3	Metode Penelitian 21
3.3.1	Studi Simulasi 22
3.3.2	Imputasi Berganda dengan Metode Regresi 23
3.3.3	Imputasi Berganda dengan Metode <i>Predictive Mean Matching</i> 23
3.3.4	Membandingkan Efisiensi Relatif 23
BAB 4 ANALISIS DAN PEMBAHASAN	
4.1	Studi Simulasi 25
4.2	Deskripsi Data Hasil Survei Industri Besar dan Sedang 31
4.3	Imputasi <i>Missing Data</i> pada Hasil Survei Industri 36
4.3.1	KIP 17115 (kain tenun, ikat) 37
4.3.1.1	Model Regresi pada KIP 17115 43
4.3.1.2	Imputasi KIP 17115 dengan Metode Regresi 43
4.3.1.3	Imputasi KIP 17115 dengan Metode PMM 45
4.3.1.4	Perbandingan Hasil Imputasi 47
4.3.2	KIP 19201 (alas kaki) 49
4.3.2.1	Model Regresi pada KIP 19201 52
4.3.2.2	Imputasi KIP 19201 dengan Metode Regresi 52
4.3.2.3	Imputasi KIP 19201 dengan Metode PMM 54
4.3.2.4	Perbandingan Hasil Imputasi 56
BAB 5 KESIMPULAN DAN SARAN	
5.1	Kesimpulan 59
5.2	Saran 59
DAFTAR PUSTAKA 61	
LAMPIRAN 63	

DAFTAR TABEL

Tabel 2.1 : Pola <i>Missing Data</i> Monoton	8
Tabel 2.2 : Pola <i>Missing Data</i> Non-Monoton	9
Tabel 2.3 : Metode Imputasi pada PROC MI	9
Tabel 2.4 : Ilustrasi Data	10
Tabel 3.1 : Pengelompokan Golongan Industri Berdasarkan Jumlah Tenaga Kerja	20
Tabel 4.1 : Banyaknya Perusahaan yang Respon dan Non Respon	32
Tabel 4.2 : Banyaknya Variabel Produksi yang <i>Missing</i>	32
Tabel 4.3 : Observasi dengan Nilai Produksi yang <i>Missing</i>	37
Tabel 4.4 : Korelasi Pearson	38
Tabel 4.5 : Koefisien Regresi pada KIP 17115	38
Tabel 4.6 : ANOVA dengan Melibatkan Variabel Tenaga Kerja	39
Tabel 4.7 : ANOVA tanpa Melibatkan Variabel Tenaga Kerja	39
Tabel 4.8 : Perusahaan KIP 17115 yang dianggap sebagai <i>outlier</i>	41
Tabel 4.9 : Hasil Uji Kolmogorov Smirnov Variabel pada KIP 17115	42
Tabel 4.10 : Koefisien Regresi pada KIP 17115	43
Tabel 4.11 : Hasil Imputasi dengan Metode Regresi	44
Tabel 4.12 : Informasi Varian dari Imputasi Berganda	45
Tabel 4.13 : Koefisien Regresi pada Metode PMM	45
Tabel 4.14 : Hasil Imputasi dengan Metode PMM	46
Tabel 4.15 : Informasi Varian dari Imputasi Berganda	47
Tabel 4.16 : Hasil Imputasi dengan Metode Regresi dan PMM	47
Tabel 4.17 : Penghitungan nilai d_j dan d_j^2	48
Tabel 4.18 : Korelasi Pearson pada Variabel KIP 19201	50
Tabel 4.19 : Output Regresi KIP 19201	50
Tabel 4.20 : ANOVA dengan Melibatkan Variabel Bahan Bakar dan Listrik	51

Tabel 4.21 : ANOVA tanpa Melibatkan Variabel Bahan Bakar dan Listrik	51
Tabel 4.22 : Observasi dengan Nilai Produksi <i>Missing</i>	51
Tabel 4.23 : Hasil Uji Kolmogorov Smirnov Residual KIP 19201	52
Tabel 4.24 : Koefisien Regresi pada KIP 19201	53
Tabel 4.25 : Hasil Imputasi dengan Metode Regresi	53
Tabel 4.26 : Informasi Varian dari Imputasi Berganda	54
Tabel 4.27 : Koefisien Regresi pada Metode PMM	54
Tabel 4.28 : Hasil Imputasi dengan Metode PMM	55
Tabel 4.29 : Informasi Varian dari Imputasi Berganda	55
Tabel 4.30 : Hasil Imputasi dengan Metode Regresi dan PMM	56
Tabel 4.31 : Output Uji Kolmogorov-Smirnov	57

DAFTAR GAMBAR

Gambar 3.1 : Diagram Alur Penelitian	24
Gambar 4.1 : Plot Data Lengkap	26
Gambar 4.2 : Plot Data setelah dimissingkan sebanyak 50%	26
Gambar 4.3 : Grafik Perbandingan MSE_S Metode Regresi dan Metode PMM pada simulasi Skenario ke-1	27
Gambar 4.4 : Grafik Perbandingan MSE_S Metode Regresi dan Metode PMM pada simulasi Skenario ke-2	28
Gambar 4.5 : Grafik Perbandingan MSE_S Metode Regresi dan Metode PMM pada simulasi Skenario ke-3	29
Gambar 4.6 : Grafik Perbandingan MSE_S Metode Regresi dan Metode PMM pada simulasi Skenario ke-4	30
Gambar 4.7 : Grafik Perbandingan MSE_S Metode Regresi dan Metode PMM pada simulasi Skenario ke-5	31
Gambar 4.8 : Plot Hubungan antara variabel produksi dengan jumlah tenaga kerja, bahan baku, tenaga listrik serta bahan bakar pada KIP 15121	33
Gambar 4.9 : Plot Hubungan antara variabel produksi dengan jumlah tenaga kerja, bahan baku, tenaga listrik serta bahan bakar pada KIP 15122	34
Gambar 4.10 : Plot Hubungan antara variabel produksi dengan jumlah tenaga kerja, bahan baku, tenaga listrik serta bahan bakar pada KIP 19201	35
Gambar 4.11 : Plot Hubungan antara variabel produksi dengan jumlah tenaga kerja, bahan baku, tenaga listrik serta bahan bakar pada KIP 17115	36
Gambar 4.12 : Plot Residual pada KIP 17115 sebelum <i>outlier</i> dikeluarkan	40
Gambar 4.13 : Plot Residual pada KIP 17115 setelah <i>outlier</i> dikeluarkan	41
Gambar 4.14 : <i>Scatter Plot</i> Uji Normal Multivariat KIP 17115	42
Gambar 4.15 : <i>Scatter Plot</i> Uji Normal Multivariat KIP 19201	57

DAFTAR LAMPIRAN

Lampiran 1 : Plot Hubungan antara variabel produksi, jumlah tenaga kerja, bahan baku, tenaga listrik, serta bahan bakar Menurut Kode Klasifikasi Industri	63
Lampiran 2 : Data Perusahaan KIP 17115 yang Memberikan Data Lengkap	79
Lampiran 3 : Daftar Perusahaan KIP 17115 Setelah Seleksi <i>Outlier</i>	80
Lampiran 4 : Makro QQ-Test	81
Lampiran 5 : Output Imputasi Menggunakan Metode Regresi pada KIP 17115	82
Lampiran 6 : Output Imputasi Menggunakan Metode PMM pada KIP 17115	84
Lampiran 7 : Output Regresi Dari 29 Perusahaan KIP 17115 Dengan Data Lengkap	86
Lampiran 8 : Output Regresi Dari 54 Perusahaan KIP 19201 Dengan Data Lengkap	87
Lampiran 9 : Output Regresi dan Uji Asumsi Standar terhadap 29 Perusahaan yang memberikan data lengkap	91
Lampiran 10 : Daftar Perusahaan KIP 19201 yang Memberikan Data Lengkap	97
Lampiran 11 : Output Imputasi Menggunakan Metode Regresi pada KIP 19201	98
Lampiran 12 : Output Imputasi Menggunakan Metode PMM pada KIP 19201	100
Lampiran 13 : Macro Imputasi Berganda Metode Regresi menggunakan SAS 9,0	102
Lampiran 14 : Macro Imputasi Berganda Metode Regresi menggunakan SAS 9,0	103
Lampiran 14 : Hasil Penghitungan MSE_s dari ke-5 Skenario	104

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Salah satu permasalahan yang sering dihadapi dalam survei adalah data yang hilang (*missing data*). Hal ini disebabkan oleh beberapa hal diantaranya adalah non respon, kelalaian saat melakukan perekaman data atau bahkan oleh hal lain yang tidak diketahui sebabnya (Qin, Zhang, Zhu, Zhang dan Zhang, 2009). Jika *missing data* dibiarkan begitu saja maka inferensi statistik dengan menggunakan metode standar untuk data lengkap tidak dapat dilakukan, terlebih lagi apabila jumlahnya banyak (Rubin, 1988).

Dalam berbagai survei, banyak faktor yang menyebabkan sampel yang mengandung *missing data* tidak bisa dilakukan penggantian sampel. Keterbatasan biaya, tingkat kesulitan yang tinggi serta keterbatasan jumlah sampel yang memungkinkan untuk diobservasi adalah beberapa faktor yang biasanya muncul. Untuk mengatasi permasalahan tersebut, metode yang umum digunakan adalah mengisi *missing data* tersebut dengan memasukkan sebuah nilai tertentu, inilah yang disebut dengan imputasi. Dengan imputasi, *missing data* yang ada bisa diisi agar diperoleh data lengkap (tidak terdapat *missing data*) sehingga analisa standar untuk data lengkap bisa dilakukan (Pedlow, Luke dan Blumberg, 2007).

Analisa statistik dengan *missing data* telah tercatat dalam literatur lebih dari 70 tahun yang lalu. Walks (1932) memulai studi pada *maximum likelihood estimation* untuk model multivariat normal pada data tidak lengkap. Sesudah itu, diskusi-diskusi tentang topik ini terus berlanjut. Referensi yang berguna untuk *parametric statistical inferences* dengan *missing data* secara umum dapat ditemukan dalam (Little dan Rubin, 1987). Berbagai prosedur kemudian diusulkan untuk melakukan imputasi, diantaranya adalah menggunakan metode yang paling sederhana yaitu dengan memasukkan nilai rata-rata dari variabel tersebut untuk mengisi *missing data*. Metode ini dikenal dengan nama imputasi "hot deck" (Madow, Olkin, Ford dan Rubin, 1983).

Prosedur lain yang bisa digunakan adalah dengan memodelkan variabel lain yang berkaitan yang terekam dalam survei untuk memprediksi *missing data* tersebut. Sebagai contoh ketika data penghasilan dari seorang responden tidak diketahui, model regresi dengan menggunakan karakteristik demografi seperti umur, jenis kelamin, pendidikan dan jabatan dari responden tersebut bisa digunakan untuk mengestimasi besarnya penghasilan (Rubin, 1988).

Imputasi tunggal yaitu mengisi dengan sebuah nilai tertentu pada setiap *missing data* ternyata tidak merefleksikan adanya ketidakpastian pada prediksi dari *missing data* tersebut. Pada kasus tertentu, misalnya ketika imputasi tunggal dilakukan dengan mengisi nilai rata-rata pada setiap *missing data* maka akan menghasilkan estimasi varian dari parameter yang bias menuju nol. Sebagai gantinya, Rubin (1987) mengenalkan sebuah prosedur yang disebut dengan prosedur imputasi berganda. Prosedur ini pada dasarnya adalah memasukkan sekumpulan data yang menggambarkan ketidakpastian dari nilai yang akan diimputasikan pada setiap *missing data*. Data set hasil imputasi ini kemudian dianalisa dengan menggunakan prosedur standar untuk data lengkap dan mengkombinasikan data set hasil imputasi tersebut untuk analisa lebih lanjut.

Dalam prakteknya, menurut Yuan (2001) inferensi imputasi berganda memuat tiga tahapan yang berbeda yaitu:

- 1 *missing data* dimasukkan sebanyak m kali untuk menghasilkan m buah data set yang lengkap,
- 2 m buah data set yang lengkap tersebut dianalisa dengan menggunakan prosedur-prosedur standar untuk data lengkap,
- 3 hasil dari m buah data lengkap tersebut kemudian digunakan untuk inferensi.

Magnani (2004) telah meninjau *missing data techniques* (MDTs) yang utama yaitu metode konvensional, imputasi global, imputasi lokal serta estimasi parameter. Pada metode konvensional, prosedur yang diambil adalah *listwise deletion* yaitu membuang observasi yang mengandung *missing data* dan hanya menggunakan observasi dengan data lengkap (tanpa *missing data*). Imputasi global pada prinsipnya adalah mengisi setiap *missing data* dengan menggunakan ukuran pemusatan data seperti median, modus atau rata-ratanya. Imputasi lokal menggunakan prosedur *hot deck imputation* yaitu mengisi *missing data* pada

sebuah observasi dengan data yang berhasil direkam pada observasi lain; sedangkan estimasi parameter dilakukan menggunakan *Expectation Maximization (EM) algorithm* menggunakan fungsi densitas dari sampel yang tidak lengkap.

Magnani telah berupaya untuk menyoroiti kekurangan dan kelebihan dari berbagai jenis mekanisme *missing data* tersebut. Sebagai contoh, Magnani mengungkapkan bahwa metode statistik sebagian besar telah dibangun dalam manajemen data survei dan terbukti sangat efektif dalam berbagai situasi. Meskipun demikian, diperlukan kehati-hatian dan ketelitian dalam melakukan imputasi dari berbagai metode tersebut.

Allison (2000) telah mengevaluasi dua algoritma untuk memproduksi imputasi berganda untuk *missing data* menggunakan data simulasi dengan software SOLAS. Software menggunakan *propensity score classifier* dengan *Approximate Bayesian Bootstrap (ABB)* menghasilkan estimasi dari koefisien regresi dengan bias yang buruk saat data pada variabel prediktor adalah *Missing at Random (MAR)* atau *Missing Completely at Random (MCAR)*. Allison juga telah menunjukkan bahwa *listwise deletion* menghasilkan estimasi regresi yang tidak bias pada saat mekanisme *missing data* tergantung hanya pada variabel prediktor, tidak pada variabel respon.

Metode imputasi berbasis kernel telah dikaji oleh Zhang, Qin, Zhu, Zhang dan Zhang (2007) serta Qin,dkk. (2009). Dari hasil kajian mereka, metode imputasi tersebut menghasilkan inferensi yang baik pada berbagai parameter, namun memerlukan perhitungan yang cukup kompleks.

Yuan (2001) juga telah meninjau tiga metode imputasi berganda untuk *missing data* yaitu metode regresi, *propensity score method* dan *MCMC method*. Metode regresi dilakukan dengan memperhatikan hubungan diantara variabel-variabel lain yang berhasil diobservasi. Pada *Propensity score method*, setiap variabel dengan *missing data* dibuat *propensity score* pada setiap observasi untuk mengestimasi peluang observasi tersebut *missing*. Observasi yang ada dikelompokkan berdasarkan *propensity score* tersebut kemudian imputasi dengan menggunakan *approximate bayesian bootstrap* diterapkan pada setiap kelompok. *MCMC method* menggunakan pendekatan bayesian dalam proses imputasinya. Dari ketiga data hasil imputasi menggunakan ketiga metode tersebut kemudian

digunakan metode statistik standar untuk mengevaluasi efisiensi dari imputasi berganda. Perhitungan Efisiensi Relatif (ER) yang dipergunakannya banyak diadopsi untuk mengevaluasi efisiensi dari sebuah metode imputasi.

Berbeda dengan *propensity score method* yang tidak memperhatikan hubungan diantara variabel lain (Schafer, 1999), imputasi berganda menggunakan metode regresi dilakukan dengan memperhatikan hubungan diantara variabel-variabel lain yang berhasil diobservasi. Kemudahan dalam menginterpretasikan hubungan antar variabel dari model yang terbentuk menyebabkan metode regresi menjadi salah satu metode imputasi yang banyak digunakan.

Metode *predictive mean matching* pada dasarnya adalah sama dengan metode regresi, yang membedakan adalah untuk setiap nilai yang hilang (*missing*) diimputasikan dari nilai observasi yang terdekat dari model (Rubin, 1987). Metode ini menjamin bahwa nilai-nilai yang diimputasikan adalah masuk akal dan kemungkinan akan lebih tepat dibandingkan metode regresi apabila asumsi kenormalan tidak terpenuhi (Horton dan Lipsitz, 2001).

Salah satu tugas rutin Badan Pusat Statistik (BPS) adalah melaksanakan survei tahunan perusahaan industri besar dan sedang. Data tersebut sangat berguna untuk berbagai analisa di bidang industri serta mengetahui perkembangan sektor industri di Indonesia dari waktu ke waktu. Dalam survei tersebut pemasukan dokumen untuk diolah tidak pernah mencapai 100 persen. Dalam survei tahun 2003 tingkat pemasukan dokumen adalah sebesar 73,53 persen. Pada tahun 2004 tingkat pemasukan dokumen turun menjadi 70,32 persen. Oleh karena metode pengumpulan data yang digunakan adalah pencacahan lengkap (*sensus*), maka terhadap perusahaan-perusahaan yang tidak masuk dokumennya dilakukan estimasi (Badan Pusat Statistik, 2005).

Secara umum ada dua kategori perusahaan non respon yaitu perusahaan lama dan perusahaan baru. Perusahaan lama non respon artinya perusahaan yang data tahun sebelumnya sudah ada tapi non respon pada tahun saat dilakukan survei. Perusahaan baru non respon adalah perusahaan yang baru pernah disurvei pada tahun bersangkutan dan nama serta alamatnya telah diperoleh saat *listing*, namun non respon.

Estimasi perusahaan lama non respon dilakukan terhadap seluruh variabel yang ada datanya pada tahun sebelumnya. Metode estimasi terhadap kelompok perusahaan ini dilakukan dengan menggunakan informasi yang diperoleh dari pertumbuhan (*growth*) nilai tambah, sedangkan variabel lainnya dilakukan estimasi dengan memperhatikan rasio antar variabel perusahaan yang respon. Bagi perusahaan baru non respon, estimasi dilakukan dengan cara imputasi data dari perusahaan lain, dengan menggunakan rasio dari perusahaan yang respon yang sama kode Klasifikasi Baku Lapangan Usaha Indonesia (KBLI) 5 digitnya atau 3 digitnya (BPS, 2005).

Pada perusahaan lama yang respon ternyata tidak sedikit terjadi kasus *data missing*. Dari kasus tersebut yang sering muncul adalah data produksi. Untuk mengatasi masalah ini bisa dilakukan dengan pendekatan menggunakan data lain yang tersedia yang berkaitan dengan produksi. Oleh karena itu, metode imputasi merupakan salah satu alternatif yang tepat guna membantu mengatasi permasalahan tersebut.

1.2 Rumusan Masalah

Berdasarkan uraian dan latar belakang di atas, maka masalah dalam penelitian ini adalah:

1. Bagaimana melakukan imputasi berganda dengan menggunakan metode regresi dan metode *predictive mean matching* pada *missing data* yang terdapat pada hasil survei industri yang dilakukan oleh Badan Pusat Statistik
2. Bagaimana perbandingan hasil imputasi berganda antara metode regresi dengan metode *predictive mean matching*.

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, maka tujuan yang ingin dicapai dalam penelitian ini adalah:

1. Melakukan imputasi berganda dengan menggunakan metode regresi dan metode *predictive mean matching* pada *missing data* yang terdapat pada hasil survei industri yang dilakukan oleh Badan Pusat Statistik
2. Melakukan perbandingan hasil imputasi berganda antara metode regresi dengan metode *predictive mean matching*.

1.4 Manfaat Penelitian

Manfaat yang ingin dicapai dari hasil penelitian adalah:

1. Dapat digunakan oleh Badan Pusat Statistik sebagai metode alternatif untuk melengkapi *missing data* pada survei industri
2. Mengembangkan wawasan keilmuan dan ilmu pengetahuan mengenai imputasi data dan penerapannya.

1.5 Batasan Permasalahan

Ada berbagai metode imputasi yang dapat digunakan untuk menangani permasalahan yang berkaitan dengan *missing data* dan ada berbagai survei yang dilakukan secara periodik oleh Badan Pusat Statistik. Dalam penelitian ini ada dua batasan permasalahan utama yang digunakan, yaitu :

1. metode imputasi yang dikaji adalah metode regresi dan metode *predictive mean matching*
2. data survei yang diteliti sebagai studi kasus adalah data Survei Industri Besar dan Sedang Propinsi Jawa Timur tahun 2007.

BAB 2

TINJAUAN PUSTAKA

Pada bagian ini akan dijelaskan berbagai hal yang berkaitan dengan imputasi berganda, diantaranya adalah asumsi yang digunakan, pola *missing data*, metode imputasi regresi dan metode imputasi *predictive mean matching* serta efisiensi dari imputasi berganda.

2.1 Asumsi pada Mekanisme *missing data*

Little dan Rubin (1987) mengklasifikasikan mekanisme *missing data* dalam tiga kategori, yaitu :

1. *Missing completely at random* (MCAR); adalah kasus dimana data lengkap/ tanpa *missing* tidak dapat dibedakan dengan kasus yang memuat *missing data* (*incomplete data*). Pada kasus ini, nilai dari sebuah variabel dapat *missing* secara acak tanpa dipengaruhi oleh nilai-nilai dari variabel lainnya.
2. *Missing at random* (MAR); adalah kasus dimana terjadi perbedaan antara data yang memuat *missing data* (*incomplete*) dengan data yang tidak memuat *missing data*, namun pola *missing data* tersebut dapat dilacak atau dapat diprediksi dari variabel-variabel lain yang ada dalam *data base*. Pada kasus ini, nilai dari sebuah variabel yang *missing* dipengaruhi oleh nilai-nilai dari variabel lainnya. Misalnya sebuah observasi dengan variabel X_1 , X_2 dan X_3 , nilai dari variabel X_3 akan *missing* jika nilai dari variabel $X_2 > 2$. Dalam kasus ini, dapat dinyatakan bahwa X_3 adalah *missing* dengan asumsi MAR X_2 .
3. *Non ignorable*; adalah kasus dimana pola dari data yang hilang tidaklah random dan tidak dapat diprediksi dari variabel-variabel lain yang ada dalam *data base*. Pada kasus ini, nilai dari sebuah variabel yang *missing* tidak dipengaruhi oleh nilai-nilai dari variabel lainnya tetapi dipengaruhi oleh nilai dari variabel yang *missing* tersebut. Misalnya sebuah observasi dengan variabel X_1 , X_2 dan X_3 , nilai dari variabel X_3 akan *missing* jika nilai dari variabel $X_3 > 20$. Dalam kasus ini, dapat dinyatakan bahwa X_3 adalah *missing* dengan asumsi *non ignorable*.

2.2 Pola Missing Data

Dalam menentukan metode imputasi yang tepat, selain asumsi *missing data* juga perlu diperhatikan pola dari *missing data* tersebut. Perbedaan pola *missing data* akan menyebabkan perbedaan dalam pemilihan metode imputasi yang tepat untuk dipergunakan.

Di dalam blog www.support.sas.com dijelaskan tentang pola *missing data* dan metode imputasi yang direkomendasikan untuk digunakan. Berikut ini adalah penjelasan tentang pola *missing data*.

2.2.1 Pola Monoton

Sebuah data set dengan variabel $X_1, X_2, \dots, X_i, X_j, \dots, X_p$ dikatakan mempunyai pola *missing data* monoton ketika kejadian hilangnya variabel X_i pada sebuah individu diikuti dengan hilangnya semua variabel X_j , dimana $j > i$ pada individu tersebut. Dengan kata lain, jika variabel X_i pada sebuah individu dapat diobservasi (datanya terekam), maka diasumsikan bahwa semua variabel sebelumnya juga dapat diobservasi pada individu tersebut.

Tabel 2.1 Pola *Missing Data* Monoton

Observasi	Variabel				
	X_1	X_2	X_3	X_4	X_5
1	x	x	x	x	x
2	x	x	x	x	*
3	x	x	x	*	*

Keterangan :

(*) menunjukkan *missing data*, (x) menunjukkan data bisa diobservasi

2.2.2 Pola Non-Monoton

Sebuah data set dengan variabel $X_1, X_2, \dots, X_i, X_j, \dots, X_p$ dikatakan mempunyai pola *missing data* non-monoton apabila kejadian hilangnya variabel X_i pada sebuah individu tidak diikuti dengan hilangnya variabel X_j , dimana $j > i$ pada individu tersebut. Dengan kata lain ketika variabel X_i pada sebuah individu

dapat diobservasi (datanya terekam), maka tidak bisa diasumsikan bahwa semua variabel sebelumnya juga dapat diobservasi pada individu tersebut.

Tabel 2.2 Pola *Missing Data* Non-Monoton

Observasi	Variabel				
	X_1	X_2	X_3	X_4	X_5
1	x	x	x	x	x
2	x	x	*	*	x
3	x	x	x	x	*
4	x	*	x	*	x

Keterangan :

(*) menunjukkan *missing data*, (x) menunjukkan data bisa diobservasi

2.2.3 Pola *Arbitrary*

Sebuah data set dengan variabel $X_1, X_2, \dots, X_i, X_j, \dots, X_p$ dikatakan mempunyai pola *arbitrary* ketika kejadian hilangnya variabel X_i pada sebuah individu tidak mengikuti pola monoton maupun non-monoton.

Berdasarkan pola missing di atas dan tipe variabel yang akan diimputasi, diberikan rekomendasi seperti pada Tabel 2.3 berikut ini (www.support.sas.com):

Tabel 2.3 Metode Imputasi pada PROC MI

Pola <i>Missing Data</i>	Tipe Variabel Yang akan diimputasi	Metode yang Direkomendasikan
Monoton	Kontinyu	* Regresi
		* <i>Predictive Mean Matching</i>
		* <i>Propensity Score</i>
Monoton	Klasifikasi (Ordinal)	* Regresi Logistik
Monoton	Klasifikasi (Nominal)	* Metode Fungsi Diskriminan
<i>Arbitrary</i>	Kontinyu	* <i>MCMC Full-Data Imputation</i>
		* <i>MCMC Monotone-Data Imputation</i>

2.3 Imputasi Berganda dengan Metode Regresi

Imputasi berganda dengan menggunakan metode regresi merupakan salah satu metode yang bisa ditempuh ketika *missing data* yang akan diimputasi memiliki pola missing monoton serta tipe variabel yang akan diimputasi adalah kontinyu. Pada metode regresi, variabel X_i yang *missing* dianggap sebagai Y_i .

Sebagai ilustrasi berikut ini diberikan gambaran struktur data dari n buah responden yang terdiri dari n_0 responden yang tidak memberikan data lengkap (terdapat *missing data*) dan n_1 responden yang memberikan data lengkap :

Tabel 2.4 Ilustrasi Data

Responden	Variabel				
	X_1	X_2	X_3	X_4	Y_i
1	x	x	x	x	*
2	x	x	x	x	*
3	x	x	x	x	*
⋮	⋮	⋮	⋮	⋮	⋮
n_0	x	x	x	x	*
1	x	x	x	x	x
2	x	x	x	x	x
3	x	x	x	x	x
⋮	⋮	⋮	⋮	⋮	⋮
n_1	x	x	x	x	x

Keterangan :

(*) menunjukkan *missing data*, (x) menunjukkan data bisa diobservasi

Metode regresi merupakan metode yang umum digunakan dalam memprediksi variabel respon Y dengan menggunakan sekumpulan prediktor X . Dalam permasalahan ini

$$Y \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2) \quad (2.1)$$

dengan $\mathbf{X} = [1 \ X_1 \ X_2 \ X_3 \dots X_q]$ dan $\boldsymbol{\beta} = [\beta_0 \ \beta_1 \ \beta_2 \ \beta_3 \dots \beta_q]'$ merupakan spesifikasi dari $f(Y_i | X_i, \theta)$, $\theta = (\boldsymbol{\beta}, \log \sigma)$, dengan $\boldsymbol{\beta}$ adalah vektor komponen $q+1$. Dalam hal ini q adalah banyaknya variabel prediktor dan σ adalah skalar

dengan mengambil asumsi bahwa *conventional improper prior* untuk θ adalah $\Pr(\theta) \propto \text{const.}$ Selain itu diasumsikan juga bahwa $n_i > q + 1$ dengan n_i adalah banyaknya responden yang memberikan data lengkap.

Menurut Box dan Tiao (1973), secara empiris σ^2 merupakan hasil bagi antara $\hat{\sigma}_1^2 (n_i - q - 1)$ dengan variabel random berdistribusi $\chi_{n_i - q - 1}^2$. Selain itu, β diberikan σ^2 adalah berdistribusi normal dengan rata-rata $\hat{\beta}_1$ serta matrik varian kovarian $\sigma^2 \mathbf{V}$ pada terminologi statistik kuadrat terkecil berdasarkan pada n_i anggota responden yang memberikan data lengkap, sehingga bisa dituliskan

$$\hat{\sigma}_1^2 = \frac{\mathbf{Y}'\mathbf{Y} - \hat{\beta}_1' \mathbf{X}'\mathbf{Y}}{n_i - q - 1} \quad (2.2)$$

$$\hat{\beta}_1 = \mathbf{V}[\mathbf{X}'\mathbf{Y}] \quad (2.3)$$

$$\mathbf{V} = [\mathbf{X}'\mathbf{X}]^{-1} \quad (2.4)$$

Ketiga nilai pada persamaan (2.2), (2.3) dan (2.4) merupakan nilai-nilai yang diperoleh dengan menggunakan data dari responden yang memberikan data lengkap (responden pada n_i).

Tahap-tahap berikut dilakukan untuk menghasilkan nilai yang akan diimputasikan pada setiap imputasi :

1. Hitung nilai σ_*^2 dengan menggunakan persamaan:

$$\sigma_*^2 = \hat{\sigma}_1^2 (n_i - q - 1) / g \quad (2.5)$$

dengan $\hat{\sigma}_1^2$ adalah varian dari data lengkap

n_i adalah banyaknya responden yang memberikan data lengkap

g adalah variabel random yang dibangkitkan dari distribusi $\chi_{n_i - q - 1}^2$

2. Tentukan $q + 1$ buah variabel independen berdistribusi $N(0,1)$ untuk membuat $q + 1$ komponen vektor \mathbf{Z} , lalu hitung nilai β_* menggunakan persamaan:

$$\beta_* = \hat{\beta}_1 + \sigma_* [\mathbf{V}]^{1/2} \mathbf{Z} \quad (2.6)$$

dengan $[\mathbf{V}]^{1/2}$ adalah akar dari \mathbf{V} yang merupakan matrik segitiga atas yang diperoleh dengan menggunakan dekomposisi Cholesky.

3. Y_{mis} (data yang *missing*) kemudian digantikan dengan :

$$\hat{Y}_i = X_i \beta + z_i \sigma \quad (2.7)$$

dengan i adalah anggota responden yang terdapat *missing data* dan z_i adalah nilai berdistribusi normal (0,1) hasil simulasi.

Nilai yang akan diimputasikan pada Y_{mis} dilakukan dengan mengambil nilai yang baru bagi parameter σ^2 . Apabila ingin dilakukan imputasi sebanyak m kali, maka ketiga langkah di atas diulang sebanyak m kali secara independen (Rubin, 1987; halaman 166-167).

2.4 Imputasi Berganda dengan Metode *Predictive Mean Matching*

Metode *predictive mean matching* merupakan metode lain disamping metode regresi yang bisa digunakan ketika *missing data* yang akan diimputasi memiliki pola missing monoton serta tipe variabel yang akan diimputasi adalah kontinyu. Pada dasarnya metode ini adalah sama dengan metode regresi, yang membedakan adalah untuk setiap nilai yang hilang (*missing*) diimputasikan dari nilai observasi yang terdekat dari model (Rubin, 1987).

Tahap-tahap berikut dilakukan untuk menghasilkan nilai yang akan diimputasikan pada setiap imputasi:

1. Hitung nilai σ^2 dengan menggunakan persamaan:

$$\sigma^2 = \hat{\sigma}_1^2 (n_1 - q - 1) / g$$

dengan $\hat{\sigma}_1^2$ adalah varian dari data lengkap

n_1 adalah banyaknya responden yang memberikan data lengkap

g adalah variabel random yang dibangkitkan dari distribusi $\chi_{n_1 - q - 1}^2$

2. Tentukan $q + 1$ buah variabel independen berdistribusi Normal (0,1) untuk membuat $q + 1$ komponen vektor \mathbf{Z} , kemudian hitung nilai β menggunakan persamaan:

$$\beta = \hat{\beta}_1 + \sigma \cdot [\mathbf{V}]^{1/2} \mathbf{Z}$$

dengan $[\mathbf{V}]^{1/2}$ adalah akar dari \mathbf{V} yang merupakan matrik segitiga atas yang diperoleh dengan menggunakan dekomposisi Cholesky.

3. Y_{mis} (data yang *missing*) diprediksi dengan :

$$\hat{Y}_i = X_i \beta \quad (2.8)$$

dengan i adalah anggota responden yang terdapat *missing data*

4. Untuk setiap \hat{Y}_i dengan i adalah anggota responden yang terdapat *missing data*, ambil responden yang mempunyai nilai Y_i dengan i adalah anggota responden yang memberikan data lengkap yang nilainya paling dekat dengan nilai \hat{Y}_i , kemudian nilai tersebut diimputasikan pada data yang *missing*.

Metode *predictive mean matching* menjamin bahwa nilai-nilai yang diimputasikan adalah masuk akal dan kemungkinan akan lebih tepat dibandingkan metode regresi apabila asumsi kenormalan tidak terpenuhi (Horton dan Lipsitz, 2001).

2.5 Mengkombinasikan Inferensi dari Kumpulan Data Hasil Imputasi

Dengan melakukan m buah imputasi, dapat dihitung m buah set yang berbeda dari estimasi titik dan estimasi varian untuk parameter Q . Anggap \hat{Q}_i dan \hat{U}_i adalah estimasi titik dan estimasi varian dari data set hasil imputasi ke- i dimana $i = 1, 2, 3, \dots, m$. Kemudian estimasi titik untuk Q dari imputasi berganda adalah rata-rata dari m buah "data lengkap" yaitu :

$$\bar{Q} = \frac{1}{m} \sum_{i=1}^m \hat{Q}_i \quad (2.9)$$

Anggap \bar{U} adalah varian *within imputation* yang merupakan rata-rata dari m buah "data lengkap", estimasinya adalah :

$$\bar{U} = \frac{1}{m} \sum_{i=1}^m \hat{U}_i \quad (2.10)$$

B adalah varian *between imputation*, estimasinya adalah :

$$B = \frac{1}{m-1} \sum_{i=1}^m (\hat{Q}_i - \bar{Q})^2 \quad (2.11)$$

Kemudian estimasi varian yang berasosiasi dengan \bar{Q} adalah total dari varians *within imputation* dan *between imputation*; yaitu:

$$T = \bar{U} + \left(1 + \frac{1}{m}\right) B \quad (2.12)$$

Statistik dari $(Q - \bar{Q}) T^{-\frac{1}{2}}$ mengikuti distribusi t dengan derajat bebas v_m (Rubin, 1987), dengan

$$v_m = (m - 1) \left[1 + \frac{\bar{U}}{(1 + m^{-1})B} \right]^2 \quad (2.13)$$

2.6 Efisiensi dari Imputasi Berganda

Besarnya derajat bebas v_m tergantung pada m dan rasio dari r yang dapat dihitung dengan :

$$r = \frac{(1 + m^{-1})B}{\bar{U}} \quad (2.14)$$

r didefinisikan sebagai pertambahan relatif dari varians yang berkaitan dengan non respon (Rubin, 1987).

Saat tidak ada informasi *missing* tentang Q , maka kedua nilai dari r dan B adalah nol. Dengan nilai m yang besar dan nilai r yang kecil maka derajat bebas dari v_m akan menjadi besar sehingga distribusinya akan mendekati normal.

Statistik lain yang amat berguna berkaitan dengan non respon adalah *fraction* dari informasi *missing* pada Q . *Fraction* merupakan sebuah nilai yang berpengaruh terhadap kecepatan konvergensi terhadap suatu nilai. Semakin besar *fraction* pada informasi *missing* maka konvergensinya semakin lambat. *Fraction* dapat dicari dengan menggunakan persamaan:

$$\hat{\lambda} = \frac{r + 2/(v_m + 3)}{r + 1} \quad (2.15)$$

Efisiensi relatif (ER) adalah efisiensi yang diperoleh dengan menggunakan m buah imputasi terbatas (*finite*) dibandingkan dengan jumlah tidak terbatas (*infinite*). Pada kasus dengan hanya sedikit informasi *missing*, hanya diperlukan imputasi dengan jumlah yang kecil. Menurut Yuan (2001), nilai efisiensi relatif biasanya berkisar antara 80 sampai 100 persen. Efisiensi Relatif diperoleh dari fungsi m dan $\hat{\lambda}$.

$$ER = \left(1 + \frac{\hat{\lambda}}{m} \right)^{-1} \times 100 \% \quad (2.16)$$

Nilai ER yang besar menunjukkan bahwa metode imputasi yang digunakan memiliki efisiensi yang besar yang artinya dengan tidak terlalu banyak melakukan pengulangan, nilai imputasi yang dihasilkan sudah mendekati hasil imputasi yang dilakukan dengan jumlah tidak terbatas (*infinite*).

2.7 Uji Korelasi

Koefisien korelasi pada populasi (ρ) menyatakan ukuran hubungan linear antara dua variabel. Nilai ρ terletak antara -1 sampai dengan 1. Nilai ρ yang mendekati 1 atau -1 menunjukkan adanya hubungan yang kuat antara kedua variabel tersebut, sedangkan nilai ρ yang mendekati 0 menunjukkan hubungan yang lemah antara kedua variabel.

Untuk menguji koefisien korelasi pada populasi (ρ) dapat digunakan koefisien korelasi pada sampel (r). Prosedur pengujian hipotesisnya adalah sebagai berikut :

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

Statistik uji untuk uji hipotesis di atas adalah :

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} \quad (2.17)$$

Tolak H_0 apabila $|t| > t_{\alpha/2, n-2}$ yang artinya terdapat korelasi yang signifikan antara kedua variabel tersebut.

2.8 Pengujian Koefisien regresi

a. Uji Serentak

Uji serentak atau juga dikenal sebagai uji statistik F digunakan untuk menguji apakah semua variabel prediktor yang dimasukkan ke dalam model mempunyai pengaruh secara bersama-sama terhadap variabel respon. Pengujian hipotesisnya adalah :

$$H_0 : \beta_0 = \beta_1 = \beta_2 = \dots = \beta_q = 0$$

$$H_1 : \text{Paling tidak ada satu } \beta_i \neq 0 ; i = 1, 2, 3, \dots, q$$

Statistik uji F hitung dapat dihitung dengan persamaan :

$$F = \frac{SSR/q-1}{SSE/n-q} \quad (2.18)$$

dengan n adalah banyaknya sampel dan q adalah banyaknya variabel prediktor.

Tolak H_0 apabila $F > F_{\alpha, q-1, n-q}$ dan disimpulkan bahwa secara bersama-sama variabel prediktor berpengaruh terhadap variabel respon.

b. Uji Parsial

Uji parsial dilakukan untuk mengetahui variabel prediktor mana saja yang berpengaruh secara signifikan terhadap variabel respon secara individu. Walaupun secara simultan variabel prediktor berpengaruh secara signifikan terhadap variabel respon, namun secara individu belum tentu semua variabel prediktor tersebut berpengaruh secara signifikan. Untuk mengujinya dilakukan prosedur sebagai berikut :

$$H_0 : \beta_i = 0$$

$$H_1 : \beta_i \neq 0, i=1,2,3,\dots,q$$

Statistik uji t dapat dihitung dengan persamaan :

$$t = \frac{\hat{\beta}_i}{\sqrt{\hat{\sigma}^2 C_{ii}}} \quad (2.19)$$

dengan C_{ii} adalah elemen diagonal dari $(X'X)^{-1}$ yang berhubungan dengan $\hat{\beta}_i$.

Tolak H_0 jika $|t| > t_{\alpha/2, n-q-1}$ dan dapat disimpulkan bahwa variabel prediktor tersebut berpengaruh secara signifikan terhadap variabel responnya.

2.9 Uji Kolmogorov Smirnov

Uji Kolmogorov Smirnov dilakukan untuk menguji apakah sebuah variabel mengikuti distribusi normal atau tidak. Pengujian ini dilakukan dengan hipotesis sebagai berikut :

$$H_0 : F_{(x)} = F_{o(x)}$$

$$H_1 : F_{(x)} \neq F_{o(x)}$$

Statistik Uji yang digunakan adalah :

$$D = \text{Sup} |S_{(x)} - F_{o(x)}| \quad (2.20)$$

dengan : $S_{(x)}$ adalah proporsi nilai pengamatan sampel yang kurang dari sama dengan x

$F_{o(x)}$ adalah fungsi yang diketahui (normal)

Daerah penolakan :

Tolak H_0 jika $D \geq$ kuantil $1 - \alpha$ pada tabel Kolmogorov Smirnov dan disimpulkan bahwa variabel yang diuji tidak mengikuti distribusi normal.

2.10 Uji t untuk Data Berpasangan

Menurut Weiss (2002), pengujian ini dilakukan untuk menguji apakah terdapat perbedaan antara dua populasi dengan data yang berpasangan. Untuk setiap pasangan, d_j dinotasikan sebagai perbedaan antara nilai-nilai dari pasangan tersebut. Selanjutnya d_j disebut dengan variabel beda berpasangan.

$$d_j = x_{1j} - x_{2j} \quad (2.21)$$

dengan d_j adalah perbedaan antara nilai-nilai dari pasangan ke- j

x_{1j} adalah data ke- j pada populasi pertama

x_{2j} adalah data ke- j pada populasi kedua

Rata-rata dari beda berpasangan adalah sama dengan perbedaan antara rata-rata dari dua populasi. Diasumsikan bahwa beda berpasangan tersebut berdistribusi normal. Pengujian hipotesisnya adalah:

$$H_0 : \mu_d = 0$$

$$H_1 : \mu_d \neq 0$$

Uji statistik untuk persamaan (2.21) diatas adalah

$$t = \frac{\bar{d}}{s_d / \sqrt{n}} \quad (2.22)$$

dengan

$$\bar{d} = \frac{\sum_{j=1}^n d_j}{n} \quad (2.23)$$

dan

$$s_d = \sqrt{\frac{\sum_{j=1}^n d_j^2 - \left[\left(\sum_{j=1}^n d_j \right)^2 / n \right]}{n-1}} \quad (2.24)$$

adalah rata-rata dan varian sampel dari perbedaan tersebut. Keputusan untuk menolak H_0 apabila $|t| > t_{\alpha/2, n-1}$.

BAB 3 METODOLOGI PENELITIAN

3.1 Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder dari Survei Tahunan Perusahaan Industri Besar dan Sedang Propinsi Jawa Timur Tahun 2007. Survei ini dilakukan untuk memperoleh berbagai informasi di bidang perusahaan industri seperti banyaknya perusahaan, jumlah pekerja, pengeluaran untuk pekerja, tenaga listrik, bahan bakar yang digunakan, biaya input, nilai output, nilai tambah dan jual beli barang modal tetap.

Pelaksanaan survei industri ini dilakukan dengan cara memberikan daftar pertanyaan (*questionare*) kepada semua perusahaan industri yang tergolong besar dan sedang yang tercatat dalam direktori BPS. Jadi pencacahan dalam survei ini dilakukan secara lengkap. Perusahaan industri yang dimaksud dalam survei ini adalah suatu unit (kesatuan) usaha yang melakukan kegiatan ekonomi, bertujuan menghasilkan barang atau jasa, terletak pada suatu bangunan atau lokasi tertentu dan mempunyai catatan administrasi tersendiri mengenai produksi dan struktur biaya serta ada seorang atau lebih yang bertanggung jawab atas usaha tersebut.

Industri Pengolahan adalah suatu kegiatan ekonomi yang melakukan kegiatan mengubah suatu barang dasar secara mekanis, kimia atau dengan tangan sehingga menjadi barang jadi atau setengah jadi dan atau barang yang kurang nilainya menjadi barang yang lebih tinggi nilainya, dan sifatnya lebih dekat kepada pemakai akhir. Termasuk dalam kegiatan ini adalah jasa industri dan pekerjaan perakitan.

Perusahaan industri pengolahan dikelompokkan menjadi 4 (empat) yaitu perusahaan besar, sedang, kecil dan rumah tangga. Pengelompokan ini hanya didasarkan kepada banyaknya tenaga kerja yang bekerja, tanpa memperhatikan apakah perusahaan tersebut menggunakan mesin tenaga atau tidak, serta tanpa memperhatikan besarnya modal perusahaan tersebut.

Tabel 3.1

Pengelompokan Golongan Industri Berdasarkan Jumlah Tenaga Kerja

Golongan Industri	Banyaknya Tenaga Kerja
Besar	100 orang atau lebih
Sedang	20 – 99
Kecil	5 – 19
Rumah Tangga	1 – 14

Sumber : Statistik Industri Besar dan Sedang BPS, 2005

3.2 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini adalah :

1. Banyaknya pekerja/karyawan (X_1) yaitu banyaknya pekerja/karyawan rata-rata perhari kerja, baik pekerja yang dibayar maupun tidak dibayar selama tahun 2007. Tidak termasuk di dalamnya orang yang dibayar hanya berdasarkan komisi, orang yang bekerja sendiri seperti konsultan dan kontraktor serta pegawai yang bekerja di sektor bukan industri pengolahan seperti pegawai unit perkebunan dan pegawai unit pertambangan.
2. Bahan bakar dan pelumas (X_2) yaitu nilai (dalam ribuan) dari seluruh bahan bakar dan pelumas yang betul-betul dipakai selama tahun 2007 seperti bensin, solar, minyak tanah, batu bara dan bahan bakar lainnya.
3. Banyaknya tenaga listrik yang digunakan (X_3) yaitu jumlah (dalam kwh) dari seluruh tenaga listrik yang digunakan pada proses produksi. Merupakan jumlah dari banyaknya tenaga listrik yang dibeli dan banyaknya tenaga listrik yang dibangkitkan sendiri dikurangi banyaknya tenaga listrik yang dijual.
4. Banyaknya bahan baku dan bahan penolong (X_4) yaitu nilai (dalam ribuan) dari seluruh bahan baku serta bahan penolong yang digunakan dalam proses produksi selama tahun 2007.
5. Produksi (Y) yaitu nilai (dalam ribuan) dari semua barang yang dihasilkan baik yang diekspor maupun tidak diekspor, tidak termasuk di dalamnya

barang yang belum selesai diolah (setengah jadi) dan barang yang tidak diproses / tidak diolah.

3.3 Metode Penelitian

Ada dua tahapan utama yang dilakukan pada penelitian ini yaitu studi simulasi dan aplikasi pada data hasil survei Industri Besar dan Sedang (IBS) yang dilakukan oleh Badan Pusat Statistik Provinsi Jawa Timur Tahun 2007. Sebelumnya, terlebih dahulu dilakukan beberapa hal antara lain :

- i. Mengumpulkan berbagai pustaka yang dapat mendukung penelitian dari berbagai sumber
- ii. Mengumpulkan data tentang semua variabel yang akan digunakan dalam penelitian
- iii. Melakukan pengelompokan terhadap perusahaan industri besar dan sedang menurut Kode Induk Perusahaan. Pengelompokan ini dilakukan guna mengumpulkan perusahaan-perusahaan industri ke dalam kelompok-kelompok yang memiliki pola hubungan yang sama pada variabel-variabel yang akan diteliti.
- iv. Mengelompokkan perusahaan industri besar dan sedang menurut pola *missing*. Dalam penelitian ini yang akan dibahas adalah perusahaan-perusahaan dengan pola *missing* monoton.
- v. Melakukan analisis statistik deskriptif pada variabel yang digunakan dalam penelitian.
- vi. Menentukan variabel apa saja yang berpengaruh secara signifikan terhadap produksi dengan memodelkan variabel pada perusahaan yang datanya lengkap (tanpa nilai *missing*) sebagai berikut:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$$

dengan Y adalah nilai produksi

X_1 adalah banyaknya tenaga kerja

X_2 adalah bahan bakar dan pelumas yang digunakan

X_3 adalah banyaknya tenaga listrik yang digunakan

X_4 adalah banyaknya bahan baku yang digunakan

- vii. Melakukan pengujian terhadap asumsi yang dipakai yaitu data yang digunakan adalah multivariat normal. Dalam hal ini diawali dengan pengujian terhadap residual dari model pada tahap vi kemudian dilanjutkan dengan uji multivariat normal pada variabel yang digunakan pada proses imputasi.

3.3.1 Studi Simulasi

Sebagai langkah awal sebelum menerapkan metode regresi dan *predictive mean matching*, terlebih dahulu dilakukan studi simulasi untuk mengetahui bagaimana kinerja kedua metode dalam melakukan imputasi. Simulasi dilakukan dengan berbagai skenario yang disesuaikan dengan kondisi data hasil survei IBS yang akan dijadikan sebagai studi aplikasi sehingga bisa disimpulkan metode mana yang lebih baik dalam melakukan imputasi terhadap *missing data*.

Secara umum, studi simulasi dilakukan dengan membangkitkan sekumpulan data berdistribusi multivariat normal serta data yang tidak berdistribusi multivariat normal. Dari data tersebut kemudian sebagian dari nilai Y (dianalogikan sebagai nilai produksi) *dimissingkan* dengan mengikuti asumsi MCAR. Dengan asumsi MCAR, kemungkinan terjadinya nilai Y yang *missing* bisa muncul pada observasi manapun (random).

Dengan menggunakan kedua metode, imputasi dilakukan terhadap *missing data* yang ada. Hasil imputasi ini digunakan untuk menghitung *Mean Square Error Simulasi* (MSE_S). MSE_S dihitung dengan menggunakan rumus sebagai berikut:

$$MSE_S = \frac{\sum_{i=1}^{n_0} (Y_{i\ miss} - Y_{i\ imp})^2}{n_0}$$

dengan n_0 adalah banyaknya observasi dengan nilai Y yang *dimissingkan*

$Y_{i\ miss}$ adalah nilai Y ke- i yang *dimissingkan*

$Y_{i\ imp}$ adalah nilai hasil imputasi untuk Y ke- i yang *dimissingkan*

Metode dengan nilai MSE_S yang lebih kecil berarti kinerja metode tersebut dalam melakukan imputasi data adalah lebih baik dibanding metode lainnya.

3.3.2 Imputasi Berganda dengan Metode Regresi

Untuk menghasilkan nilai yang akan diimputasikan pada setiap imputasi dengan menggunakan metode regresi, dilakukan tahapan sebagai berikut :

1. Menghitung varians $\sigma_*^2 = \hat{\sigma}_1^2 (n_1 - q) / g$ seperti pada persamaan (2.5)
2. Menentukan parameter baru β_* menggunakan persamaan (2.6)
3. Melakukan imputasi pada *missing data* dengan : $\hat{Y}_{i*} = X_i \beta_* + z_i \sigma_*$

nilai z_i diperoleh dengan membangkitkan data berdistribusi normal (0,1)

3.3.3 Imputasi Berganda dengan Metode *predictive mean matching*

Untuk menghasilkan nilai yang akan diimputasikan dengan menggunakan metode *predictive mean matching*, dilakukan tahapan berikut :

1. Menghitung varian $\sigma_*^2 = \hat{\sigma}_1^2 (n_1 - q) / g$ seperti pada persamaan (2.5)
2. Menentukan parameter baru β_* menggunakan persamaan (2.6)
3. Hitung prediksi untuk setiap nilai yang hilang dengan \hat{Y}_{i*} menggunakan persamaan (2.8)
4. Ambil responden yang memiliki data lengkap dengan nilai Y_i yang terdekat dengan \hat{Y}_{i*}
5. Melakukan imputasi pada *missing data* dengan menggunakan nilai dari responden yang diambil pada tahap ke-4.

3.3.4 Membandingkan Efisiensi Relatif

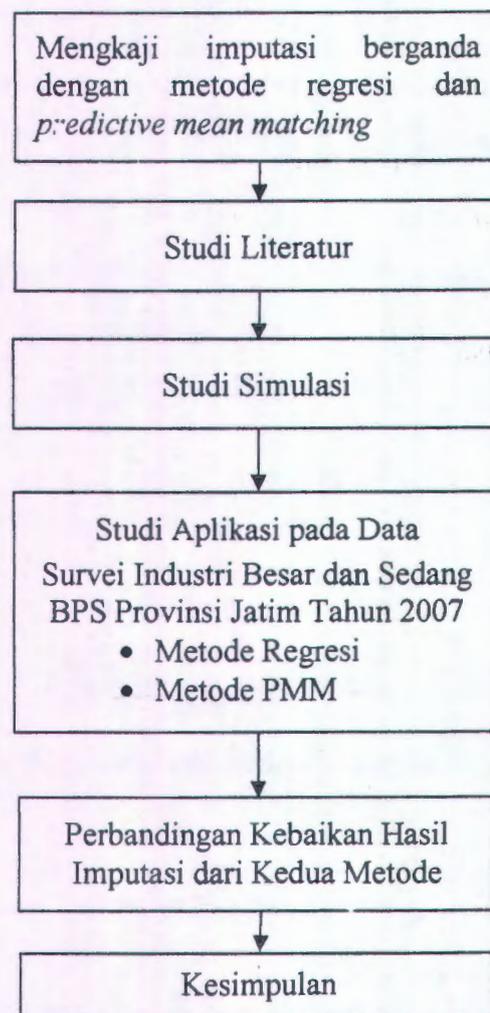
Selanjutnya guna membandingkan Efisiensi Relatif dari kedua metode imputasi dilakukan langkah-langkah sebagai berikut :

1. Hitung estimasi titik untuk Q menggunakan persamaan (2.9)
2. Hitung varian *within imputation* (\bar{U}) dan *between imputation* (B) seperti pada persamaan (2.10) dan (2.11)
3. Hitung besarnya derajat bebas (v_m) menggunakan persamaan (2.13)

4. Hitung besarnya nilai (r) menggunakan persamaan (2.14)
5. Hitung $\hat{\lambda}$ yaitu *fraction* dari informasi *missing* pada Q menggunakan persamaan (2.15)
6. Hitung besarnya efisiensi relatif (ER) menggunakan persamaan (2.16).

Dalam penelitian ini digunakan paket program SAS 9.0 dalam menerapkan imputasi dengan kedua metode.

Untuk mempermudah gambaran metode penelitian yang digunakan, dapat ditampilkan dalam diagram alur sebagai berikut :



Gambar 3.1 Diagram Alur Penelitian

BAB 4

ANALISIS DAN PEMBAHASAN

4.1 Studi Simulasi

Sebelum menerapkan metode regresi dan metode *predictive mean matching* untuk melakukan imputasi pada *missing data* yang terdapat di dalam data hasil Survei Industri Besar dan Sedang yang telah dilakukan oleh Badan Pusat Statistik Provinsi Jawa Timur Tahun 2007, terlebih dahulu dilakukan simulasi guna mengetahui kinerja dari kedua metode. Dalam tiap simulasi, imputasi masing-masing dilakukan sebanyak lima kali sebab imputasi umumnya cukup dilakukan antara tiga sampai lima kali (Rubin, 1996 ; halaman 480). Simulasi dilakukan dengan berbagai skenario sebagai berikut :

4.1.1 Skenario ke-1

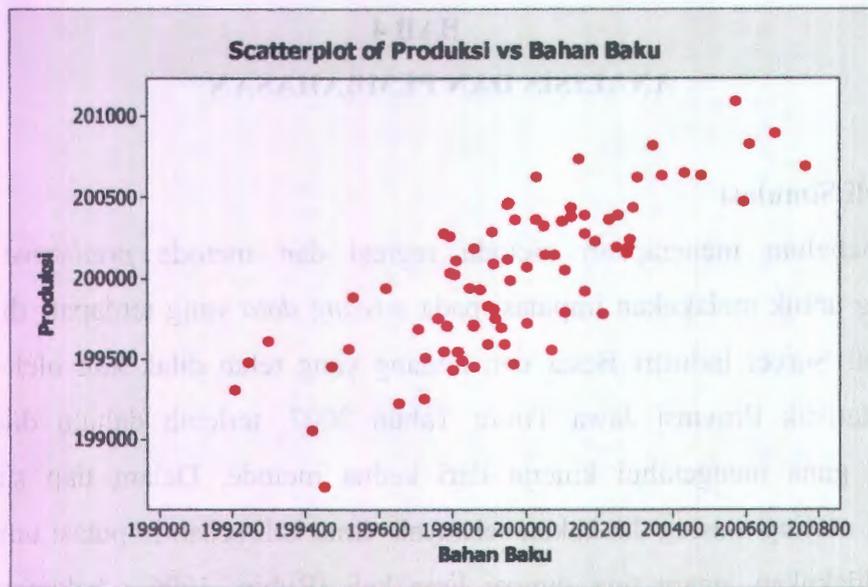
Simulasi dilakukan dengan membangkitkan data normal multivariat dimana X_1 , X_2 dan ϵ masing-masing memiliki rata-rata 30, 200.000 dan 0. Matrik varian-kovariannya adalah

$$\Sigma = \begin{bmatrix} 25 & 0,5 & 0 \\ 0,5 & 100.000 & 0 \\ 0 & 0 & 81.000 \end{bmatrix}$$

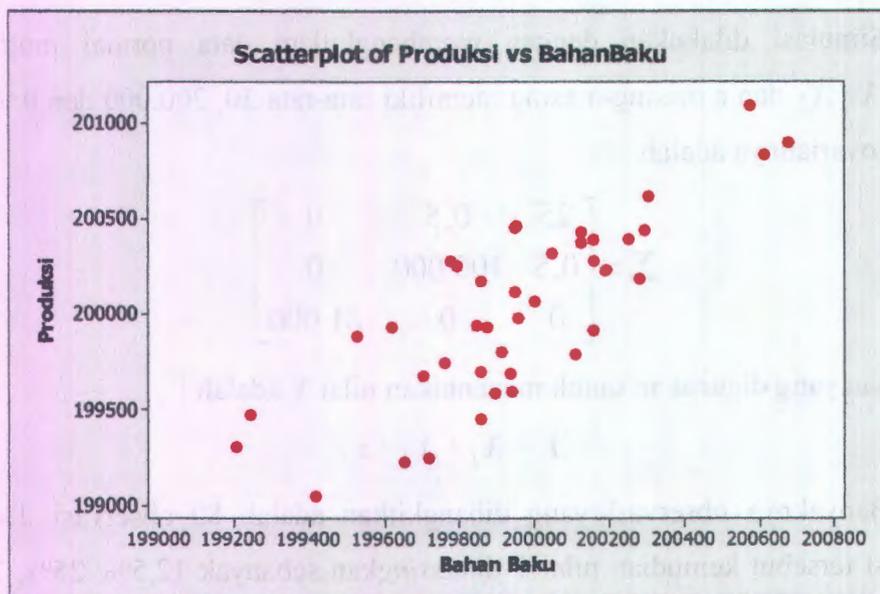
Persamaan yang digunakan untuk menentukan nilai Y adalah :

$$Y = X_1 + X_2 + \epsilon .$$

Banyaknya observasi yang dibangkitkan adalah 80 observasi. Dari 80 observasi tersebut kemudian nilai Y *dimissingkan* sebanyak 12,5%, 25%, 37,5% dan 50% dengan mengikuti asumsi MCAR. Dalam asumsi MCAR nilai Y *dimissingkan* secara random dan tidak terpengaruh pada besarnya nilai variabel X_1 , X_2 maupun nilai dari variabel Y itu sendiri. Secara teknis, hal ini dilakukan dengan membangkitkan angka random sebanyak 10, 20, 30 dan 40 yang menunjukkan nomor urut observasi, kemudian dari observasi terpilih tersebut nilai Y *dimissingkan*. Sebagai ilustrasi, berikut ini ditampilkan plot dari 80 data produksi yang lengkap dan plot data produksi setelah *dimissingkan* sebanyak 50% :



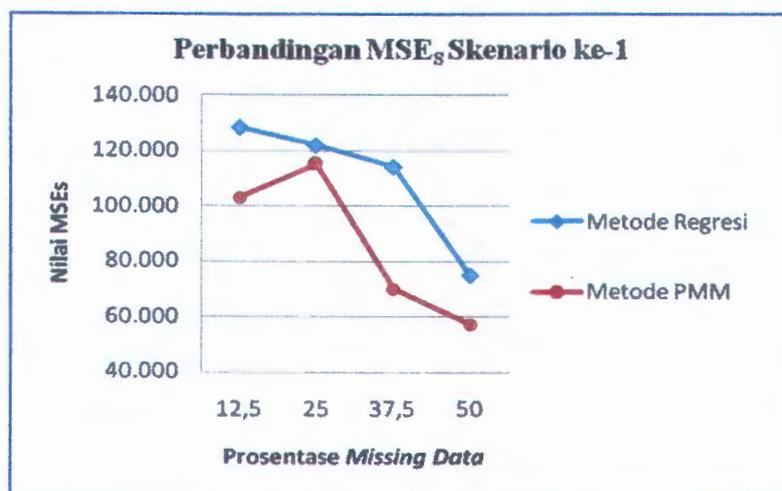
Gambar 4.1 Plot Data Lengkap



Gambar 4.2 Plot Data setelah dimissingkan sebanyak 50%

Dengan menggunakan metode regresi dan metode *predictive mean matching* (PMM) dilakukan imputasi terhadap data yang *missing* tersebut masing-masing sebanyak 5 kali menggunakan paket program SAS 9,0 *seed* 1979. *Seed* dengan nilai tertentu digunakan supaya nilai imputasi yang dihasilkan adalah sama ketika dilakukan pengecekan ulang.

Hasil imputasi dari kedua metode kemudian dihitung nilai *Mean Square Error* Simulasi (MSE_S) untuk dibandingkan keakuratannya. Hasil simulasi disajikan pada Gambar 4.3. Dari Gambar 4.3 terlihat bahwa MSE_S dari metode PMM secara umum lebih kecil dibandingkan dengan metode regresi.



Gambar 4.3 Grafik Perbandingan MSE_S Metode Regresi dan Metode PMM pada Simulasi Skenario ke-1

4.1.2 Skenario ke-2

Pada simulasi yang kedua ini dilakukan dengan cara membangkitkan data normal multivariat dimana X_1, X_2, X_3 dan ε masing-masing memiliki rata-rata 30, 25.000, 250.000 dan 0. Matrik varian-kovariannya adalah:

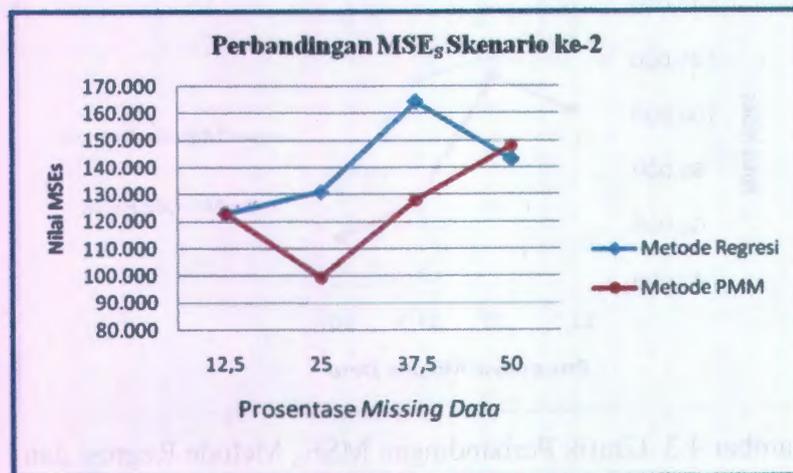
$$\Sigma = \begin{bmatrix} 20 & 0,5 & 0,5 & 0 \\ 0,5 & 15.000 & 0 & 0 \\ 0,5 & 0 & 100.000 & 0 \\ 0 & 0 & 0 & 100.000 \end{bmatrix}$$

Nilai Y ditentukan dengan menggunakan persamaan :

$$Y = X_1 + X_2 + X_3 + \varepsilon.$$

Sebanyak 80 observasi dibangkitkan dengan mengikuti kondisi tersebut. Dari 80 observasi tersebut kemudian nilai Y di~~miss~~ingkan masing-masing sebanyak 12,5%, 25%, 37,5% dan 50% dengan mengikuti asumsi MCAR. Imputasi terhadap *missing data* tersebut dilakukan dengan paket program SAS 9,0 *seed* 1500 menggunakan metode regresi dan metode PMM. Dari hasil imputasi kedua

metode kemudian dihitung nilai *Mean Square Error* Simulasi (MSE_S) untuk dibandingkan keakuratannya. Dari hasil simulasi pada Gambar 4.4 di bawah ini terlihat bahwa nilai MSE_S yang dihasilkan oleh metode regresi tidak selalu lebih kecil dibandingkan metode PMM. Artinya dari kedua metode tidak satupun yang selalu menghasilkan nilai imputasi yang lebih akurat dibandingkan dengan metode lainnya.



Gambar 4.4 Grafik Perbandingan MSE_S Metode Regresi dan Metode PMM pada Simulasi Skenario ke-2

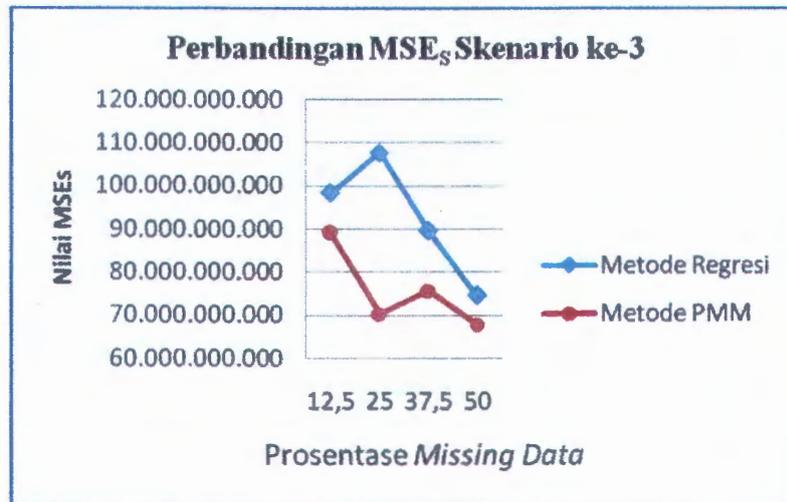
4.1.3 Skenario ke-3

Berbeda dengan kedua simulasi sebelumnya yang membangkitkan data berdistribusi normal multivariat, pada simulasi yang ketiga ini dilakukan dengan membangkitkan data berdistribusi uniform dimana X_1 , X_2 dan ϵ masing-masing memiliki nilai berkisar antara 20 s/d 45, 750.000 s/d 1.500.000 dan -300.000 s/d 700.000. Persamaan yang digunakan dalam menentukan nilai Y sama dengan persamaan pada skenario ke-1 yaitu :

$$Y = X_1 + X_2 + \epsilon .$$

Banyaknya observasi yang dibangkitkan pada simulasi ini adalah sebanyak 80 observasi. Dengan prosedur yang sama, dari 80 observasi tersebut nilai Y dimissingkan dengan mengikuti asumsi MCAR masing-masing sebanyak 12,5%, 25%, 37,5% dan 50%. Dengan menggunakan metode regresi dan metode PMM dilakukan imputasi terhadap *missing data* tersebut masing-masing sebanyak 5 kali

dengan menggunakan *seed* 79. Nilai MSE_S yang dihasilkan oleh kedua metode disajikan pada Gambar 4.5.



Gambar 4.5 Grafik Perbandingan MSE_S Metode Regresi dan Metode PMM pada Simulasi Skenario ke-3

Dari Gambar 4.5 di atas terlihat bahwa MSE_S dari metode PMM selalu lebih kecil dibandingkan dengan MSE_S metode regresi. Pada simulasi yang ketiga ini nampak bahwa metode PMM memberikan hasil imputasi yang lebih baik dibandingkan metode regresi.

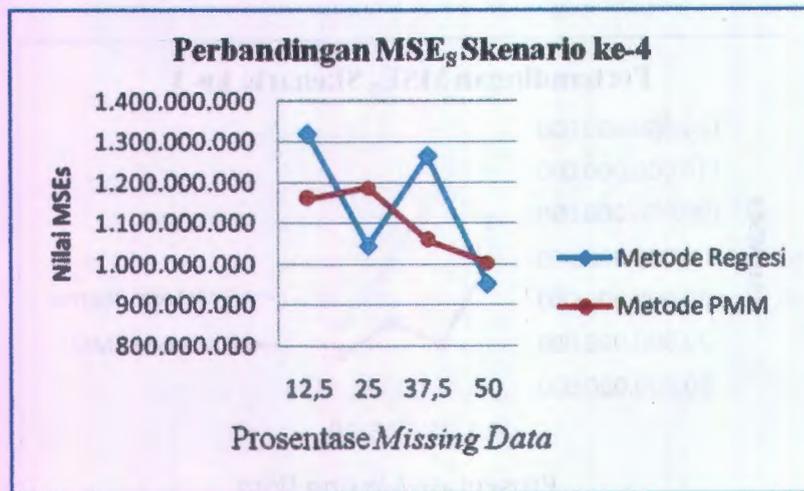
4.1.4 Skenario ke-4

Pada simulasi yang keempat ini dilakukan dengan membangkitkan data berdistribusi uniform dimana X_1 , X_2 , X_3 dan ε masing-masing memiliki nilai berkisar antara 25 s/d 60, 500 s/d 2.000, 100.000 s/d 500.000 dan -70.000 s/d 30.000. Persamaan yang digunakan dalam menentukan nilai Y adalah

$$Y = X_1 + X_2 + X_3 + \varepsilon .$$

Seperti pada simulasi sebelumnya, banyaknya observasi yang dibangkitkan sebanyak 80 observasi. Dengan mengikuti asumsi MCAR, dari 80 observasi tersebut kemudian nilai Y di *missing*kan sebanyak 12,5%, 25%, 37,5% dan 50%. Dengan menggunakan metode regresi dan metode PMM dilakukan imputasi terhadap *missing data* tersebut masing-masing sebanyak 5 kali menggunakan paket program SAS 9,0 dengan *seed* 80. Untuk membandingkan keakuratan hasil

imputasi dari kedua metode kemudian dihitung nilai *Mean Square Error* Simulasi (MSE_s). Hasil simulasi disajikan pada Gambar 4.6 berikut ini :



Gambar 4.6 Grafik Perbandingan MSE_s Metode Regresi dan Metode PMM pada Simulasi Skenario ke-4

Dari hasil simulasi pada Gambar 4.6 di atas bahwa nilai MSE_s yang dihasilkan oleh metode regresi tidak selalu lebih kecil dibandingkan metode PMM. Artinya, apabila dilihat dari nilai MSE_s yang dihasilkan tidak bisa dinyatakan bahwa metode regresi selalu menghasilkan nilai imputasi yang lebih akurat dibandingkan dengan metode PMM.

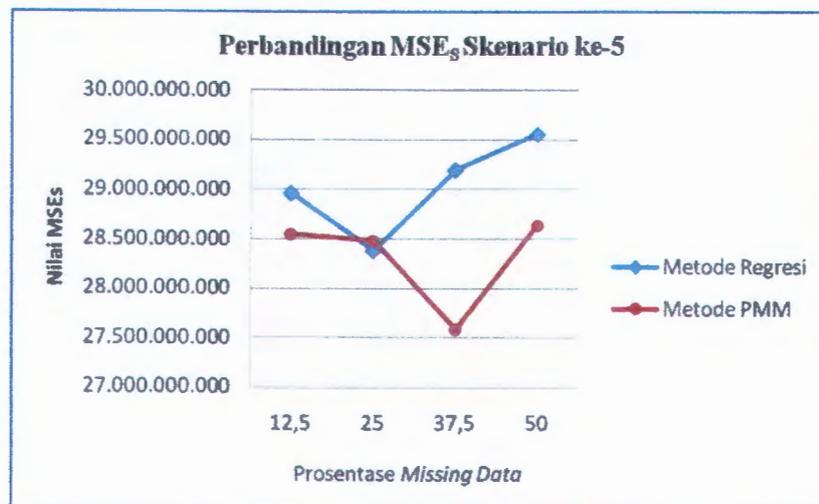
4.1.5 Skenario ke-5

Simulasi kali ini dilakukan dengan kondisi mirip dengan skenario ke-4 yaitu membangkitkan data berdistribusi uniform. Yang membedakan adalah variabel X hanya ada dua yaitu X_1 dan X_2 . Nilai-nilai X_1 dan X_2 dan ϵ masing-masing berkisar antara 25 s/d 60, 100.000 s/d 500.000 dan -70.000 s/d 30.000. Persamaan yang digunakan dalam menentukan nilai Y adalah :

$$Y = 0.5X_1 + 0,75X_2 + \epsilon.$$

Banyaknya observasi yang dibangkitkan adalah 80 observasi. Dari 80 observasi tersebut kemudian nilai Y di *missing*kan sebanyak 12,5%, 25%, 37,5% dan 50% dengan mengikuti asumsi MCAR. Dengan menggunakan metode regresi dan metode PMM dilakukan imputasi terhadap *missing data* tersebut masing-masing sebanyak 5 kali menggunakan paket program SAS 9,0 dengan *seed* 80.

Hasil imputasi dari kedua metode kemudian dihitung nilai *Mean Square Error* Simulasi (MSE_S) untuk dibandingkan keakuratannya. Hasilnya disajikan pada Gambar 4.7 berikut :



Gambar 4.7 Grafik Perbandingan MSE_S Metode Regresi dan Metode PMM pada Simulasi Skenario ke-5

Dari Gambar 4.7 di atas terlihat bahwa MSE_S dari metode PMM secara umum lebih kecil dibandingkan dengan metode regresi. Selain itu terdapat hal menarik yang muncul pada simulasi ini yaitu munculnya nilai-nilai negatif yang dihasilkan oleh metode regresi. Kondisi tersebut tidak terjadi pada metode PMM. Dengan demikian pada kondisi ini nilai imputasi yang dihasilkan oleh metode PMM adalah lebih masuk akal dibandingkan metode regresi.

4.2 Deskripsi Data Hasil Survei Industri Besar dan Sedang

Sebelum dilakukan imputasi terhadap *missing data*, terlebih dahulu dilakukan eksplorasi terhadap data hasil Survei Industri Besar dan Sedang yang telah dilakukan oleh Badan Pusat Statistik Provinsi Jawa Timur Tahun 2007. Hal ini dilakukan guna mengetahui gambaran awal dari data yang akan digunakan dalam penelitian ini.

Secara umum, hasil Survei Industri Besar dan Sedang di Provinsi Jawa Timur tahun 2007 tidak berbeda jauh dengan tahun-tahun sebelumnya dimana tingkat pemasukan dokumen dalam survei ini tidak mencapai seratus persen. Tingkat pemasukan dokumen yang tidak mencapai seratus persen ini disebabkan karena tidak semua perusahaan yang menjadi responden pada survei ini memberikan respon terhadap *surveyor*. Banyaknya perusahaan yang respon dan non respon bisa dilihat pada Tabel 4.1 berikut ini:

Tabel 4.1 Banyaknya Perusahaan yang Respon dan Non Respon

Perusahaan Industri	Jumlah	Prosentase
Respon	4.305	69,07
Non Respon	1.928	30,93
Total	6.233	100,00

Sumber : Survei Industri Besar dan Sedang BPS Provinsi Jawa Timur Tahun 2007 (Diolah)

Dari Tabel 4.1 di atas terlihat bahwa banyaknya perusahaan industri besar dan sedang di Provinsi Jawa Timur tahun 2007 yang non respon cukup tinggi yaitu sebanyak 1.928 perusahaan (30,93%). Dari 4.305 buah perusahaan yang respon, ternyata sebanyak 265 perusahaan (6,20 %) diantaranya tidak memberikan datanya pada variabel jumlah produksi yang dihasilkan seperti tampak pada Tabel 4.2 di bawah ini :

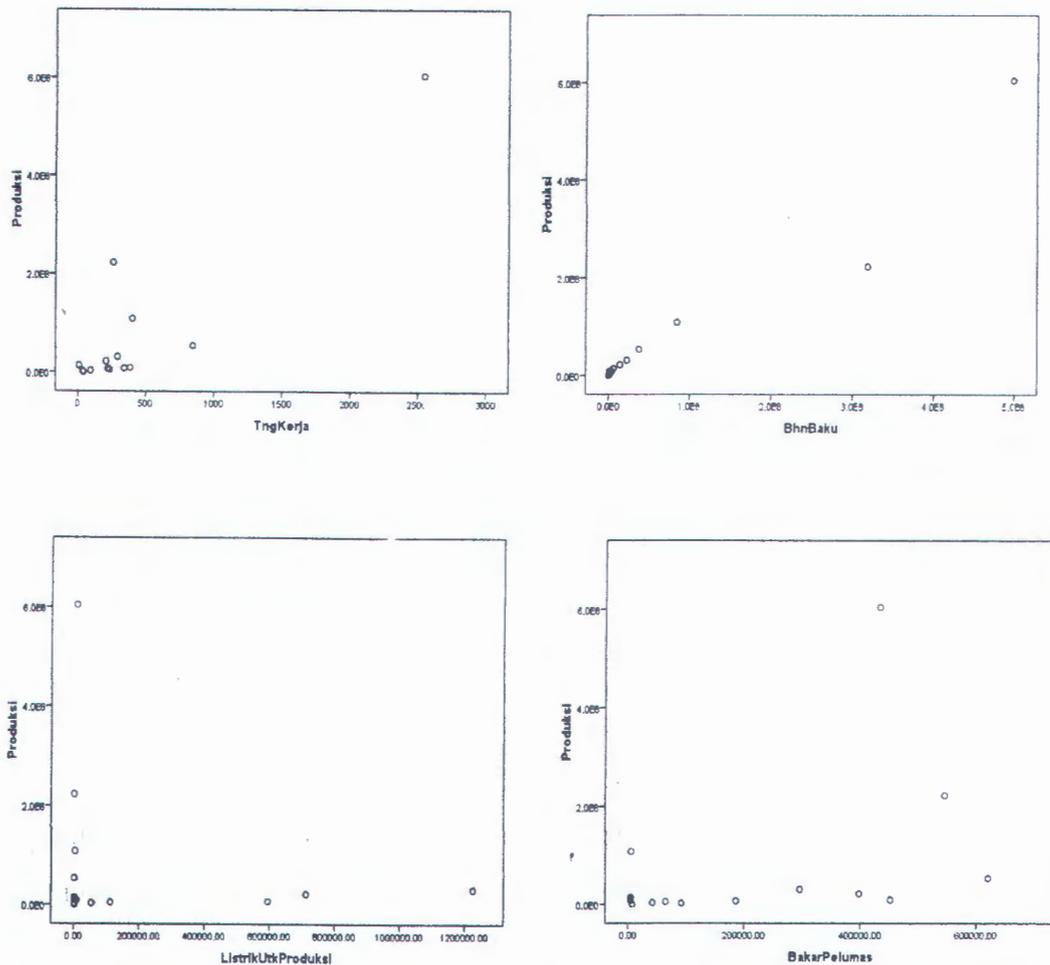
Tabel 4.2 Banyaknya Variabel Produksi yang Missing

Variabel Produksi	Jumlah	Prosentase
<i>Missing</i>	265	6,20
Tersedia	4.040	93,80
Total	4.305	100,00

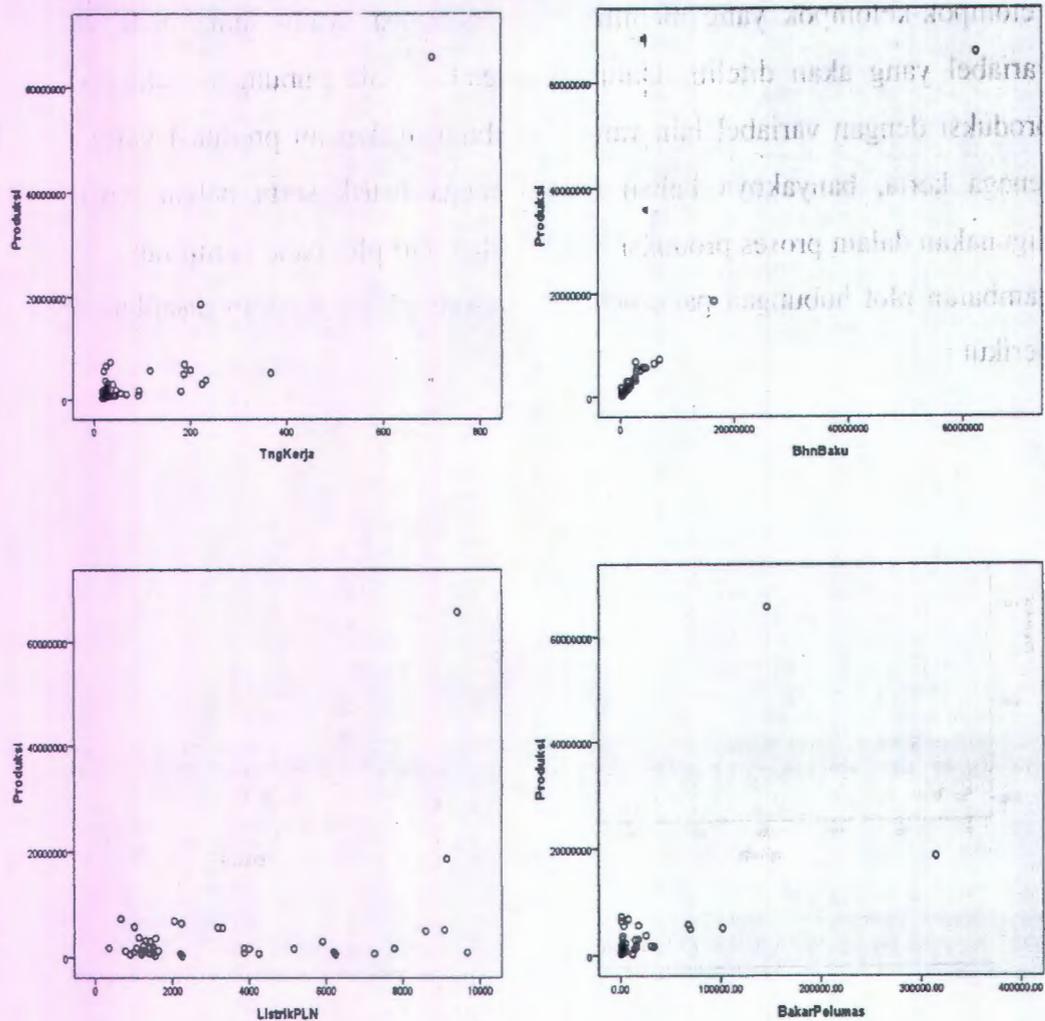
Sumber : Survei Industri Besar dan Sedang BPS Provinsi Jawa Timur Tahun 2007 (Diolah)

Selanjutnya, dilakukan pengelompokan terhadap perusahaan industri besar dan sedang berdasarkan Kode Induk Perusahaan (KIP). Pengelompokan ini dilakukan guna mengumpulkan perusahaan-perusahaan industri ke dalam

kelompok-kelompok yang memiliki pola hubungan yang sama pada variabel-variabel yang akan diteliti. Untuk mengetahui pola hubungan antara variabel produksi dengan variabel lain yang berhubungan dengan produksi yaitu jumlah tenaga kerja, banyaknya bahan baku, tenaga listrik serta bahan bakar yang digunakan dalam proses produksi dapat dilihat dari plot pada Lampiran 1. Sebagai gambaran plot hubungan pada beberapa kelompok perusahaan disajikan sebagai berikut :

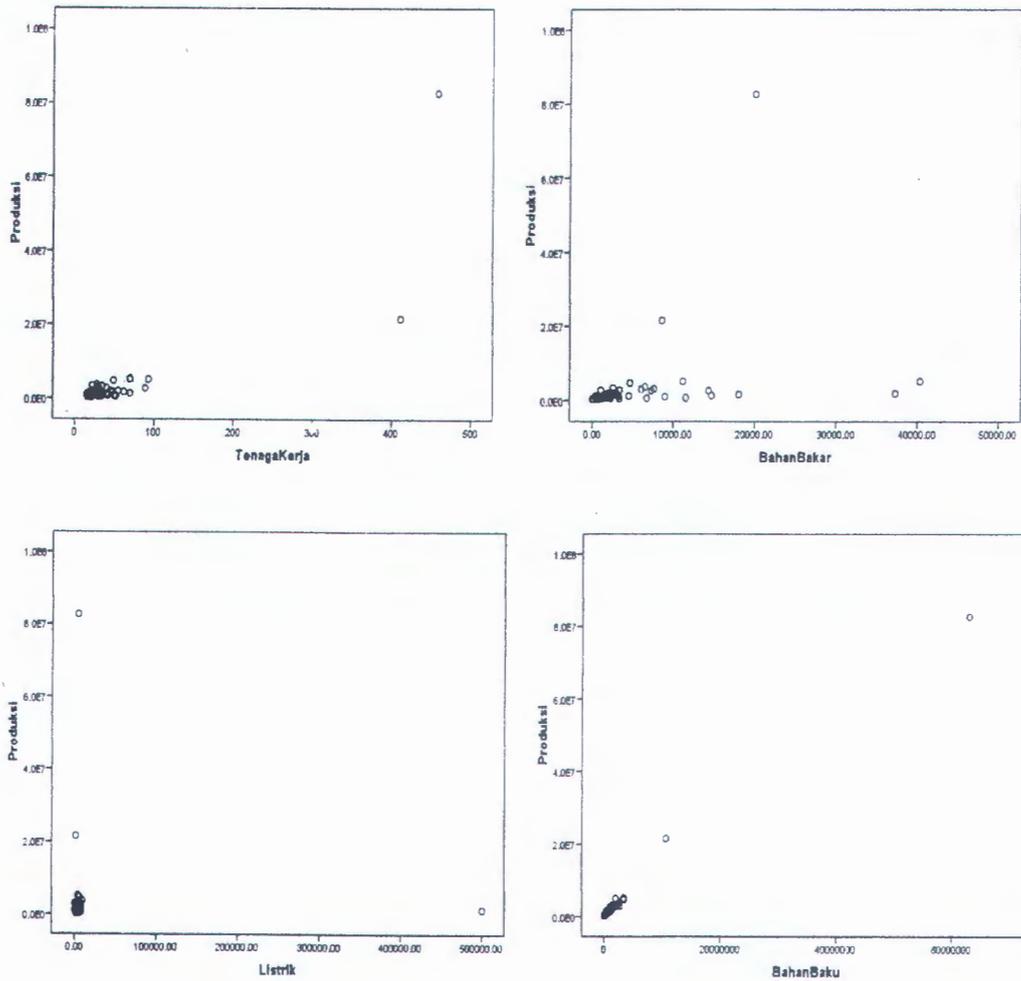


Gambar 4.8 Plot Hubungan antara variabel produksi dengan jumlah tenaga kerja, bahan baku, tenaga listrik serta bahan bakar pada KIP 15121



Gambar 4.9 Plot Hubungan antara variabel produksi dengan jumlah tenaga kerja, bahan baku, tenaga listrik serta bahan bakar pada KIP 15122

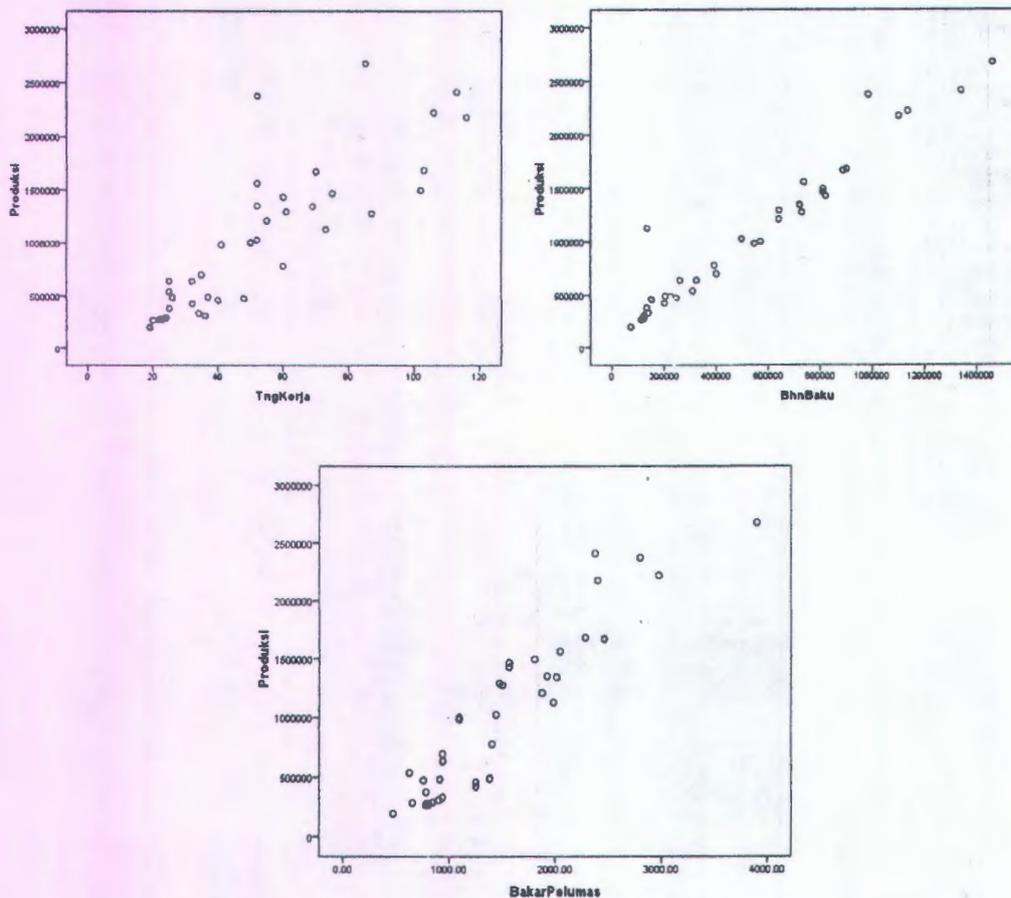
Dari Gambar 4.8 dan Gambar 4.9 di atas terlihat bahwa pola hubungan antara variabel produksi dengan keempat variabel lainnya menunjukkan pola yang hampir mirip. Pola hubungan antara produksi dengan bahan baku dan jumlah tenaga kerja secara umum menunjukkan hubungan linier sedangkan pola hubungan antara bahan baku dengan jumlah tenaga listrik dan bahan bakar tidak menunjukkan pola hubungan linier.



Gambar 4.10 Plot Hubungan antara variabel produksi dengan jumlah tenaga kerja, bahan baku, tenaga listrik serta bahan bakar pada KIP 19201

Dari Gambar 4.10 di atas terlihat bahwa pola hubungan antara bahan baku dengan produksi dan jumlah tenaga kerja dengan produksi secara umum diperkirakan linier, sedangkan hubungan antara bahan bakar terhadap produksi polanya tidak beraturan. Hubungan antara penggunaan listrik dengan jumlah produksinya tidak menunjukkan pola hubungan tertentu.

Pada KIP 17115 tidak terdapat plot hubungan antara produksi dengan tenaga listrik. Perusahaan pada kelompok ini tidak menggunakan listrik sebagai salah satu komponen produksinya seperti tampak pada plot di bawah ini :



Gambar 4.11 Plot Hubungan antara variabel produksi dengan jumlah tenaga kerja, bahan baku serta bahan bakar pada KIP 17115

Tahap berikutnya, dilakukan pemilihan Kelompok Industri yang akan dianalisa berdasarkan pada banyaknya kasus *missing data* pada kelompok tersebut. Selain itu pola hubungan antara variabel produksi dengan variabel lain yang berhubungan dengan produksi yaitu jumlah tenaga kerja, banyaknya bahan baku, tenaga listrik serta bahan bakar yang digunakan dalam proses produksi juga menjadi pertimbangan yang lain. Berdasarkan beberapa pertimbangan tersebut diambil dua kelompok perusahaan yaitu perusahaan KIP 17115 (kain tenun ikat) dan KIP 19201 (alas kaki).

4.3 Imputasi *Missing Data* pada Hasil Survei Industri Besar dan Sedang

Pada bagian ini akan dibahas imputasi *missing data* pada hasil survei industri besar dan sedang pada KIP 17115 (kain tenun ikat) dan KIP 19201 (alas kaki).

4.3.1 KIP 17115 (kain tenun ikat)

Pada kelompok ini terdapat 75 responden perusahaan industri besar dan sedang. Perusahaan-perusahaan tersebut ada yang memberikan data secara lengkap dan ada pula yang hanya memberikan datanya pada variabel tertentu (terdapat *missing data*). Dari ke-75 responden tersebut kemudian dipisahkan antara 39 perusahaan dengan data lengkap dan 36 perusahaan yang terdapat *missing data*. Selanjutnya diambil perusahaan-perusahaan yang hanya *missing* pada nilai produksinya seperti pada Tabel 4.3. Dari tabel tersebut nampak bahwa terdapat sembilan perusahaan yang *missing* pada nilai produksinya.

Tabel 4.3 Observasi dengan nilai Produksi *Missing*

Perusahaan Ke-	Tenaga Kerja	Bahan Bakar	Bahan Baku	Produksi
5	20	1.000	123.550	*
7	21	1.200	150.870	*
13	24	720	133.430	*
23	30	1.200	233.256	*
26	30	1.230	223.550	*
28	32	936	22.360	*
30	40	1.100	539.755	*
36	90	1.404	813.456	*

Sumber : Survei Industri Besar dan Sedang BPS Provinsi Jawa Timur Tahun 2007 (Diolah)

Untuk melakukan imputasi terhadap nilai *missing* tersebut, terlebih dahulu dilakukan pengujian terhadap variabel tenaga kerja, bahan bakar dan bahan baku pada perusahaan yang memberikan data lengkap apakah ketiga variabel tersebut semuanya diikuti dalam pembentukan model atau tidak. Untuk menentukan variabel mana saja yang akan diikutsertakan dalam model maka terlebih dahulu dilakukan uji korelasi antar variabel dengan prosedur sebagai berikut :

Hipotesis untuk uji ini adalah :

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

Tabel 4.4 Korelasi Pearson

		Produksi	Tenaga Kerja	Bahan Bakar	Bahan Baku
Produksi	Korelasi	1	0,849	0,933	0,971
	Signifikansi		0,000	0,000	0,000
Tenaga Kerja	Korelasi	0,849	1	0,780	0,841
	Signifikansi	0,000		0,000	0,000
Bahan Bakar	Korelasi	0,933	0,780	1	0,879
	Signifikansi	0,000	0,000		0,000
Bahan Baku	Korelasi	0,971	0,841	0,879	1
	Signifikansi	0,000	0,000	0,000	

Dari Tabel 4.4 di atas tampak bahwa signifikansi korelasi antara produksi dengan tenaga kerja, bahan bakar dan bahan baku masing-masing sebesar 0,000. Dengan demikian dapat disimpulkan bahwa H_0 yang menyatakan bahwa korelasi antar variabel adalah nol ditolak yang artinya bahwa terdapat hubungan yang kuat antara produksi dengan tenaga kerja, bahan bakar dan bahan baku. Besarnya koefisien korelasi masing-masing adalah 0,849, 0,933 dan 0,971.

Untuk mengetahui variabel prediktor mana saja yang akan disertakan dalam model dilakukan uji parsial terhadap variabel prediktor tersebut. Prosedur pengujiannya adalah sebagai berikut :

$$H_0 : \beta_i = 0$$

$$H_1 : \beta_i \neq 0, i=1,2,3,\dots,q$$

Tolak H_0 jika $|t| > t_{\alpha/2, n-q-1}$ dan dapat disimpulkan bahwa variabel prediktor tersebut berpengaruh secara signifikan terhadap variabel responnya.

Tabel 4.5 Koefisien Regresi pada KIP 17115

Prediktor	Koefisien	SE Koefisien	<i>t</i>	<i>p-value</i>
Konstanta	-98.823	51.260	-1,93	0,062
Tenaga Kerja	1.536	1.316	1,17	0,251
Bahan Bakar	311,64	55,45	5,62	0,000
Bahan Baku	1,1358	0,1283	8,85	0,000

Dari tabel 4.5 di atas terlihat nilai t pada variabel bahan bakar dan bahan baku sebesar 5,62 dan 8,85 lebih besar dari $t_{0,025;35}$ sebesar 2,011. Dengan demikian kedua variabel ini berpengaruh secara signifikan terhadap variabel produksi. Hal yang berbeda tampak pada variabel tenaga kerja dengan nilai t sebesar 1,17 kurang dari $t_{0,025;35}$ sebesar 2,011 dan disimpulkan tidak berpengaruh secara signifikan terhadap produksi. Dengan demikian variabel tenaga kerja bisa dikeluarkan dari model.

Tabel 4.6 ANOVA dengan Prediktor Bahan Baku, Bahan Bakar dan Tenaga Kerja

Sumber	DF	<i>Sum Square</i>	<i>Mean Square</i>	F	P
Regresi	3	17.945.723.733.973	5.981.907.911.324	392,52	0,000
<i>Error</i>	35	525.365.103.450	15.010.431.527		
Total	38	18.471.088.837.424			

Tabel 4.7 ANOVA dengan Prediktor Bahan Baku dan Bahan Bakar

Sumber	DF	<i>Sum Square</i>	<i>Mean Square</i>	F	P
Regresi	2	17.925.271.813.525	8.962.635.906.762	591,14	0,000
<i>Error</i>	36	545.817.023.898	15.161.583.997		
Total	38	18.471.088.837.424			

Pada saat variabel tenaga kerja dikeluarkan dari model ternyata nilai *Mean Square Error* meningkat dari 15.010.431.527 menjadi 15.161.583.997 seperti nampak pada Tabel 4.6 dan Tabel 4.7. Selain itu menurut Rubin (1996), dalam imputasi data diupayakan untuk memasukkan sebanyak mungkin variabel prediktor yang bisa meningkatkan validitas dari analisis serta meningkatkan ketepatan nilai imputasi. Schafer (1997), menyatakan bahwa dalam imputasi asumsi MAR akan semakin masuk akal apabila variabel yang dimasukkan semakin banyak. Oleh karena itu diputuskan bahwa variabel tenaga kerja tetap dimasukkan dalam analisa selanjutnya.

Langkah selanjutnya dilakukan uji asumsi terhadap 39 perusahaan yang memberikan data lengkap (Lampiran 2). Pengujian asumsi ini dilakukan dengan

cara melakukan uji normalitas terhadap residual dari regresi antara produksi sebagai variabel prediktor dengan tenaga kerja, bahan bakar dan bahan baku sebagai variabel respon.

Uji Kolmogorov Smirnov dilakukan untuk menguji apakah residual mengikuti distribusi normal atau tidak. Pengujian ini dilakukan dengan hipotesis sebagai berikut :

$$H_0 : F_{(x)} = F_{o(x)}$$

$$H_1 : F_{(x)} \neq F_{o(x)}$$

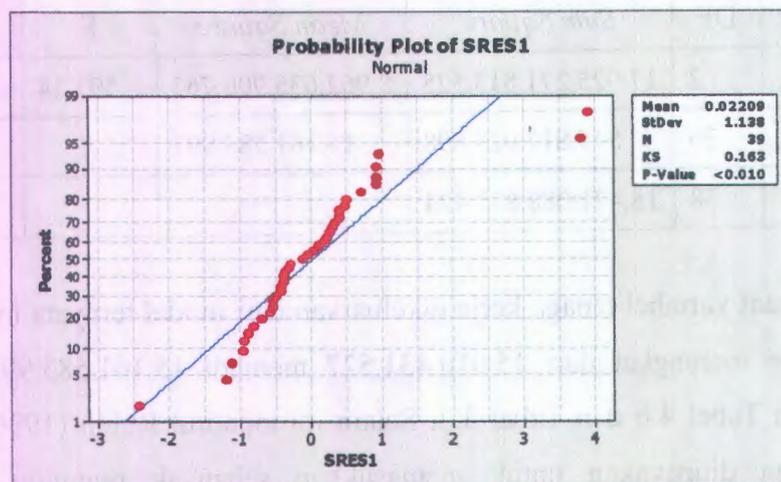
Statistik Uji :

$$D = \text{Sup} |S_{(x)} - F_{o(x)}|$$

dengan : $S_{(x)}$ adalah proporsi nilai pengamatan sampel yang kurang dari sama dengan x

$F_{o(x)}$ adalah fungsi yang diketahui (normal)

Daerah penolakan : Tolak H_0 jika $D \geq$ kuantil $1-\alpha$ pada tabel Kolmogorov Smirnov atau $p\text{-value} < \alpha$.



Gambar 4.12 Plot Residual pada KIP 17115 sebelum *outlier* dikeluarkan

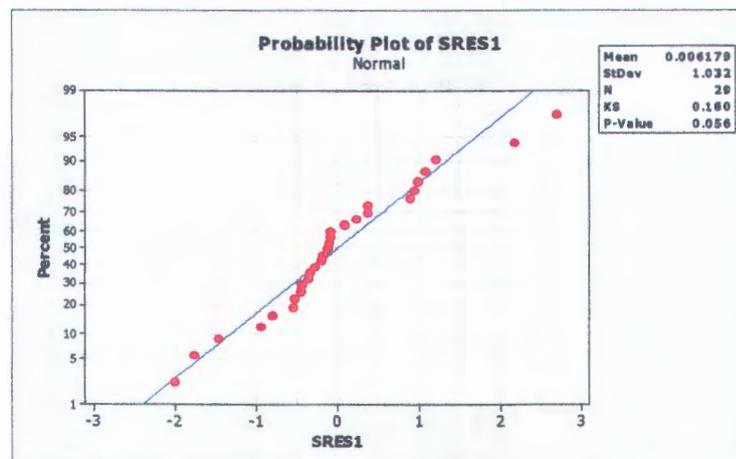
Dari output di atas terlihat bahwa $p\text{-value} < 0,010$ yang artinya bahwa H_0 yang menyatakan bahwa residual berdistribusi normal ditolak. Dengan demikian bisa disimpulkan bahwa residual tidak berdistribusi normal. Ketidaknormalan ini disebabkan adanya perusahaan yang dianggap sebagai *outlier*. Perusahaan *outlier* di sini diartikan sebagai perusahaan-perusahaan yang menghasilkan residual

standar dengan nilai yang relatif besar. Menurut Hair (1998) untuk kasus sampel kecil (kurang dari 80), maka standar skor dengan nilai ± 2 bisa dianggap sebagai *outlier*. Perusahaan-perusahaan tersebut disajikan pada Tabel 4.8.

Berdasarkan informasi tersebut, perusahaan-perusahaan yang dianggap sebagai *outlier* dikeluarkan dalam analisa selanjutnya. Setelah perusahaan-perusahaan tersebut dikeluarkan dan hanya menyertakan 29 perusahaan, residual dari model regresi diuji kembali dengan uji Kolmogorov Smirnov. Berdasarkan output pada Gambar 4.13 terlihat nilai *p-value* yang dihasilkan sebesar 0,056 dan disimpulkan bahwa residual sudah berdistribusi normal. Data ke-29 perusahaan tersebut dapat dilihat pada Lampiran 3.

Tabel 4.8 Perusahaan KIP 17115 yang dianggap sebagai *Outlier*

Perusahaan Ke-	Observasi	Residual	Standardized Residual
59	23	401.159	3,88
73	37	352.783	3,79
75	39	-277.336	-2,45
68	31	125.495	2,54
71	34	125.446	2,56
52	16	106.081	2,52
62	25	106.275	2,50
45	9	73.422	2,23
58	21	67.305	2,01
72	31	65.088	2,07



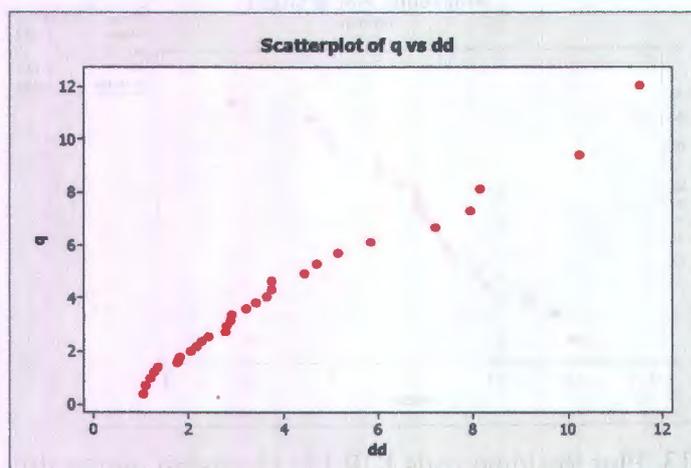
Gambar 4.13 Plot Residual pada KIP 17115 setelah *outlier* dikeluarkan

Syarat data yang diperlukan untuk menggunakan metode imputasi regresi dan *predictive mean matching* adalah normal multivariat. Pada data normal multivariat semua variabel berdistribusi normal secara individu. Oleh karena itu, dilakukan pengujian terhadap semua variabel yang akan dilibatkan dalam proses imputasi. Dengan prosedur yang sama seperti pada bagian sebelumnya dilakukan uji normalitas Kolmogorov Smirnov. Dari Tabel 4.9 di bawah ini terlihat bahwa *p-value* dari semua variabel nilainya lebih besar dari $\alpha(0,05)$ sehingga bisa disimpulkan bahwa semua variabel yang ada masing-masing berdistribusi normal.

Tabel 4.9 Hasil Uji Kolmogorov Smirnov Variabel pada KIP 17115

Keterangan	Tenaga Kerja	Bahan Bakar	Bahan Baku	Produksi
Rata-rata	48,76	1.280,07	443.589,38	846.000
Standar Deviasi	26,744	559,909	332.600	576.400
Nilai Ekstrim Absolut	0,166	0,179	0,172	0,186
Positif	0,166	0,179	0,172	0,186
Negatif	-0,133	-0,086	-0,132	-0,130
Kolmogorov Smirnov Z	0,893	0,963	0,925	1,001
Signifikansi (dua sisi)	0,402	0,312	0,359	0,269

Untuk menguji apakah variabel-variabel tersebut adalah normal multivariat dilakukan uji *QQ-test* (makro pada Lampiran 4). *Scatter plot* uji normal multivariat pada KIP 17115 ditampilkan pada Gambar 4.14. Dari Gambar 4.14 tersebut disimpulkan bahwa data telah memenuhi asumsi normal multivariat sehingga telah memenuhi syarat untuk menerapkan imputasi dengan metode regresi dan metode PMM.



Gambar 4.14 Scatter Plot Uji Normal Multivariat pada KIP 17115

4.3.1.1 Model Regresi pada KIP 17115 (Kain tenun ikat)

Dari ke-29 perusahaan yang memberikan data lengkap yang telah lolos dari deteksi *outlier* model regresi berganda yang terbentuk adalah :

$$\hat{Y} = -7.722 + 569 X_1 + 123 X_2 + 1,51 X_4 \quad (4.1)$$

dengan Y adalah nilai produksi

X_1 adalah banyaknya tenaga kerja

X_2 adalah nilai bahan bakar

X_4 adalah nilai bahan baku yang digunakan

4.3.1.2 Imputasi pada KIP 17115 (Kain tenun ikat) dengan Metode Regresi

Dengan menggunakan Paket Program SAS 9.0 PROC MI dengan menggunakan *seed* 2009 dilakukan imputasi sebanyak 5 kali. Imputasi umumnya cukup dilakukan antara tiga sampai lima kali (Rubin, 1996; halaman 480). Output lengkap pada Lampiran 5.

a. Model pada Setiap Imputasi

Tabel 4.10 Koefisien Regresi untuk Metode Monoton

Efek	Data Lengkap	Imputasi ke-				
		1	2	3	4	5
β_0	-0,10357	-0,094763	-0,112025	-0,086043	-0,114444	-0,102610
X_1	0,02588	0,021304	0,029412	0,014320	0,013770	0,018130
X_2	0,10801	0,113942	0,099467	0,119672	0,085909	0,101820
X_4	0,84414	0,839621	0,854730	0,846432	0,866335	0,869460

Dari Tabel di atas dapat diperoleh informasi model regresi yang diperoleh pada responden dengan data lengkap adalah:

$$\hat{Y} = -0,10357 + 0,02588 X_1 + 0,10801 X_2 + 0,84414 X_4 \quad (4.2)$$

Model ini berbeda dari model regresi pada (4.1). Hal ini disebabkan pada metode imputasi, penghitungan parameter dilakukan dengan menormalstandarkan variabel-variabel yang ada. Oleh karena itu nilai dari parameter yang dihasilkan berbeda dengan model regresi pada umumnya.

Berdasarkan informasi yang ada pada Tabel 4.10 di atas dapat diketahui model yang dipergunakan dalam setiap imputasi. Model-model tersebut adalah:

Model imputasi ke-1 :

$$\hat{Y}_p = -0,094763 + 0,02130 X_1 + 0,113942 X_2 + 0,839621 X_4 \quad (4.3)$$

Model imputasi ke-2 :

$$\hat{Y}_p = -0,112025 + 0,029412 X_1 + 0,09946 X_2 + 0,854730 X_4 \quad (4.4)$$

Model imputasi ke-3 :

$$\hat{Y}_p = -0,086043 + 0,014320 X_1 + 0,119672 X_2 + 0,846432 X_4 \quad (4.5)$$

Model imputasi ke-4 :

$$\hat{Y}_p = -0,114444 + 0,013770 X_1 + 0,085909 X_2 + 0,866335 X_4 \quad (4.6)$$

Model imputasi ke-5 :

$$\hat{Y}_p = -0,102610 + 0,018130 X_1 + 0,101820 X_2 + 0,869460 X_4 \quad (4.7)$$

b. Nilai yang diimputasikan

Dengan menggunakan model pada persamaan 4.3 sampai dengan 4.7 diperoleh nilai-nilai yang akan diimputasikan. Nilai yang diimputasikan pada nilai produksi yang *missing* ditampilkan pada Tabel 4.11 di bawah ini:

Tabel 4.11 Hasil Imputasi dengan Metode Regresi

Perusahaan Ke-	Nilai Imputasi ke-					Rata-rata
	1	2	3	4	5	
5	251.453	319.598	334.225	271.634	335.772	302.536
7	417.196	351.837	422.933	390.809	359.361	388.427
13	321.458	311.578	311.443	334.851	321.980	320.262
23	535.055	486.835	550.272	494.038	493.956	512.031
26	539.294	465.987	511.244	480.623	454.809	490.391
28	167.434	128.896	202.724	125.996	151.532	155.317
30	1.013.849	948.534	966.862	944.217	967.288	968.150
36	1.428.567	1.460.280	1.441.069	1.446.065	1.487.084	1.452.613

c. Efisiensi Relatif

Dari Tabel 4.12 di bawah ini, bisa diperoleh beberapa informasi sebagai berikut:

Tabel 4.12 Informasi Varian dari Imputasi Berganda

Varian			Fraction dari informasi Missing
Between	Within	Total	
9.877.576	8.300.345.820	8.312.198.911	0,001427

Varian *between imputation* sebesar 9.877.576 yang artinya adalah variasi antar imputasi satu dengan yang lainnya sebesar 9.877.576. Varian *within imputation* sebesar 8.300.345.820 yang artinya adalah variasi rata-rata dari 5 buah data lengkap yang tersedia sebesar 8.300.345.820. Nilai varian totalnya adalah sebesar 8.312.198.911.

Dari Tabel 4.12 di atas nilai efisiensi relatif dapat dihitung menggunakan persamaan 2.16. Besarnya nilai efisiensi relatif pada metode regresi dengan melakukan imputasi sebanyak lima kali adalah :

$$ER = \left(1 + \frac{\lambda}{m}\right)^{-1} \times 100\% = \left(1 + \frac{0,001427}{5}\right)^{-1} \times 100\% = 99,9715\%$$

4.3.1.3 Imputasi pada KIP 17115 (Kain tenun ikat) dengan Metode PMM

Dengan menggunakan Paket Program SAS 9.0 PROC MI dengan menggunakan *seed* 2009 dilakukan imputasi sebanyak 5 kali. Output lengkap pada Lampiran 6.

a. Model pada Setiap Imputasi

Tabel 4.13 Koefisien Regresi pada Metode PMM

Efek	Data Lengkap	Imputasi ke				
		1	2	3	4	5
β_0	-0,10357	-0,094763	-0,096688	-0,108278	-0,098677	-0,112275
X_1	0,02588	0,021304	0,041038	0,043001	0,038388	0,007553
X_2	0,10801	0,113942	0,126861	0,082288	0,096448	0,101372
X_4	0,84414	0,839621	0,812684	0,854794	0,850251	0,873147

Dari Tabel 4.13 di atas dapat diperoleh informasi model PMM pada responden dengan data lengkap adalah:

$$\hat{Y} = -0,10357 + 0,02588 X_1 + 0,10801 X_2 + 0,84414 X_4 \quad (4.8)$$

Model yang dipergunakan dalam setiap imputasi adalah:

Imputasi ke-1 :

$$\hat{Y}_1 = -0,094763 + 0,021304 X_1 + 0,113942 X_2 + 0,839621 X_4 \quad (4.9)$$

Imputasi ke-2 :

$$\hat{Y}_2 = -0,096688 + 0,041038 X_1 + 0,126861 X_2 + 0,812684 X_4 \quad (4.10)$$

Imputasi ke-3 :

$$\hat{Y}_3 = -0,108278 + 0,043001 X_1 + 0,082288 X_2 + 0,854794 X_4 \quad (4.11)$$

Imputasi ke-4 :

$$\hat{Y}_4 = -0,098677 + 0,038388 X_1 + 0,096448 X_2 + 0,850251 X_4 \quad (4.12)$$

Imputasi ke-5 :

$$\hat{Y}_5 = -0,112275 + 0,007553 X_1 + 0,101372 X_2 + 0,873147 X_4 \quad (4.13)$$

b. Nilai yang diimputasikan

Nilai yang diimputasikan pada nilai produksi yang *missing* ditampilkan pada Tabel 4.14 di bawah ini:

Tabel 4.14 Hasil Imputasi dengan Metode PMM

Perusahaan Ke-	Nilai Imputasi ke-					Rata-rata
	1	2	3	4	5	
5	331.500	309.000	309.000	309.000	309.000	313.500
7	424.320	424.320	331.500	331.500	331.500	368.628
13	309.000	289.800	289.800	289.800	289.800	293.640
23	538.200	538.200	488.750	488.750	488.750	508.530
26	488.750	488.750	484.380	488.750	488.750	487.876
28	196.560	196.560	196.560	196.560	196.560	196.560
30	982.800	982.800	982.800	982.800	982.800	982.800
36	1.430.000	1.430.000	1.465.100	1.465.100	1.430.000	1.444.400

c. Efisiensi Relatif

Tabel 4.15 Informasi Varian dari Imputasi Berganda

Varian			Fraction dari informasi Missing
Between	Within	Total	
4.660.022	8.282.457.604	8.288.049.630	0,000675

Varian *between imputation* sebesar 4.660.022 yang artinya adalah variasi antar imputasi satu dengan yang lainnya sebesar 4.660.022. Varian *within imputation* sebesar 8.282.457.604 yang artinya adalah variasi rata-rata dari 5 buah data lengkap yang tersedia sebesar 8.282.457.604. Nilai varian totalnya adalah sebesar 8.288.049.630.

Besarnya nilai Efisiensi Relatif pada metode PMM dengan melakukan imputasi sebanyak lima kali adalah :

$$ER = \left(1 + \frac{\lambda}{m}\right)^{-1} \times 100 \% = \left(1 + \frac{0,000675}{5}\right)^{-1} \times 100 \% = 99,9865 \%$$

4.3.1.4 Perbandingan Hasil Imputasi pada KIP 17115 (Kain tenun ikat) antara Metode Regresi dengan Metode PMM

Tabel 4.16 Hasil Imputasi dengan Metode Regresi dan PMM

Perusahaan Ke-	Tenaga Kerja	Bahan Bakar	Bahan Baku	Hasil Imputasi	
				Metode Regresi	Metode PMM
5	20	1.000	123.550	302.536	313.500
7	21	1.200	150.870	388.427	368.628
13	24	720	133.430	320.262	293.640
23	30	1.200	233.256	512.031	508.530
26	30	1.230	223.550	490.391	487.876
28	32	930	22.360	155.317	196.560
30	40	1.100	539.755	968.150	982.800
36	90	1.404	813.456	1.452.613	1.444.040

Secara umum nilai yang diimputasikan dengan menggunakan metode regresi maupun PMM tidak berbeda jauh. Untuk menguji apakah rata-rata nilai

imputasi dari kedua metode memiliki perbedaan yang signifikan dilakukan prosedur uji t data berpasangan dengan prosedur sebagai berikut :

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

Statsitik uji yang digunakan adalah :

$$t = \frac{\bar{d}}{s_d / \sqrt{n}}$$

Keputusan untuk menolak $H_0 : \mu_1 = \mu_2$ apabila $|t| > t_{\alpha/2, n-1}$.

Tabel 4.17 Penghitungan nilai d_j dan d_j^2

Metode Regresi	Metode PMM	d_j	d_j^2
302.536	313.500	-10.964	120.209.296
388.427	368.628	19.799	392.000.401
320.262	293.640	26.622	708.730.884
512.031	508.530	3.501	12.257.001
490.391	487.876	2.515	6.325.225
155.317	196.560	-41.243	1.700.985.049
968.150	982.800	-14.650	214.622.500
1.452.613	1.444.040	8.573	73.496.329
Total		-5.847	3.228.626.685
Rata-rata		-730,88	403.578.335,63

$$s_d = \sqrt{\frac{\sum_{j=1}^n d_j^2 - \left[\left(\sum_{j=1}^n d_j \right)^2 / n \right]}{n-1}}$$

$$= \sqrt{\frac{3.228.826.685 - (-5.847)^2 / 8}{7}}$$

$$= 21.462,10$$

statistik ujinya adalah

$$t = \frac{\bar{d}}{s_d / \sqrt{n}} = \frac{-730,88}{21.462,10 / \sqrt{8}} = -0,10$$

karena $t_{0,025;7} = 2,365$ diputuskan gagal menolak H_0 dan disimpulkan bahwa kedua metode memiliki rata-rata yang sama. Hal ini menunjukkan bahwa pada kondisi normal multivariat kedua metode sama-sama bekerja dengan baik.

Untuk membandingkan kebaikan dari kedua metode bisa dilihat dari nilai relatif efisiensi dari masing-masing metode. Pada bagian sebelumnya telah dibahas besarnya relatif efisiensi pada metode PMM adalah 99,9865 % adapun pada metode regresi sebesar 99,9715 %. Dari kedua nilai tersebut juga terlihat bahwa kedua metode memiliki efisiensi relatif yang hampir sama.

4.3.2 KIP 19201 (alas kaki)

Pada kelompok ini terdapat 79 responden perusahaan industri besar dan sedang. Dari ke-79 responden tersebut terdapat 54 perusahaan yang memberikan data lengkap. Adapun perusahaan yang hanya *missing* pada nilai produksinya sebanyak 10 perusahaan.

Apabila diperhatikan plot pada Gambar 4.10, hubungan antara produksi dengan variabel lain tidak semuanya memiliki hubungan dengan pola tertentu yang jelas bentuk hubungannya. Oleh karena itu dilakukan seleksi awal untuk mengetahui variabel mana saja yang memberikan pengaruh signifikan terhadap produksi dengan melihat koefisien korelasi antar variabel. Prosedur pengujian hipotesisnya adalah sebagai berikut :

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

Statistik uji untuk uji hipotesis di atas adalah :

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$$

Tolak H_0 apabila $|t| > t_{\alpha/2, n-2}$ yang artinya terdapat korelasi yang signifikan antara kedua variabel tersebut. Dari Tabel 4.18 di bawah ini nampak bahwa variabel listrik tidak mempunyai korelasi yang signifikan terhadap produksi.

Tabel 4.18 Korelasi Pearson Pada Variabel KIP 19201

		Produksi	Tenaga Kerja	Bahan Bakar	Listrik	Bahan Baku
Produksi	Korelasi	1	0,866	0,304	-0,027	0,995
	Signifikansi		0,000	0,013	0,423	0,000
Tenaga Kerja	Korelasi	0,866	1	0,357	-0,038	0,819
	Signifikansi	0,000		0,004	0,394	0,000
Bahan Bakar	Korelasi	0,304	0,357	1	-0,035	0,286
	Signifikansi	0,013	0,004		0,400	0,018
Listrik	Korelasi	-0,027	-0,038	-0,035	1	-0,023
	Signifikansi	0,423	0,394	0,400		0,435
Bahan Baku	Korelasi	0,995	0,819	0,286	-0,023	1
	Signifikansi	0,000	0,000	0,018	0,435	

Tahap selanjutnya, dengan meregresikan produksi sebagai variabel respon dengan keempat variabel lain sebagai variabel prediktor diperoleh output sebagai berikut :

Tabel 4.19 Output Regresi KIP 19201

Model	Koefisien Tidak Distandarisasi		Koefisien Distandarisasi	t	p-value
	Beta	Standar Error	Beta		
Konstanta	-199670,824	75151,827		-2,657	0,011
Tenaga Kerja	22417,149	1258,121	0,155	17,818	0,000
Bahan Bakar	0,064	7,427	0,000	0,009	0,993
Listrik	-0,275	0,819	-0,002	-0,336	0,738
Bahan Baku	1,154	0,011	0,869	102,582	0,000

Berdasarkan output pada Tabel 4.19 diperoleh informasi bahwa variabel bahan bakar dan listrik mempunyai *p-value* lebih besar dari α (0,05) sehingga H_0 yang menyatakan bahwa variabel tersebut tidak berpengaruh secara signifikan gagal ditolak. Artinya kedua variabel tersebut listrik tidak berpengaruh secara signifikan terhadap variabel produksi.

Berdasarkan kesimpulan bahwa variabel bahan bakar dan listrik tidak berpengaruh secara signifikan maka kedua variabel tersebut dikeluarkan dari model sehingga hanya menyertakan variabel tenaga kerja dan bahan baku. Dengan mengeluarkan kedua variabel dari model ternyata menurunkan nilai MSE dari 161.996.232.046 menjadi 156.003.723.561 seperti pada Tabel di bawah ini :

Tabel 4.20 ANOVA dengan Prediktor Listrik, Bahan Baku, Bahan Bakar dan Tenaga Kerja

Sumber	DF	SS	MS	F	<i>p-value</i>
Regresi	4	6.852.634.889.478.142	1.713.158.722.369.535	10.575	0,000
<i>Error</i>	49	7.937.815.370.273	161.996.232.046		
Total	53	6.860.572.704.848.415			

Tabel 4.21 ANOVA dengan Prediktor Bahan Baku dan Tenaga Kerja

Sumber	DF	SS	MS	F	<i>p-value</i>
Regresi	2	6.852.616.514.946.794	3.426.308.257.473.397	21.962	0,000
<i>Error</i>	51	7.956.189.901.621	156.003.723.561		
Total	53	6.860.572.704.848.415			

Dengan hanya menggunakan dua variabel prediktor yaitu tenaga kerja dan bahan baku, kemudian diambil perusahaan-perusahaan yang hanya *missing* pada nilai produksinya. Perusahaan-perusahaan tersebut ditampilkan pada Tabel 4.22 di bawah ini :

Tabel 4.22 Observasi dengan Nilai Produksi *Missing*

Perusahaan Ke-	Tenaga Kerja	Bahan Baku	Produksi
3	17	335.812	*
10	19	727.730	*
11	19	861.020	*
13	20	149.570	*
14	20	134.300	*
15	20	275.400	*
50	32	573.907	*
73	58	700.559	*
75	20	355.424	*
79	21	834.820	*

Untuk melakukan imputasi terhadap nilai *missing* tersebut, terlebih dahulu dilakukan uji asumsi terhadap 54 perusahaan yang memberikan data lengkap

(lampiran 10). Pengujian asumsi ini dilakukan dengan cara melakukan uji normalitas (uji Kolmogorov Smirnov) terhadap residual dari regresi antara produksi sebagai variabel respon dengan tenaga kerja dan bahan baku sebagai variabel prediktor.

Pada Tabel 4.23 terlihat bahwa hasil uji Kolmogorov Smirnov terhadap residual regresi pada KIP 19201 menghasilkan *p-value* sebesar 0,409. Nilai ini lebih besar dari nilai α (0,05) sehingga H_0 yang menyatakan bahwa residual berdistribusi normal gagal ditolak dan dapat disimpulkan bahwa residual adalah berdistribusi normal.

Tabel 4.23 Hasil Uji Kolmogorov Smirnov Residual KIP 19201

Keterangan		Residual
N		54
Parameter Normal	Rata-rata	0,000
	Standar Deviasi	387,449
Nilai Ekstrim	Absolut	0,121
	Positif	0,092
	Negatif	- 0,121
Kolmogorov Smirnov Z		0,888
<i>p-value</i>		0,409

4.3.2.1 Model Regresi pada KIP 19201 (alas kaki)

Dari ke-54 perusahaan yang memberikan data lengkap yang telah lolos dari deteksi *outlier* model regresi berganda yang terbentuk adalah :

$$\hat{Y} = -203.494,147 + 22.433,228 X_1 + 1,154 X_4 \quad (4.14)$$

dengan Y adalah nilai produksi

X_1 adalah banyaknya tenaga kerja

X_4 adalah nilai bahan baku yang digunakan

4.3.2.2 Imputasi pada KIP 19201 (alas kaki) dengan Metode Regresi

Dengan menggunakan Paket Program SAS 9.0 PROC MI dengan menggunakan *seed* 12304 dilakukan imputasi sebanyak 5 kali diperoleh output sebagai berikut :

a. Model pada Setiap Imputasi

Tabel 4.24 Koefisien Regresi pada KIP 19201

Efek	Data Lengkap	Imputasi ke-				
		1	2	3	4	5
β_0	-0,03844	-0,030957	-0,034817	-0,037645	-0,049291	-0,040784
X_1	-0,16083	-0,155125	-0,144012	-0,169886	-0,150607	-0,144616
X_4	1,05678	1,056089	1,042866	1,059752	1,052169	1,041333

Dari output di atas dapat dituliskan model regresi untuk data lengkap adalah :

$$\hat{Y} = -0,03844 - 0,16083 X_1 + 1,05678 X_4. \quad (4.15)$$

Model yang dipergunakan dalam setiap imputasi :

$$\text{Imputasi ke-1 : } \hat{Y}_{i*} = -0,030957 - 0,155125 X_1 + 1,056089 X_4 \quad (4.16)$$

$$\text{Imputasi ke-2 : } \hat{Y}_{i*} = -0,034817 - 0,144012 X_1 + 1,056089 X_4 \quad (4.17)$$

$$\text{Imputasi ke-3 : } \hat{Y}_{i*} = -0,037645 - 0,169886 X_1 + 1,059752 X_4 \quad (4.18)$$

$$\text{Imputasi ke-4 : } \hat{Y}_{i*} = -0,049291 - 0,150607 X_1 + 1,052169 X_4 \quad (4.19)$$

$$\text{Imputasi ke-5 : } \hat{Y}_{i*} = -0,040784 - 0,144616 X_1 + 1,041333 X_4 \quad (4.20)$$

b. Nilai yang diimputasikan

Nilai yang diimputasikan pada nilai produksi yang *missing* adalah :

Tabel 4.25 Hasil Imputasi dengan Metode Regresi

Perusahaan Ke-	Nilai Imputasi ke-					Rata-rata
	1	2	3	4	5	
3	133.000	224.212	-555.587	106.642	296.704	40.994
10	57.308	540.755	335.275	250.863	20.928	241.026
11	321.212	792.114	693.433	1.184.809	451.581	688.630
13	444.338	-231.975	-329.079	35.105	25.977	-11.127
14	637.370	-136.947	-222789	95.758	317.836	138.245
15	-421.066	346.133	-79.063	-199.534	56.054	-59.495
50	546.058	-76.754	-117.603	223.293	-164.001	82.199
73	-391.039	-389.123	-65.005	476.567	-444.487	-162.617
75	507.791	21.050	861.144	-712.955	-26.969	130.012
79	646.666	582.807	1.200.756	898.067	586.602	782.980

c. Efisiensi Relatif

Dari Tabel 4.26 di bawah ini, bisa diperoleh beberapa informasi sebagai berikut:

Tabel 4.26 Informasi Varian dari Imputasi Berganda

Varian			Fraction dari informasi Missing
Between	Within	Total	
75.449.087	972.521.789.175	972.612.328.080	0,000093093

Besarnya nilai Efisiensi Relatif pada metode regresi dengan melakukan imputasi sebanyak lima kali adalah :

$$ER = \left(1 + \frac{\lambda}{m}\right)^{-1} \times 100\% = \left(1 + \frac{0,000093093}{5}\right)^{-1} \times 100\% = 99,9981\%$$

4.3.2.3 Imputasi pada KIP 19201 (alas kaki) dengan Metode PMM

Dengan menggunakan Paket Program SAS 9.0 PROC MI dengan menggunakan *seed* 12304 dilakukan imputasi sebanyak 5 kali. Output lengkap pada Lampiran 5.

a. Model pada Setiap Imputasi

Tabel 4.27 Koefisien Regresi pada Metode PMM

Efek	Data Lengkap	Imputasi ke-				
		1	2	3	4	5
β_0	-0,03844	-0,030957	-0,044142	-0,030219	-0,038436	-0,036939
X_1	-0,16083	-0,155125	-0,167484	-0,166063	-0,152132	-0,157532
X_4	1,05678	1,056089	1,068716	1,067404	1,060134	1,057835

Dari output di atas dapat diperoleh informasi bahwa model untuk data lengkap adalah :

$$\hat{Y} = -0,03844 - 0,16083 X_1 + 1,05678 X_4 \quad (4.21)$$

Model yang dipergunakan dalam setiap imputasi :

$$\text{Imputasi ke-1 : } \hat{Y}_* = -0,030957 - 0,155125 X_1 + 1,056089 X_4 \quad (4.22)$$

$$\text{Imputasi ke-2 : } \hat{Y}_* = -0,044142 - 0,167484 X_1 + 1,068716 X_4 \quad (4.23)$$

$$\text{Imputasi ke-3 : } \hat{Y}_{i*} = -0,030219 - 0,166063 X_1 + 1,067404 X_4 \quad (4.24)$$

$$\text{Imputasi ke-4 : } \hat{Y}_{i*} = -0,038436 - 0,152132 X_1 + 1,060134 X_4 \quad (4.25)$$

$$\text{Imputasi ke-5 : } \hat{Y}_{i*} = -0,036939 - 0,157532 X_1 + 1,057835 X_4 \quad (4.26)$$

b. Nilai yang diimputasikan

Persamaan 4.21 sampai dengan persamaan 4.26 digunakan untuk menentukan nilai yang akan diimputasikan pada *missing data* yang terdapat pada variabel produksi pada KIP 19201. Nilai-nilai yang diimputasikan pada variabel produksi yang *missing* ditampilkan pada Tabel 4.28 di bawah ini:

Tabel 4.28 Hasil Imputasi dengan Metode PMM

Perusahaan Ke-	Nilai Imputasi ke-					Rata-rata
	1	2	3	4	5	
3	189.447	98.270	189.447	98.270	164.124	147.912
10	475.560	384.376	514.296	408.237	425.388	441.571
11	604.940	514.296	604.940	514.296	558.678	559.430
13	31.821	31.821	31.821	31.821	31.821	31.821
14	31.821	31.821	31.821	31.821	31.821	31.821
15	55.748	31.821	98.270	31.821	31.821	49.896
50	98.270	31.821	98.270	33.552	55.748	63.532
73	31.821	31.821	31.821	31.821	31.821	31.821
75	164.124	33.552	164.124	55.748	98.270	103.164
79	536.048	425.388	558.678	452.175	475.560	489.570

c. Efisiensi Relatif

Dari output di bawah ini, dapat diperoleh beberapa informasi sebagai berikut:

Tabel 4.29 Informasi Varian dari Imputasi Berganda

Varian			Fraction dari informasi Missing
Between	Within	Total	
24.063.598	972.097.832.390	972.126.708.707	0,000029705

Besarnya nilai Efisiensi Relatif pada metode PMM dengan melakukan imputasi sebanyak lima kali adalah :

$$ER = \left(1 + \frac{\lambda}{m}\right)^{-1} \times 100 \%$$

$$= \left(1 + \frac{0,000029705}{5}\right)^{-1} \times 100 \% = 99,9994 \%$$

4.3.2.4 Perbandingan Hasil Imputasi pada KIP 19201 (alas kaki) antara Metode Regresi dengan Metode PMM

Tabel 4.30 Hasil Imputasi dengan Metode Regresi dan PMM

Perusahaan Ke-	Tenaga Kerja	Bahan Baku	Hasil Imputasi	
			Metode Regresi	Metode PMM
3	17	335.812	40.994	147.912
10	19	727.730	241.026	441.571
11	19	861.020	688.630	559.430
13	20	149.570	-11.127	31.821
14	20	134.300	138.245	31.821
15	20	275.400	-59.495	49.896
50	32	573.907	82.199	63.532
73	58	700.559	-162.617	31.821
75	20	355.424	130.012	103.164
79	21	834.820	782.980	489.570

Apabila diperhatikan nilai yang diimputasikan dengan menggunakan metode regresi ternyata muncul nilai-nilai negatif pada responden 13, 15 dan 73. Tentunya nilai ini menjadi aneh sebab nilai tersebut adalah nilai produksi yang dihasilkan oleh sebuah perusahaan dengan tenaga kerja dan bahan baku seperti tercantum dalam Tabel 4.30 di atas.

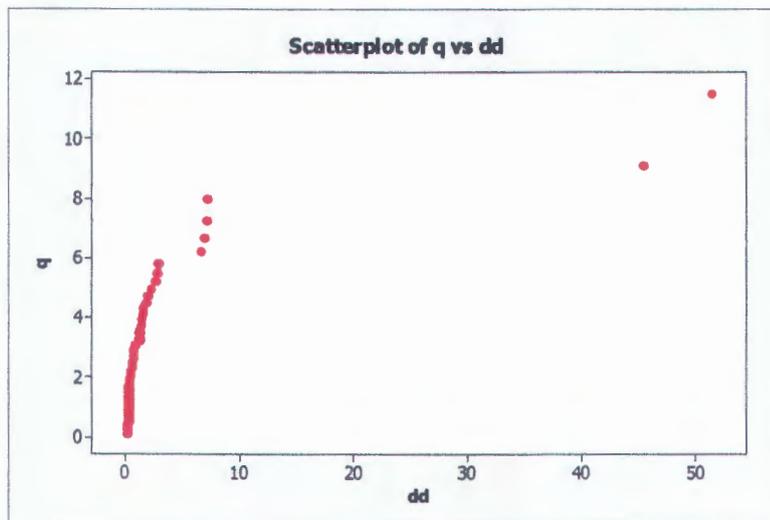
Berbeda dengan metode regresi yang muncul nilai-nilai negatif, pada metode PMM tidak muncul nilai negatif. Tentunya ini lebih masuk akal bila dibandingkan dengan metode regresi. Meskipun demikian, terjadi kesamaan nilai-nilai yang diimputasikan pada responden 13, 14 dan 73.

Keanehan nilai-nilai yang diimputasikan dari kedua metode ternyata disebabkan oleh tidak terpenuhinya asumsi normal multivariat yang disyaratkan. Hal ini terlihat dari hasil uji normalitas Kolmogorov Smirnov pada Tabel 4.31 yang menunjukkan bahwa *p-value* dari semua variabel kurang dari nilai α (0,05) sehingga disimpulkan dibawah tidak satupun variabel berdistribusi normal.

Tabel 4.31 Output Uji Kolmogorov Smirnov Variabel pada KIP 19201

Keterangan	Tenaga Kerja	Bahan Baku	Produksi
Parameter Normal Rata-rata	48,352	3.458.900	2.232.900
Standar Deviasi	78,573	11.377.400	8.560.590
Nilai Ekstrim Absolut	0,337	0,403	0,410
Positif	0,337	0,403	0,410
Negatif	-0,336	-0,387	-0,399
Kolmogorov Smirnov Z	2,474	2,960	3,012
<i>p-value</i>	0,000	0,000	0,000

Scatter Plot uji normal multivariat dengan menggunakan *QQ-test* (makro pada Lampiran 4) ditampilkan pada Gambar 4.15. Dari Gambar 4.15 tersebut terlihat bahwa data tidak berdistribusi normal multivariat. Dengan demikian dapat disimpulkan bahwa data pada KIP 19201 tidak berdistribusi normal multivariat.



Gambar 4.15 *Scatter Plot* Uji Normal Multivariat KIP 19201

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan dapat diambil beberapa kesimpulan sebagai berikut :

1. Dari hasil studi simulasi, pada data yang berdistribusi normal multivariat metode imputasi regresi dan metode *predictive mean matching* secara umum menunjukkan kinerja yang sama baik. Namun, pada data yang tidak berdistribusi normal multivariat metode PMM secara umum menunjukkan kinerja yang lebih baik dibandingkan metode imputasi regresi.
2. Dari hasil studi kasus, pada data KIP 17115 yang berdistribusi normal multivariat hasil imputasi dari kedua metode tidak terdapat perbedaan yang signifikan. Artinya kinerja kedua metode dalam melakukan imputasi terhadap *missing data* adalah sama baik. Sedangkan pada data KIP 19201 yang tidak berdistribusi normal multivariat, hasil imputasi dengan metode PMM lebih masuk akal karena hasil imputasi dari metode imputasi regresi muncul nilai-nilai negatif. Artinya kinerja metode PMM dalam kasus ini adalah lebih baik bila dibandingkan dengan metode imputasi regresi.

5.2 Saran

Berdasarkan penelitian yang telah dilakukan diberikan beberapa saran sebagai berikut :

1. Kedua metode imputasi akan memberikan hasil yang baik pada kondisi normal multivariat. Kondisi ini cukup rentan dengan adanya data *outlier*. Oleh karena itu disarankan untuk melakukan kajian terhadap metode imputasi yang *robust* terhadap *outlier* serta tidak membutuhkan asumsi kenormalan.
2. Pada penelitian ini baru difokuskan pada observasi dengan *missing data* berpola monoton sehingga perlu dilakukan kajian terhadap observasi dengan *missing data* berpola non monoton dan *arbitrary*.
3. Perlu dilakukan penelitian seberapa besar prosentase *missing data* yang masih bisa ditangani oleh kedua metode.

DAFTAR PUSTAKA

- Allison, P.D. (2000), *Multiple Imputation for Missing Data: A Cautionary Tale*, University of Pensilvania, 3718 Locust Walk, Philadelphia.
- Badan Pusat Statistik. (2005), *Statistik Industri Besar dan Sedang*, BPS, Jakarta.
- Box, G.E.P., dan Tiao, G.C. (1973), *Bayesian Inference in Statistical Analysis* Reading Mass: addison-Wesley.
- Buuren, S.V. (2007), Multiple Imputation of Discrete and Continuous Data by Fully Conditional Specification. *Journal of Statistical Methods in Medical Research*, 16, 219-242.
- Draper, N.R., dan Smith, H. (1981), *Applied Regression Analysis* 2nd edition. John Willey and Sons.
- Hair, J. (1998), *Multivariate Data analysis*. Prentice Hall.
- Hines, W.W., dan Montgomery, D.C., (1989), *Probability and Statistics in Engineering and Management Science*. John Willey and Sons.
- Homecillo, L.V. (2004), Treatment of Missing Data in the Annual Survey of Establishment, *Paper presented at 9th National Convention on Statistic (NCS) EDSA Shangri-La Hotel*, October 4-5, 2004.
- Horton, N.J., and Lipsitz, S.R. (2001), Multiple Imputation in practice: Comparison of Software Package for Regression Model With Missing Variables, *Journal of the American Statistik Association*, 55, 244-254.
- Little, R.J.A., dan Rubin, DA (1987), *Statistical Analysis with Missing Data*. New York: John Willey and Sons.
- Madow, W.G., Olkin I., Ford, B.L., dan Rubin, D.B., (1983), *An Overview of hot-deck procedures in complete data surveys*. Academic Press, New York, pp.185-207.
- Magnani, M. (2004), *Techniques for Dealing with Missing Data in Knowledge Discovery Tasks*. <http://magnanim.web.cs.unibo.it/index.html>. Didownload pada tanggal 12 Mei 2009.
- Maksum, C., dan Warde, W.D. (1991), *Comparison of Several Imputation Methods*: Ph.D.thesis, Oklahoma State University.
- Pedlow, S., Luke. J.V., dan Blumberg S.J. (2007), *Multiple Imputation of Missing Household Poverty Level Values from the National Survey of Children*

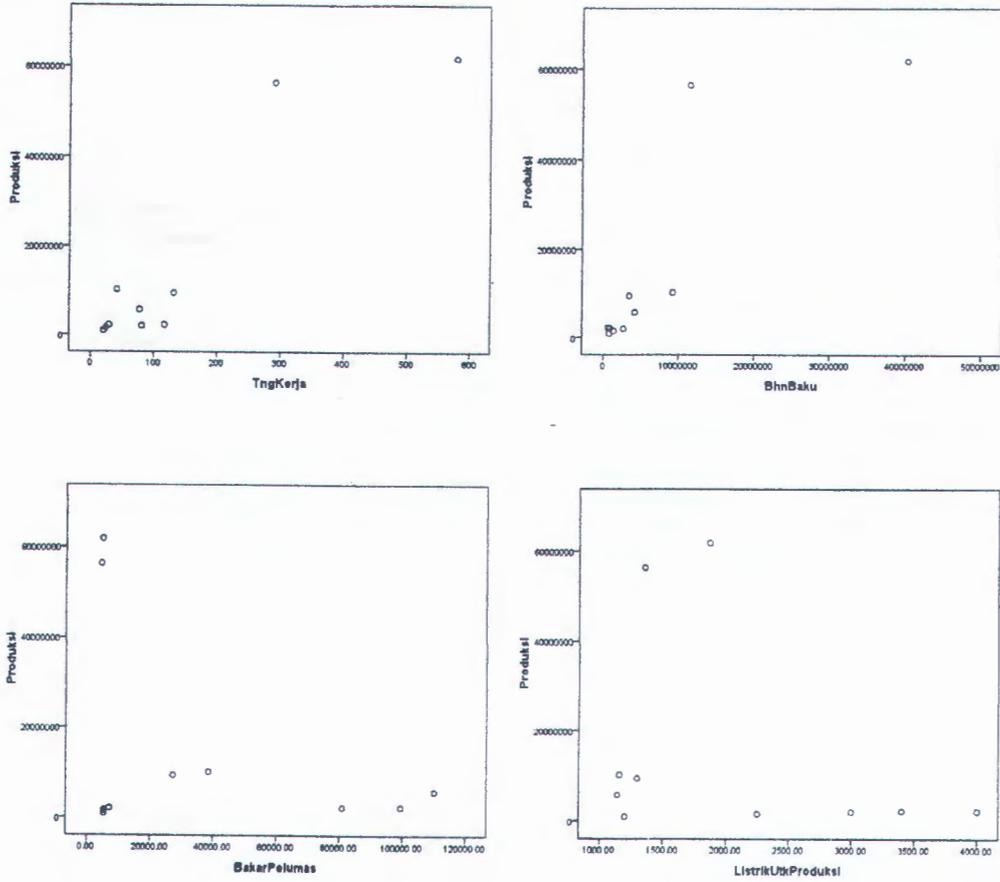
with Special Health Care Needs, 2001, and the National Survey of Children's Health, 2003. Departemen Of Health And Human Service Centers for Disease Control and Prevention, NSCH.

- Qin, Y., Zhang, S., Zhu, X., Zhang, J., dan Zhang, C. (2009), POP Algorithm: Kernel based Imputation to Treat Missing Values in Knowledge Discovery from Database, *Journal of Expert System with Applications*, 36, 2794-2804.
- Rubin, D.B. (1988) *An Overview of Multiple Imputation*, Harvard University One Oxford Street, Cambridge, MA 02138.
- Rubin, D.B. (1987), *Multiple Imputation for Non Response in Surveys*. New York: John Willey and Sons.
- Rubin, D.B. (1996), Multiple Imputation After 18+ Years, *Journal of The American Statistical Association*, 91, 473-489.
- Schafer, J. (1999), *NORM: Multiple Imputation of Incomplete Multivariate Data under a Normal Model. Version 2. Software for Windows 95/98/NT*. Available from: <http://www.stat.psu.edu/~jls/misoftwa.html>. Didownload pada tanggal 12 Mei 2009.
- Schoier, G (2001), *On Partial Non Response Situation : Hot Deck Imputation Method*, Universita di Trieste, Italia.
- Walks, S. (1932), Moments and Distributions of Estimates of Population Parameters from Fragmentary Samples. *Annals of Mathematical Statistics*, 3, 163-203.
- Wayman, J. C. (2003), *Multiple Imputation for Missing Data: What Is It And How Can I Use It ?*, Paper Presented at the 2003 Annual Meeting of the American Educational Research Association, Chicago, IL.
- Weiss, N.A. (2002), *Elementary Statistics*, 5th edition, Pearson Education, Inc. United States of America
- Yuan, Y.C. (2001), *Multiple Imputation for Missing Data : Concept and New Development SAS/STAT 8.2*. Cary NC: SAS Institute Inc. <http://www.sas.com/statistics>
- Zhang, S, Qin Y., Zhu, X., Zhang, J., dan Zhang, C.(2007), *Kernel Based Multi-Imputation for Missing Data*, University of Technology Sidney.

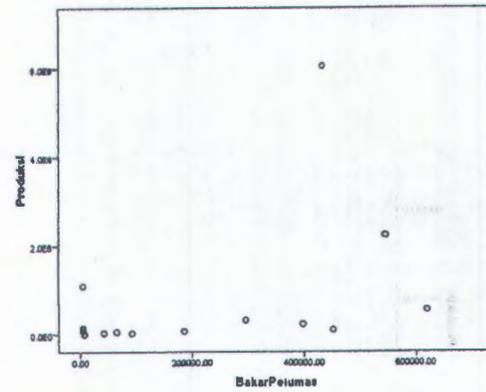
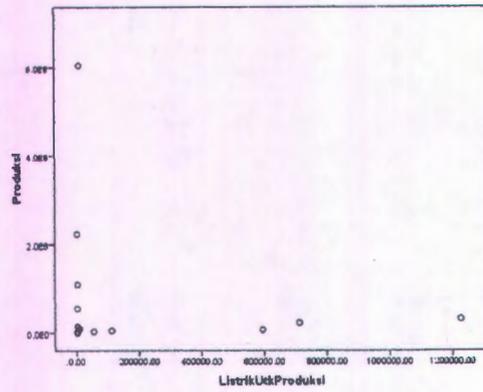
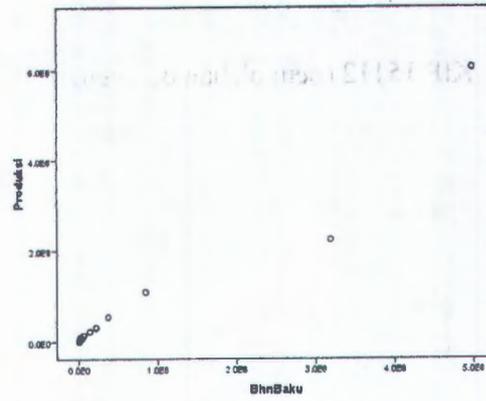
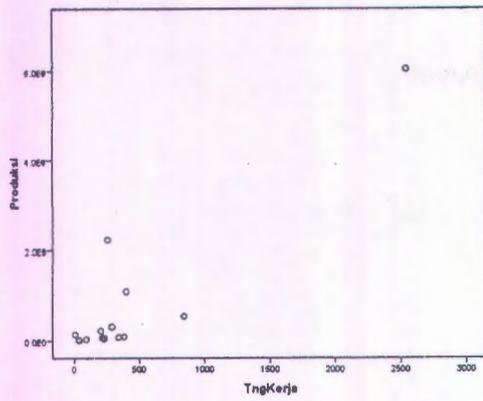
LAMPIRAN 1

Plot Hubungan antara variabel produksi dengan jumlah tenaga kerja, bahan baku, tenaga listrik, serta bahan bakar Menurut Kode Induk Perusahaan (KIP)

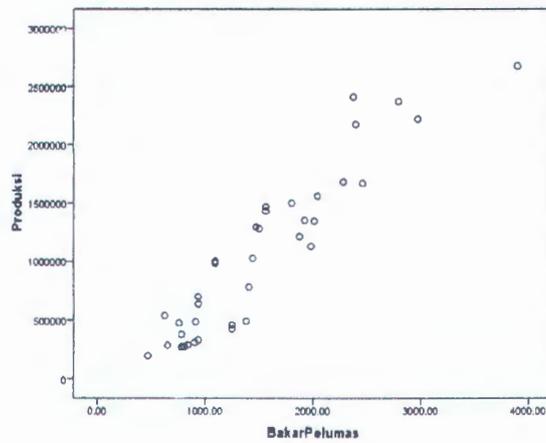
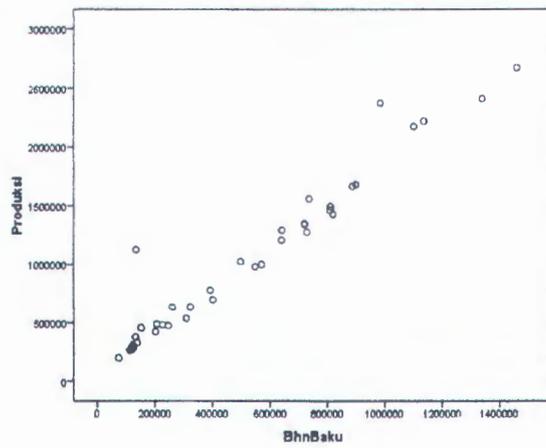
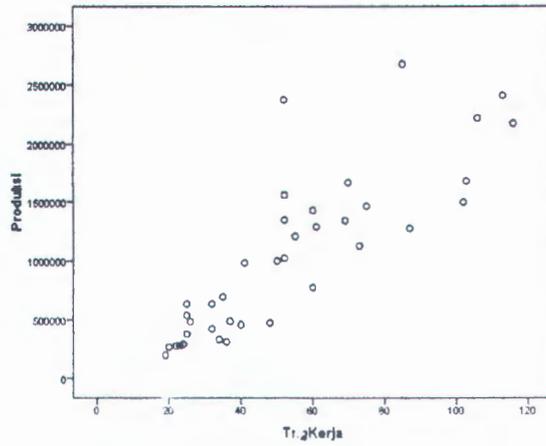
KIP 15112 (pengolahan dan pengawetan daging)



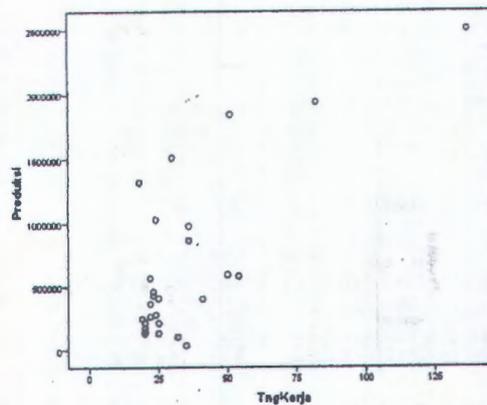
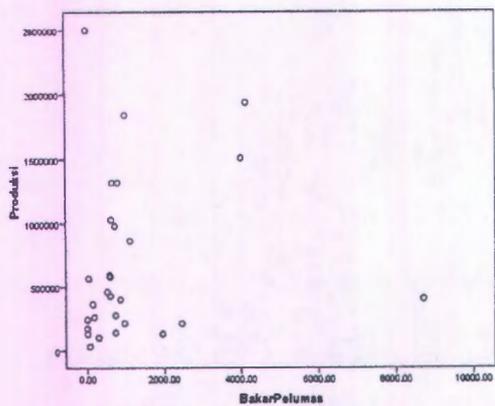
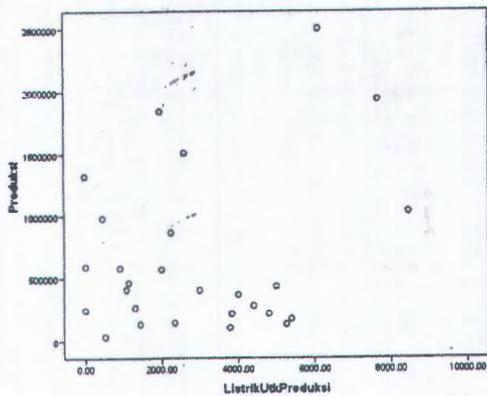
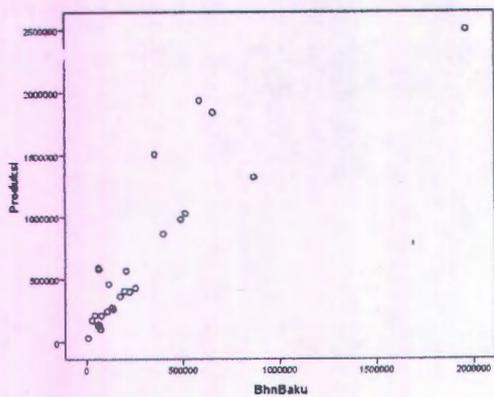
KIP 15121 (pengawetan ikan dan biota perairan lainnya)



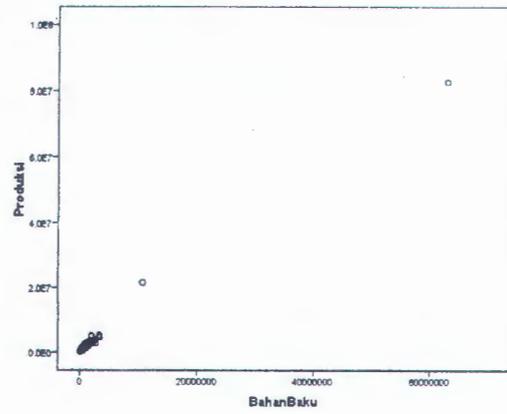
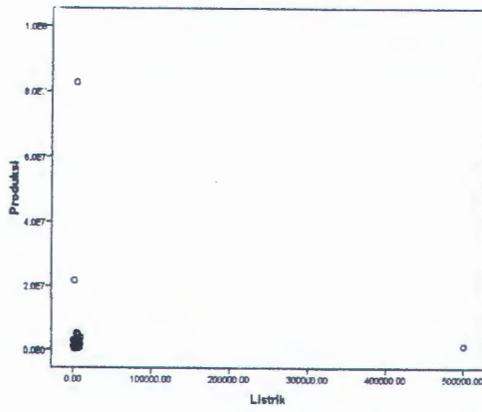
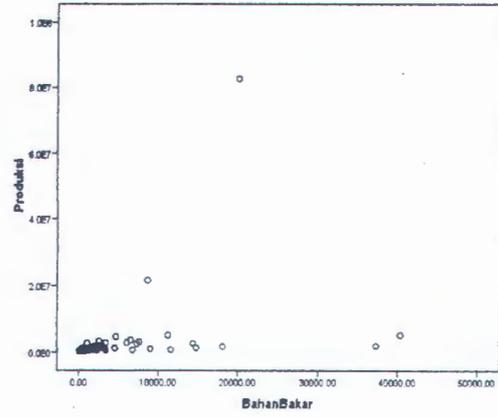
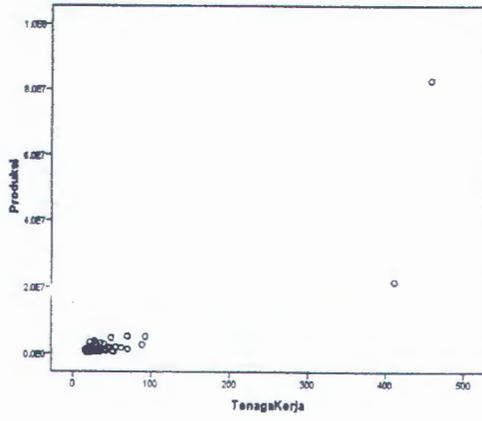
KIP 17115 (kain tenun ikat)



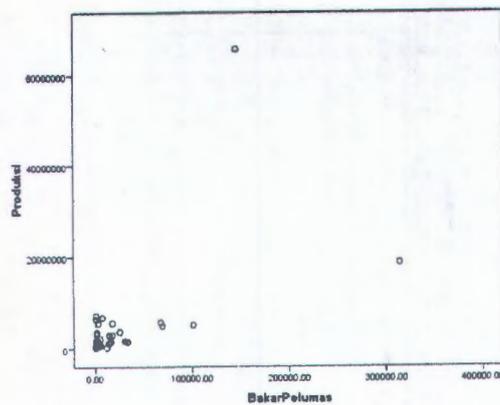
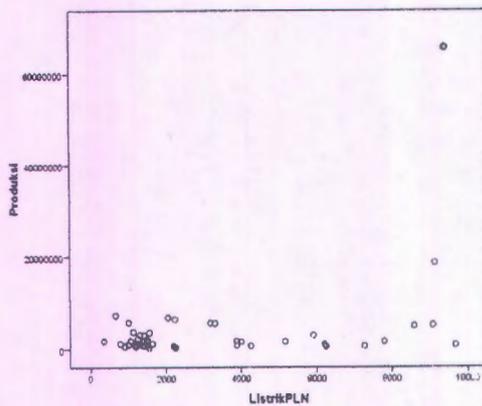
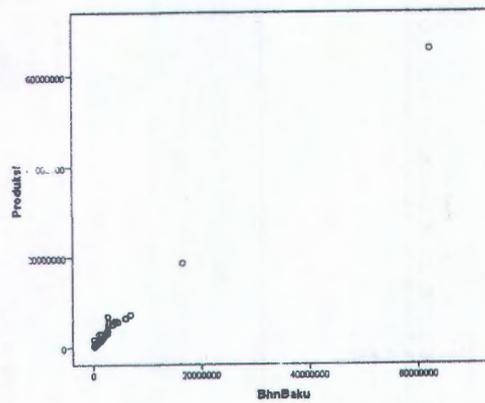
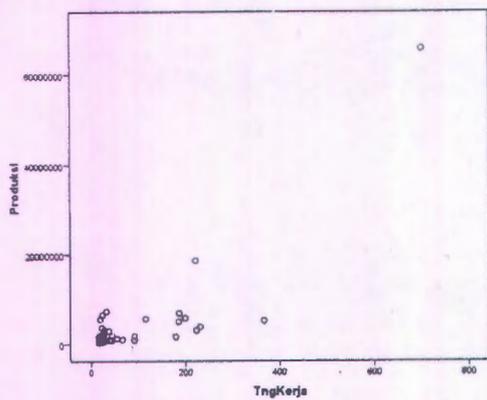
KIP 17293 (bordir/sulaman)



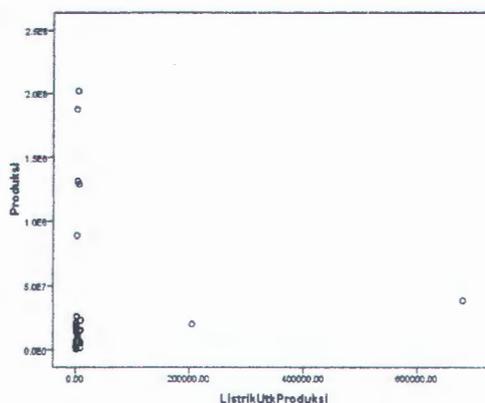
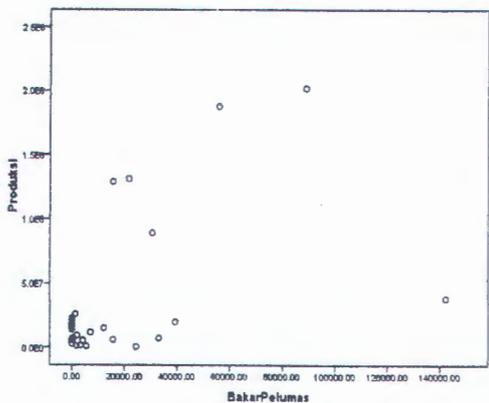
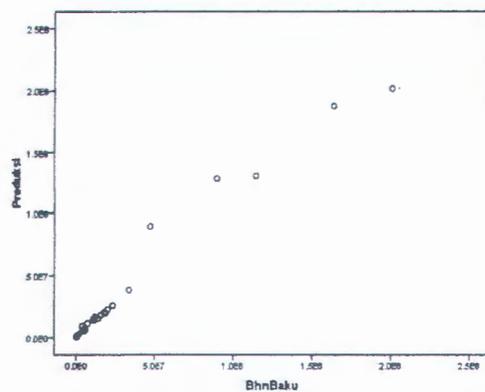
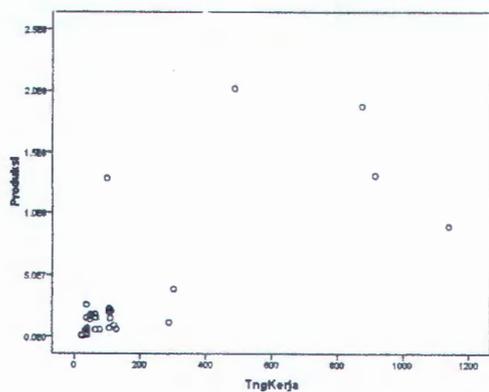
KIP 19201 (alas kaki)



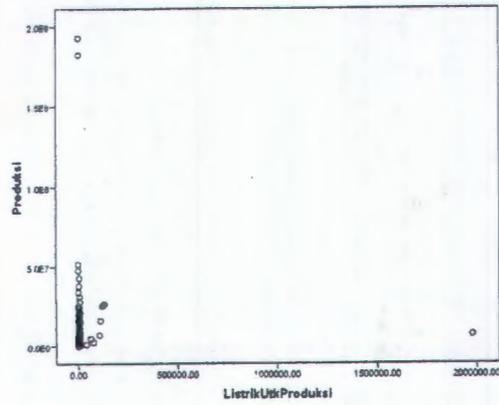
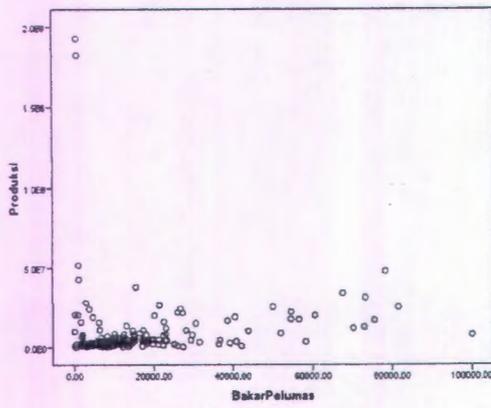
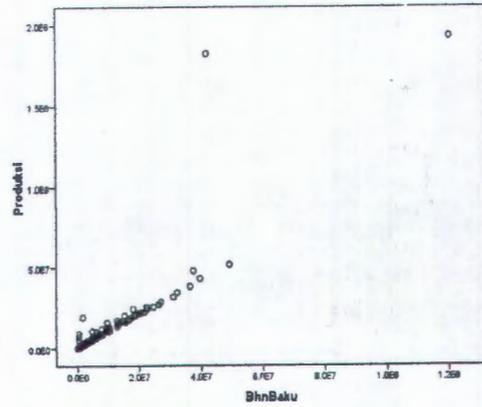
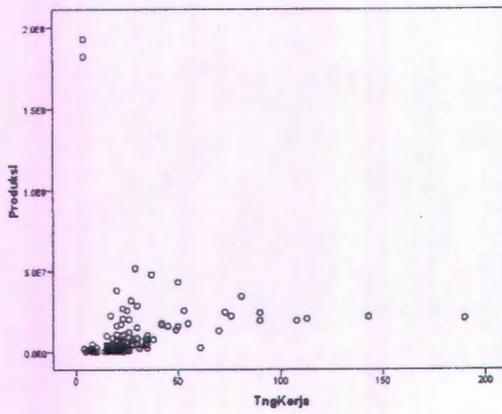
KIP 15122 (penggaraman/pengeringan ikan dan biota perairan lainnya)



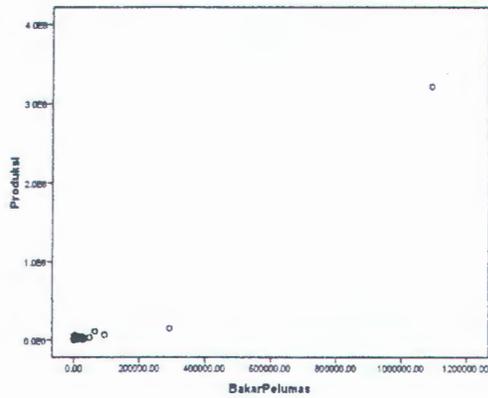
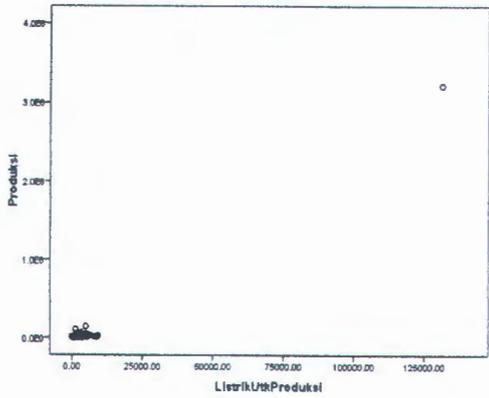
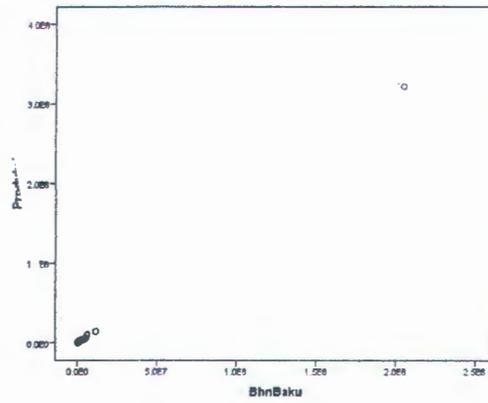
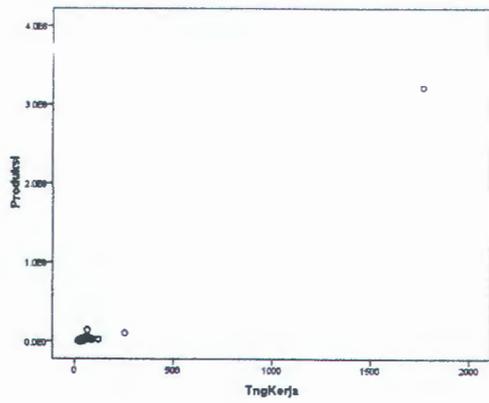
KIP 15124 (pembekuan ikan dan biota perairan lainnya)



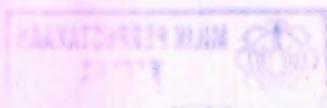
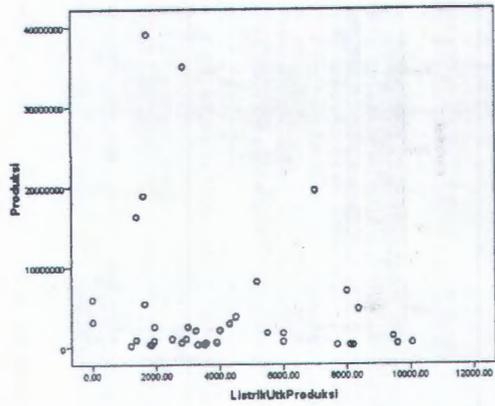
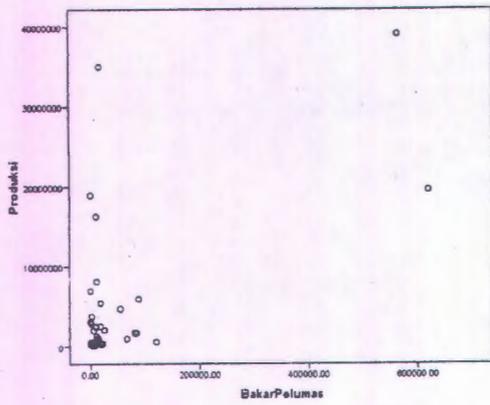
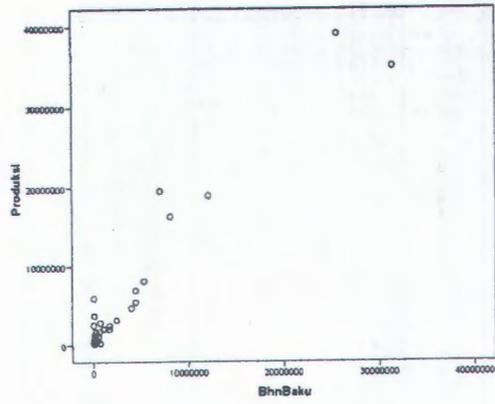
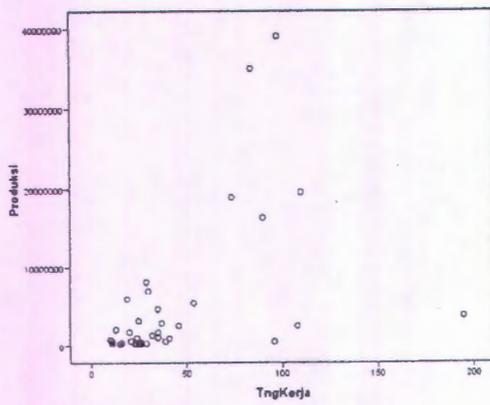
KIP 15311 (penggilingan padi dan penyosohan beras)



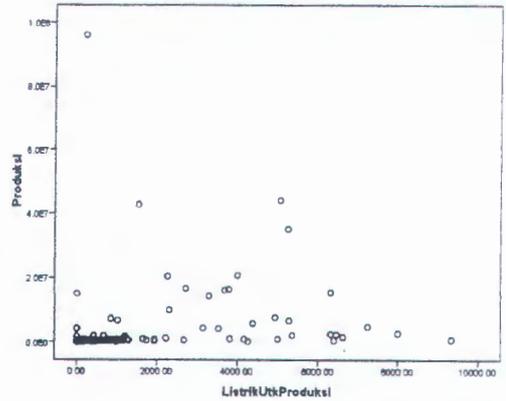
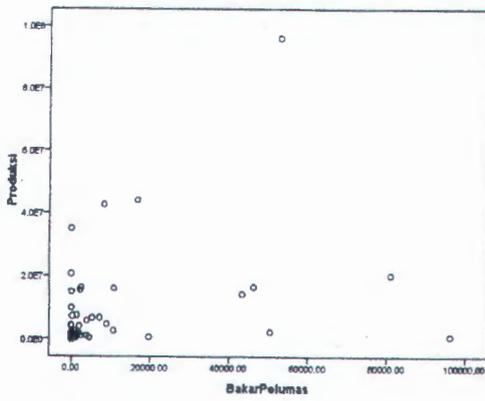
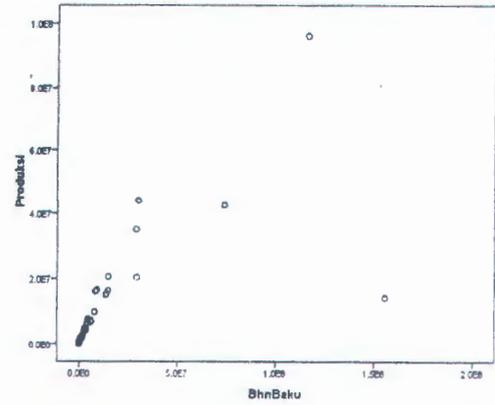
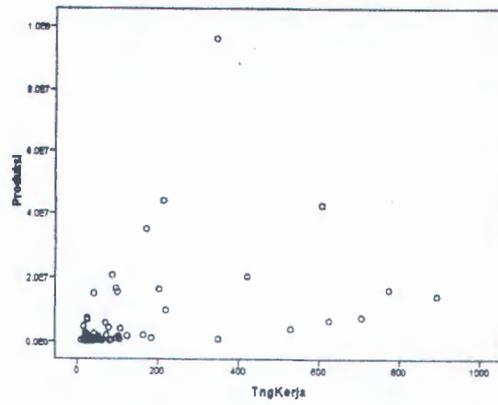
KIP 15440 (makaroni, mie, bihun dan sejenisnya)



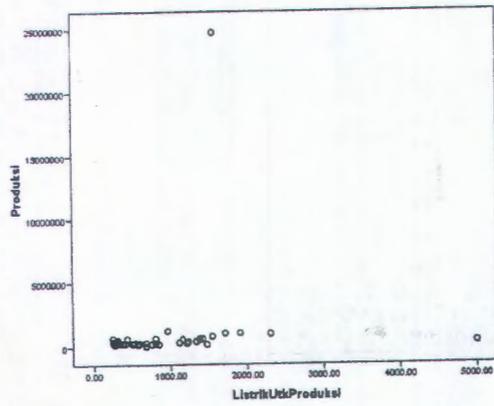
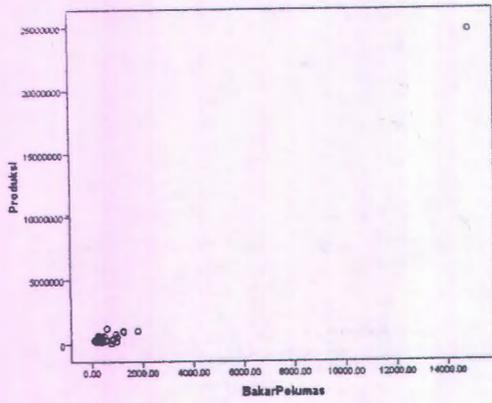
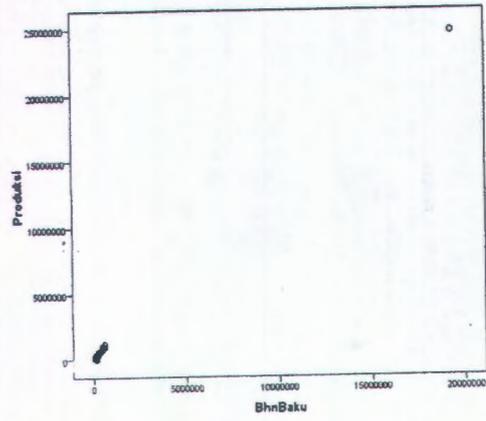
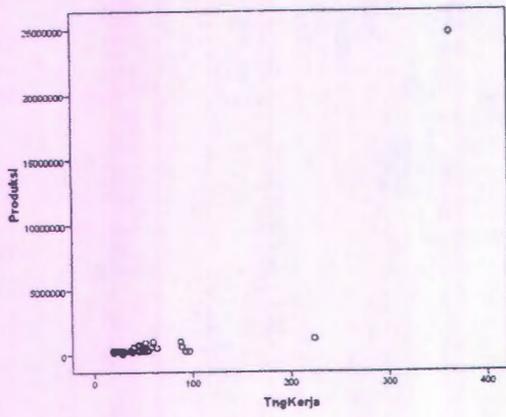
KIP 15540 (minuman ringan)



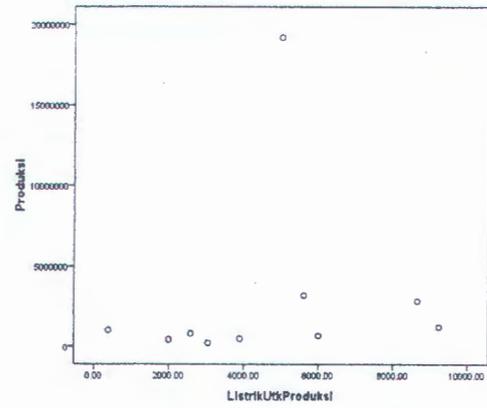
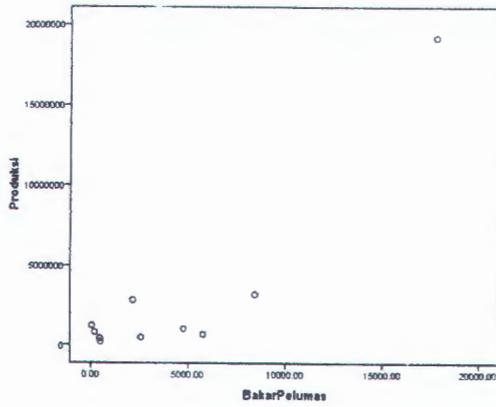
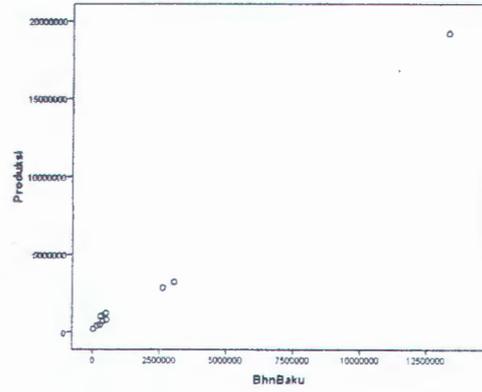
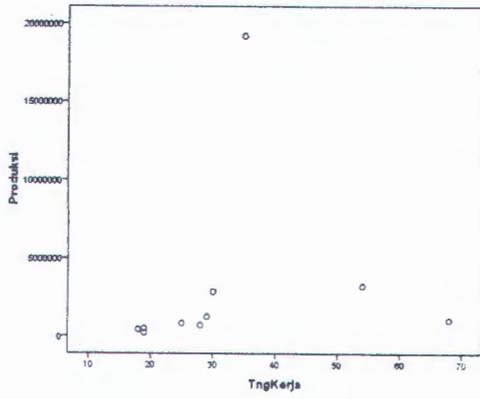
KIP 16001 (pengeringan dan pengolahan tembakau dan bumbu rokok)



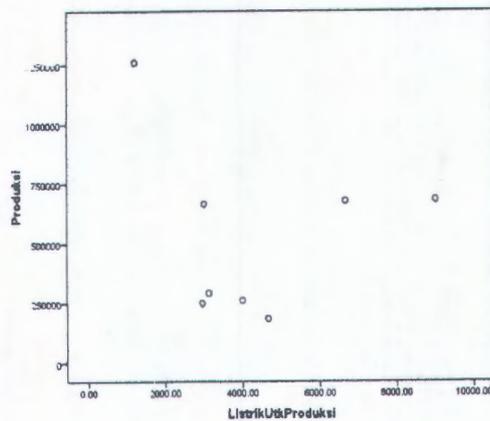
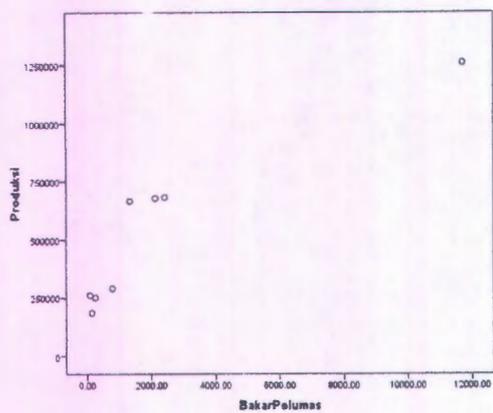
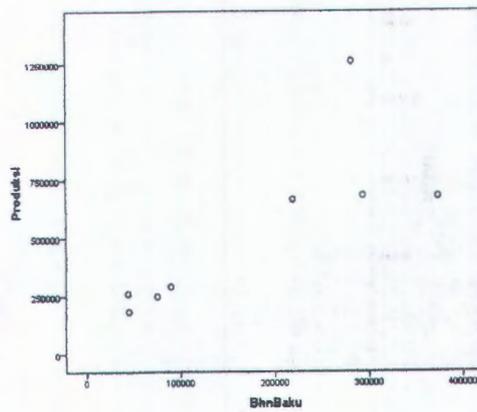
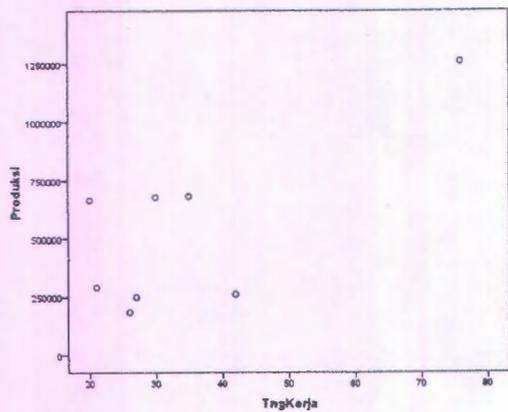
KIP 17114 (pertenunan)



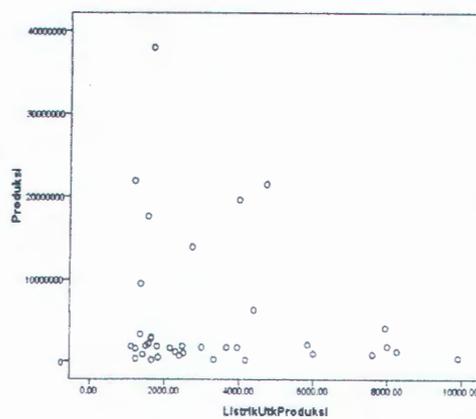
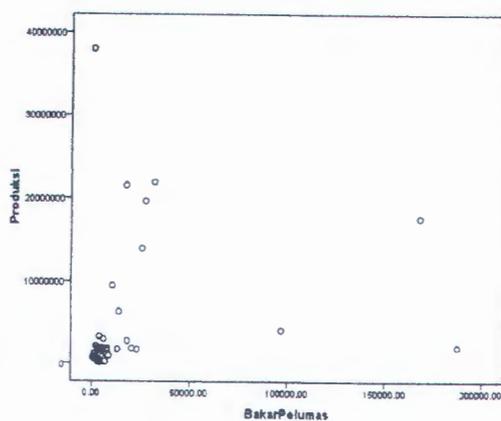
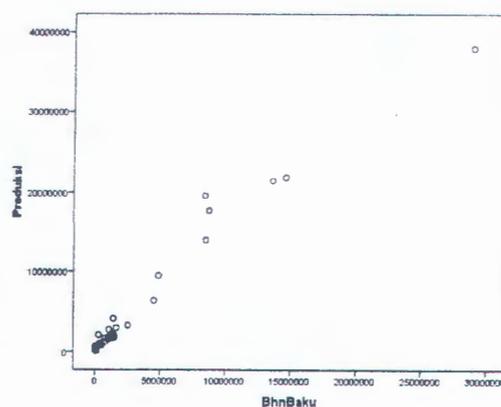
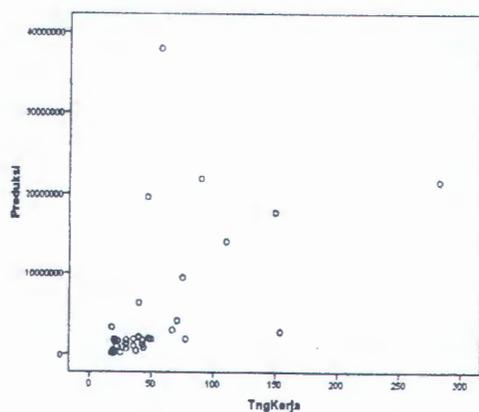
KIP 17299 (tekstil yang tidak dikelompokkan di tempat lain)



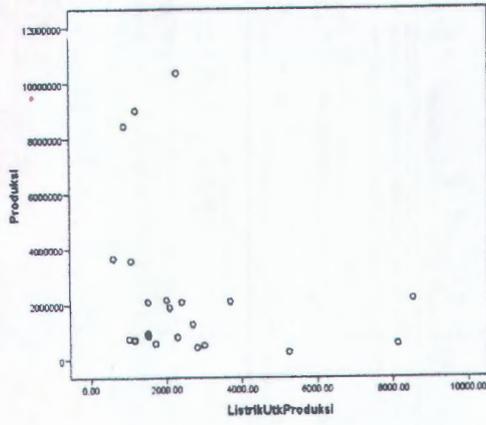
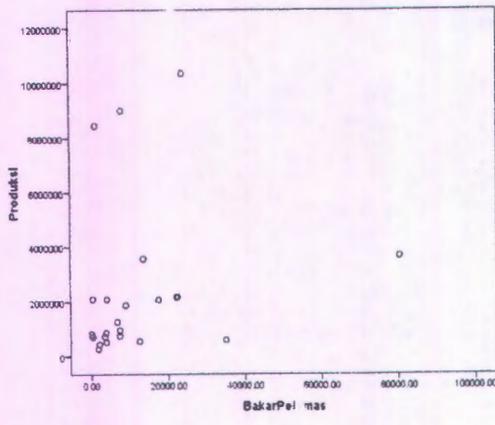
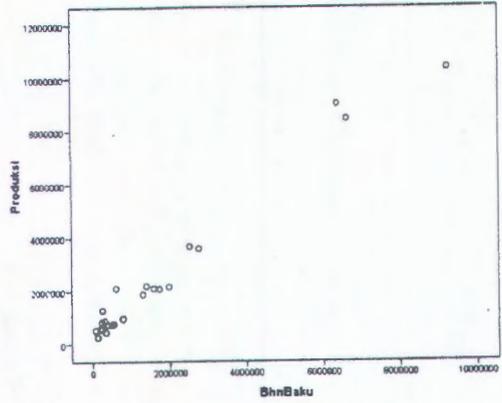
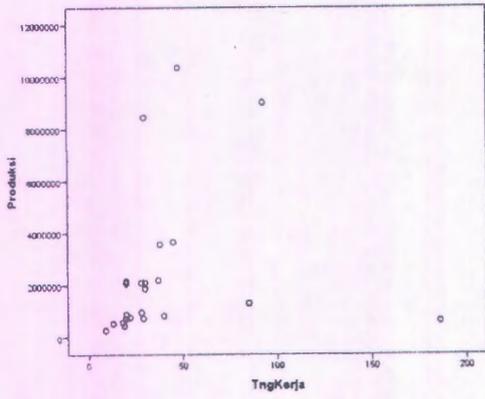
KIP 18102 (pakaian jadi lainnya dari tekstil)



KIP 25209 (barang-barang plastik lainnya)



KIP 37200 (daur ulang barang-barang bukan logam)



LAMPIRAN 2

Data 39 Perusahaan KIP 17115 yang Memberikan Data Lengkap

No	Perusahaan Ke-	Tenaga Kerja	Bahan Bakar	Bahan Baku	Produksi
1	37	19	468	71.604	196.560
2	38	20	780	111.644	265.200
3	39	22	804	119.810	273.360
4	40	23	780	119.210	276.250
5	41	24	650	118.320	287.300
6	42	24	840	125.488	289.800
7	43	36	900	122.200	309.000
8	44	34	936	137.832	331.500
9	45	25	780	131.690	377.600
10	46	32	1.248	200.824	424.320
11	47	40	1.248	151.216	457.600
12	48	48	756	246.284	475.200
13	49	26	910	225.472	484.380
14	50	37	1.380	204.688	488.750
15	51	25	624	308.880	538.200
16	52	25	936	261.040	635.700
17	53	32	936	323.752	635.700
18	54	35	936	401.336	696.800
19	55	60	1.404	392.730	777.660
20	56	41	1.092	547.196	982.800
21	57	50	1.092	570.128	1.001.000
22	58	52	1.440	498.080	1.025.800
23	59	73	1.980	133.784	1.128.000
24	60	55	1.872	639.600	1.210.560
25	61	87	1.500	727.332	1.277.100
26	62	61	1.470	640.500	1.293.600
27	63	69	2.010	719.250	1.342.875
28	64	52	1.920	719.220	1.349.200
29	65	60	1.560	818.688	1.430.000
30	66	75	1.560	808.288	1.465.100
31	67	102	1.800	809.320	1.495.600
32	68	52	2.040	735.608	1.558.800
33	69	70	2.460	886.632	1.667.500
34	70	103	2.280	898.460	1.680.000
35	71	116	2.400	1.099.760	2.176.500
36	72	106	2.976	1.134.960	2.221.000
37	73	52	2.796	983.080	2.373.240
38	74	113	2.376	1.338.688	2.410.200
39	75	85	3.900	1.458.072	2.675.000

LAMPIRAN 3

Daftar 29 Perusahaan KIP 17115 yang memberikan data Lengkap Setelah Seleksi *Outlier*

No	Perusahaan Ke-	Tenaga Kerja	Bahan Bakar	Bahan Baku	Produksi
1	37	19	468	71.604	196.560
2	38	20	780	111.644	265.200
3	39	22	804	119.810	273.360
4	40	23	780	119.210	276.250
5	41	24	650	118.320	287.300
6	42	24	840	125.488	289.800
7	43	36	900	122.200	309.000
8	44	34	936	137.832	331.500
9	46	32	1.248	200.824	424.320
10	47	40	1.248	151.216	457.600
11	48	48	756	246.284	475.200
12	49	26	910	225.472	484.380
13	50	37	1.380	204.688	488.750
14	51	25	624	308.880	538.200
15	53	32	936	323.752	635.700
16	54	35	936	401.336	696.800
17	55	60	1.404	392.730	777.660
18	56	41	1.092	547.196	982.800
19	57	50	1.092	570.128	1.001.000
20	60	55	1.872	639.600	1.210.560
21	61	87	1.500	727.332	1.277.100
22	63	69	2.010	719.250	1.342.875
23	64	52	1.920	719.220	1.349.200
24	65	60	1.560	818.688	1.430.000
25	66	75	1.560	808.288	1.465.100
26	67	102	1.800	809.320	1.495.600
27	69	70	2.460	886.632	1.667.500
28	70	103	2.280	898.460	1.680.000
29	74	113	2.376	1.338.688	2.410.200

LAMPIRAN 4

Makro QQ-Test

```
macro
qq x.1-x.p
mconstant i n p t chis
mcolumn d x.1-x.p dd pi q ss tt
mmatrix s sinv ma mb mc md
let n=count(x.1)
cova x.1-x.p s
invert s sinv
do i=1:p
  let x.i=x.i-mean(x.i)
enddo
do i=1:n
  copy x.1-x.p ma;
  use i.
  transpose ma mb
  multiply ma sinv mc
  multiply mc mb md
  copy md tt
  let t=tt(1)
  let d(i)=t
enddo
set pi
  1:n
end
let pi=(pi-0,5)/n
sort d dd
invcdf pi q;
chis p.
plot q*dd
invcdf 0.5 chis;
chis p.
let ss=dd<chis
let t=sum(ss)/n
print t
if t>=0,5
  note distribusi data multinormal
endif
if t<0,5
  note distribusi data bukan multinormal
endif
endmacro
```

LAMPIRAN 5

Output Imputasi Menggunakan Metode Regresi pada KIP 17115

The MI Procedure

Model Information

Data Set	WORK.KIP17115
Method	Monotone
Number of Imputations	5
Seed for random number generator	2009

Monotone Model Specification

	Imputed
Method	Variables
Regression	Produksi

Missing Data Patterns

Group	Tenaga Kerja	Bahan Bakar	Bahan Baku	Produksi	Freq	Percent
1	X	X	X	X	20	78.38
2	X	X	X	.	8	21.62

Missing Data Patterns

-----Group Means-----				
Group	TenagaKerja	BahanBakar	BahanBaku	Produksi
1	48.758621	1280.068966	443589	845501
2	35.875000	1098.750000	280028	.

Regression Models for Monotone Method

Imputed Variable	Effect	Obs-Data
Produksi	Intercept	-0.10357
Produksi	TenagaKerja	0.02588
Produksi	BahanBakar	0.10801
Produksi	BahanBaku	0.84414

Regression Models for Monotone Method

Imputed Variable	Effect	-----Imputation-----				
		1	2	3	4	5
Produksi	Intercept	-0.094763	-0.112025	-0.086043	-0.114444	-0.102610
Produksi	TenagaKerja	0.021304	0.029412	0.014320	0.013770	0.018130
Produksi	BahanBakar	0.113942	0.099467	0.119672	0.085909	0.101820
Produksi	BahanBaku	0.839621	0.854730	0.846432	0.866335	0.869460

Multiple Imputation Variance Information

Variable	-----Variance-----			DF	Relative Increase in Variance	Fraction Missing Information
	Between	Within	Total			
Produksi	9877576	8300345820	8312198911	34.105	0.001428	0.001427

Multiple Imputation Parameter Estimates

Variable	Mean	Std Error	95% Confidence Limits		DF
Produksi	786736	91171	601474.9	971997.7	34.105

Multiple Imputation Parameter Estimates

Variable	Minimum	Maximum	Mu0	t for H0:	
				Mean=Mu0	Pr > t
Produksi	783596	790819	750000	0.40	0.6895

Nilai yang diimputasikan pada setiap imputasi

Nilai Imputasi ke-					Rata-rata
1	2	3	4	5	
251.453,4	319.597,6	334.225,2	271.633,5	335.771,8	302.536
417.196,1	351.837,2	422.933,2	390.809,3	359.361,1	388.427
321.457,6	311.578,2	311.443,3	334.850,6	321.980,5	320.262
535.054,7	486.835,1	550.271,9	494.038,5	493.956,0	512.031
539.294,3	465.987,2	511.244,3	480.622,6	454.808,7	490.391
167.434,4	128.895,7	202.724,5	125.996,5	151.531,7	155.317
1.013.849,0	948.534,5	966.861,6	944.217,3	967.288,5	968.150
1.428.567,0	1.460.280,0	1.441.069,0	1.446.065,0	1.487.084,0	1.452.613

LAMPIRAN 6

Output Imputasi Menggunakan Metode PMM pada KIP 17115

The MI Procedure

Model Information

Data Set	WORK.KIP17115
Method	Monotone
Number of Imputations	5
Seed for random number generator	2009

Monotone Model Specification

Method	Imputed Variables
Regression(PMM)	Produksi

Missing Data Patterns

Group	Tenaga Kerja	Bahan Bakar	Bahan Baku	Produksi	Freq	Percent
1	X	X	X	X	29	78.38
2	X	X	X	.	8	21.62

Missing Data Patterns

Group	-----Group Means-----			
	TenagaKerja	BahanBakar	BahanBaku	Produksi
1	48.758621	1280.068966	443589	845501
2	35.875000	1098.750000	280028	.

Regression Models for Monotone Predicted Mean Matching Method

Imputed Variable	Effect	Obs Data
Produksi	Intercept	-0.10357
Produksi	TenagaKerja	0.02588
Produksi	BahanBakar	0.10801
Produksi	BahanBaku	0.84414

Regression Models for Monctone Predicted Mean Matching Method

Imputed Variable	Effect	-----Imputation-----				
		1	2	3	4	5
Produksi	Intercept	-0.094763	-0.096688	-0.108278	-0.098677	-0.112275
Produksi	TenagaKerja	0.021304	0.041038	0.043001	0.038388	0.007553
Produksi	BahanBakar	0.113942	0.126861	0.082288	0.096448	0.101372
Produksi	BahanBaku	0.839621	0.812684	0.854794	0.850251	0.873147

Multiple Imputation Variance Information

Variable	-----Variance-----			DF	Relative Increase in Variance	Fraction Missing Information
	Between	Within	Total			
Produksi	4660022	8282457604	8288049630	34.131	0.000675	0.000675

Multiple Imputation Parameter Estimates

Variable	Mean	Std Error	95% Confidence Limits		DF
Produksi	786894	91039	601907.4	971881.2	34.131

Multiple Imputation Parameter Estimates

Variable	Minimum	Maximum	Mu0	Mean=Mu0	t for H0: Pr > t
Produksi	764775	789747	0	8.64	<.0001

Nilai yang diimputasikan pada setiap imputasi

Nilai Imputasi ke-					Rata-rata
1	2	3	4	5	
331.500	309.000	309.000	309.000	309.000	313.500
424.320	424.320	331.500	331.500	331.500	368.628
309.000	289.800	289.800	289.800	289.800	293.640
538.200	538.200	488.750	488.750	488.750	508.530
488.750	488.750	484.380	488.750	488.750	487.876
196.560	196.560	196.560	196.560	196.560	196.560
982.800	982.800	982.800	982.800	982.800	982.800
1.430.000	1430.000	1.465.100	1465.100	1.430.000	1.444.040

LAMPIRAN 7

Output Regresi Dari 29 Perusahaan KIP 17115 Dengan Data Lengkap

Regression Analysis: Produksi versus Tenaga Kerja, Bahan Bakar, ...

The regression equation is

$$\text{Produksi} = -7722 + 569 \text{ Tenaga Kerja} + 123 \text{ Bahan Bakar} + 1.51 \text{ Bahan Baku}$$

Predictor	Coef	SE Coef	T	P
Constant	-7722	13852	-0.56	0.582
Tenaga Kerja	569.4	425.6	1.34	0.193
Bahan Bakar	122.52	17.84	6.87	0.000
Bahan Baku	1.50730	0.03716	40.56	0.000

S = 24910.5 R-Sq = 99.8% R-Sq(adj) = 99.8%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	9.28620E+12	3.09540E+12	4988.27	0.000
Residual Error	25	15513383739	620535357		
Total	28	9.30171E+12			

Source	DF	Seq SS
Tenaga Kerja	1	7.73273E+12
Bahan Bakar	1	5.32474E+11
Bahan Baku	1	1.02100E+12

Unusual Observations

Obs	Tenaga Kerja	Produksi	Fit	SE Fit	Residual	St Resid
10	40	457600	395889	9770	61711	2.69R
21	87	1277100	1321908	11004	-44808	-2.00R
29	113	2410200	2365540	14010	44660	2.17R

R denotes an observation with a large standardized residual.

LAMPIRAN 8

Output Regresi Dari 54 Perusahaan KIP 19201 Dengan Data Lengkap

N=54

MODEL AWAL

Variables Entered/Removed^b

Model	Variables Entered	Variables Removed	Method
1	Listrik, BahanBaku, BahanBakar, TenagaKerja ^a		Enter

a. All requested variables entered.

b. Dependent Variable: Produksi

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.999 ^a	.999	.999	402487.555	1.741

a. Predictors: (Constant), Listrik, BahanBaku, BahanBakar, TenagaKerja

b. Dependent Variable: Produksi

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6.853E15	4	1.713E15	1.058E4	.000 ^a
	Residual	7.938E12	49	1.620E11		
	Total	6.861E15	53			

a. Predictors: (Constant), Listrik, BahanBaku, BahanBakar, TenagaKerja

b. Dependent Variable: Produksi

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
1 (Constant)	-199670.824	75151.827		-2.657	.011		
TenagaKerja	22417.149	1258.121	.155	17.818	.000	.313	3.197
BahanBaku	1.154	.011	.869	102.582	.000	.329	3.037
BahanBakar	.064	7.427	.000	.009	.993	.872	1.147
Listrik	-.275	.819	-.002	-.336	.738	.998	1.002

a. Dependent Variable:

Produksi

Karena variabel bahan bakar dan listrik tidak signifikan maka kedua variabel tersebut dikeluarkan dari model.

MODEL KEDUA

Variables Entered/Removed^b

Model	Variables Entered	Variables Removed	Method
1	BahanBaku, TenagaKerja ^a		Enter

a. All requested variables entered.

b. Dependent Variable: Produksi

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.999 ^a	.999	.999	394973.067	1.749

a. Predictors: (Constant), BahanBaku, TenagaKerja

b. Dependent Variable: Produksi

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6.853E15	2	3.426E15	2.196E4	.000 ^a
	Residual	7.956E12	51	1.560E11		
	Total	6.861E15	53			

a. Predictors: (Constant), BahanBaku, TenagaKerja

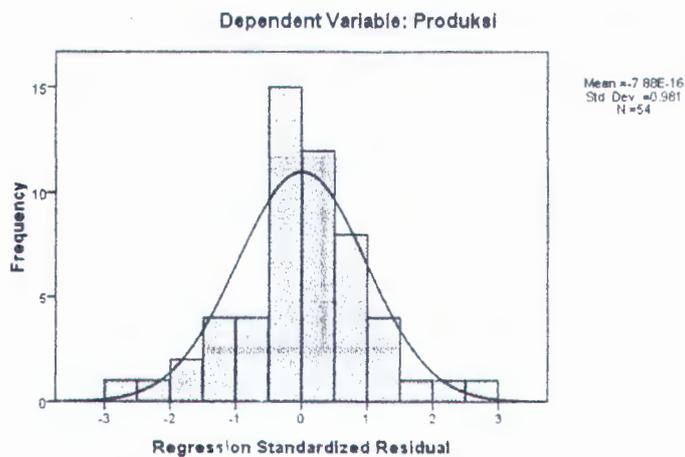
b. Dependent Variable: Produksi

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-203494.147	67317.923		-3.023	.004		
	TenagaKerja	22433.228	1203.067	.155	18.647	.000	.329	3.036
	BahanBaku	1.154	.011	.869	104.546	.000	.329	3.036

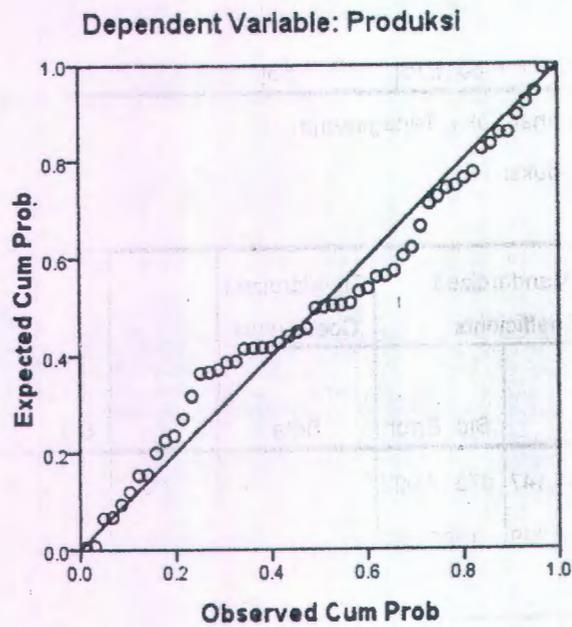
a. Dependent Variable:
Produksi

Uji Normalitas

Histogram



Normal P-P Plot of Regression Standardized Residual



Uji Kolmogorov-Smirnov

One-Sample Kolmogorov-Smirnov Test

		Unstandardized Residual
N		54
Normal Parameters ^a	Mean	.0000000
	Std. Deviation	3.87449081E5
Most Extreme Differences	Absolute	.121
	Positive	.092
	Negative	-.121
Kolmogorov-Smirnov Z		.888
Asymp. Sig. (2-tailed)		.409

a. Test distribution is Normal.

LAMPIRAN 9

Output Regresi dan Uji Asumsi Standar terhadap 29 Perusahaan KIP 17115
yang Memberikan Data Lengkap

N=29

Variables Entered/Removed^b

Model	Variables Entered	Variables Removed	Method
1	BahanBaku, BahanBakar, TenagaKerja ^a		Enter

a. All requested variables entered.

b. Dependent Variable: Produksi

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.999 ^a	.998	.998	24910.547

a. Predictors: (Constant), BahanBaku, BahanBakar, TenagaKerja

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	9.286E12	3	3.095E12	4.988E3	.000 ^a
	Residual	1.551E10	25	6.205E8		
	Total	9.302E12	28			

a. Predictors: (Constant), BahanBaku, BahanBakar, TenagaKerja

b. Dependent Variable: Produksi

1 Uji Multikolinieritas

(untuk menguji apakah ada korelasi antar variabel bebas)

Coefficient Correlations^a

Model			BahanBaku	BahanBakar	TenagaKerja
1	Correlations	BahanBaku	1.000	-.469	-.632
		BahanBakar	-.469	1.000	-.283
		TenagaKerja	-.632	-.283	1.000
Covariances	BahanBaku	.001	-.311	-9.993	
	BahanBakar	-.311	318.094	-2147.608	
	TenagaKerja	-9.993	-2147.608	181127.611	

a. Dependent Variable: Produksi

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-7722.343	13851.805		-.557	.502		
	TenagaKerja	569.409	425.591	.026	1.338	.193	.171	5.846
	BahanBakar	122.521	17.835	.119	6.870	.000	.222	4.500
	BahanBaku	1.507	.037	.870	40.563	.000	.145	6.891

a. Dependent Variable: Produksi

Melihat hasil besaran korelasi antara variabel independen nampak bahwa korelasi antara tenaga kerja dan bahan baku cukup besar yaitu -0,632 atau sekitar 63% namun karena masih dibawah 95% maka bisa dikatakan tidak terjadi multikolinieritas yang serius.

Nilai VIF juga tidak melebihi 10 yang artinya tidak terjadi multikolinieritas yang serius.

2 Uji autokorelasi

(untuk menguji apakah dalam model ada korelasi antara kesalahan pengganggu pada periode t dengan t-1 (periode sebelumnya)).

Dengan Uji Durbin-Watson

H0 : tidak ada autokorelasi ($r=0$)

H1 : ada autokorelasi ($r\neq 0$)

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.999 ^a	.998	.998	24910.547	2.333

a. Predictors: (Constant), BahanBaku, BahanBakar, TenagaKerja

b. Dependent Variable: Produksi

Nilai tabel D-W dengan $n=29$, $k=3$ adalah $dl=0,988$ dan $dv=1,418$. Jarena nilai D-W 2,333 diluar dari rentang (daerah penolakan) maka keputusannya adalah terima Ho yang artinya tidak terjadi autokorelasi.

3. Uji Heterokedastisitas

(untuk menguji apakah terjadi ketidaksamaan varian dari residual satu pengamatan ke pengamatan yang lain)

Uji Park

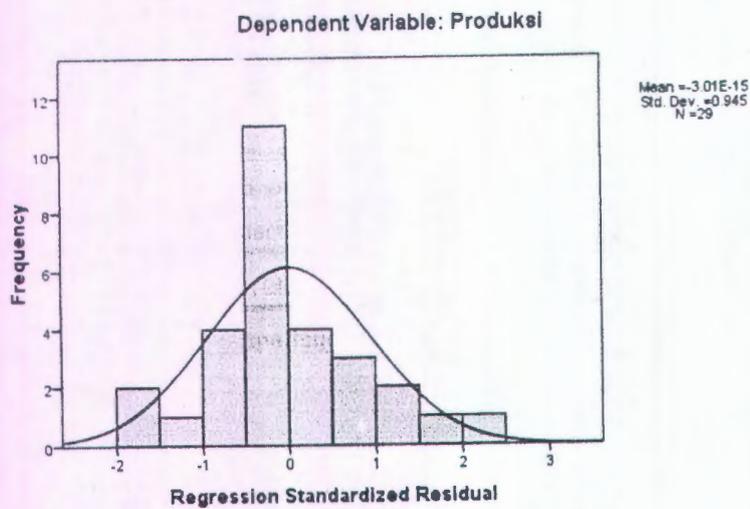
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	19.815	1.127		17.582	.000
	TenagaKerja	-.002	.035	-.022	-.049	.961
	BahanBakar	-.002	.001	-.676	-1.710	.100
	BahanBaku	4.649E-6	.000	.752	1.538	.137

a. Dependent Variable: LnU2i

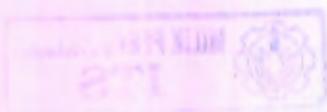
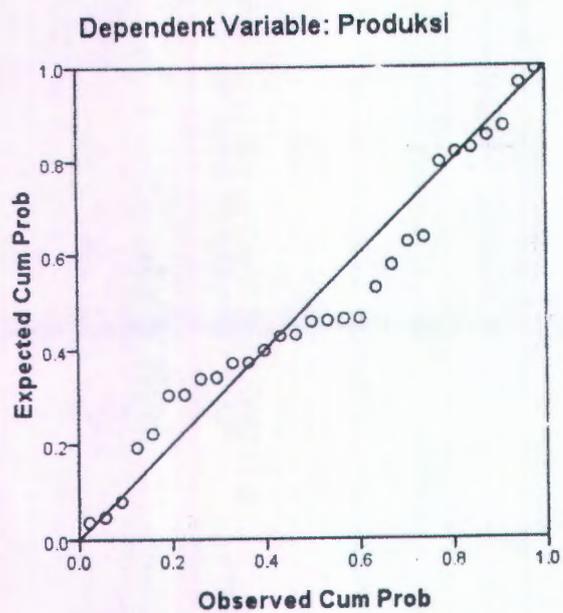
Dari output ternyata kedua variabel independent tidak ada yang signifikan artinya asumsi homokedastisitas terpenuhi.

4 Uji Normalitas

Histogram



Normal P-P Plot of Regression Standardized Residual



Uji Kolmogorov-Smirnov

One-Sample Kolmogorov-Smirnov Test

		Unstandardized Residual
N		29
Normal Parameters ^a	Mean	.0000000
	Std. Deviation	2.35382544E4
Most Extreme Differences	Absolute	.158
	Positive	.158
	Negative	-.119
Kolmogorov-Smirnov Z		.851
Asymp. Sig. (2-tailed)		.464

a. Test distribution is Normal.

Kesimpulan : NORMAL (signifikansi $0,409 > 0,05$)

5. Uji Linearitas

(untuk menguji apakah model linier sudah tepat)

Uji Lagrange Multiplier

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.228 ^a	.052	-.062	24251.55647641

a. Predictors: (Constant), BBaku2, BBakar2, TK2

b. Dependent Variable: Unstandardized Residual

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	8.099E8	3	2.700E8	.459	.713 ^a
	Residual	1.470E10	25	5.881E8		
	Total	1.551E10	28			

a. Predictors: (Constant), BBaku2, BBakar2, TK2

b. Dependent Variable: Unstandardized Residual

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1522.235	7440.899		.205	.840
	TK2	-1.711	2.965	-.244	-.577	.569
	BBakar2	-.003	.005	-.227	-.606	.550
	BBaku2	3.286E-8	.000	.552	1.126	.271

a. Dependent Variable: Unstandardized Residual

H0 : Model Linear

H1 : Model tidak linear

Hasil output menunjukkan bahwa nilai R^2 sebesar 0,052 dengan jumlah n observasi 29, besarnya nilai c^2 hitung = $29 \times 0,052 = 1,508$. Nilai ini dibandingkan dengan nilai c^2 tabel dengan df 25 dan tingkat signifikansi 0,05 sebesar 37,65. Karena nilai c^2 hitung lebih kecil dari c^2 tabel maka keputusannya adalah terima H0 dan dapat disimpulkan bahwa model adalah benar model linier.

LAMPIRAN 10

Daftar Perusahaan KIP 19201 yang Memberikan Data Lengkap

No	Prshn. Ke-	Tenaga Kerja	Bahan Baku	Produksi	No	Prshn. Ke-	Tenaga Kerja	Bahan Baku	Produksi
1	1	15	1.003.750	664.490	28	41	28	3.670.380	2.493.960
2	2	16	390.000	98.270	29	42	29	2.974.750	2.157.345
3	4	17	635.000	393.280	30	44	30	441.000	324.000
4	5	17	894.000	475.560	31	47	30	2.821.500	2.153.883
5	6	17	1.423.254	534.740	32	49	31	2.576.000	1.440.661
6	8	18	1.013.775	514.296	33	51	32	1.209.600	810.275
7	12	19	177.408	33.552	34	53	34	481.000	189.447
8	16	20	575.000	285.000	35	54	35	1.277.400	343.904
9	17	20	691.200	384.376	36	55	35	3.150.000	1.336.150
10	18	20	720.000	347.916	37	56	36	848.680	558.678
11	19	20	823.680	473.549	38	58	40	1.200.000	386.000
12	20	20	1.260.000	452.175	39	59	40	2.800.000	2.697.419
13	21	21	1.150.000	559.244	40	60	41	812.577	536.048
14	22	21	1.420.571	1.061.450	41	61	42	720.000	55.748
15	23	21	1.638.000	699.134	42	62	45	1.854.000	1.310.222
16	24	21	1.950.000	1.093.300	43	63	48	1.764.000	870.090
17	25	22	244.288	31.821	44	64	49	4.662.350	3.374.725
18	26	22	3.370.250	2.338.277	45	65	50	674.200	303.542
19	27	23	661.375	295.500	46	66	52	552.000	164.124
20	29	23	1.150.000	822.100	47	67	55	1.926.800	850.371
21	33	25	1.012.100	604.340	48	68	62	1.699.880	389.700
22	34	25	1.112.865	941.721	49	69	70	1.305.000	771.965
23	35	25	1.342.000	1.040.968	50	70	70	5.181.547	3.306.484
24	36	25	1.950.000	1.423.500	51	71	89	2.682.035	931.801
25	38	27	607.426	425.388	52	72	93	5.085.583	1.954.160
26	39	28	1.159.515	408.237	53	76	411	21.600.000	10.678.280
27	40	28	1.251.533	835.150	54	77	458	82.683.062	62.951.900

LAMPIRAN 11

Output Imputasi Menggunakan Metode Regresi pada KIP 19201

The MI Procedure

Model Information

Data Set	WORK.KIP19201
Method	Monotone
Number of Imputations	5
Seed for random number generator	12304

Monotone Model Specification

	Imputed
Method	Variables
Regression	Produksi

Missing Data Patterns

Group	Tenaga Kerja	Bahan Baku	Produksi	Freq	Percent	-----Group Means-----		
						TenagaKerja	BahanBaku	Produksi
1	X	X	X	54	84.38	48.351852	3458895	2232941
2	X	X	.	10	15.63	24.600000	494854	.

Regression Models for Monotone Method

Imputed Variable	Effect	Obs-Data
Produksi	Intercept	-0.03844
Produksi	TenagaKerja	-0.16083
Produksi	BahanBaku	1.05678

Regression Models for Monotone Method

Imputed Variable	Effect	-----Imputation-----				
		1	2	3	4	5
Produks	Intercept	-0.030957	-0.034817	-0.037645	-0.049291	-0.040784
Produksi	TenagaKerja	-0.155125	-0.144012	-0.169886	-0.150607	-0.144616
Produksi	BahanBaku	1.056089	1.042866	1.059752	1.052169	1.041333

The M1 Procedure

Multiple Imputation Variance Information

Variable	-----Variance-----			DF	Relative Increase in Variance	Fraction Missing Information
	Between	Within	Total			
Produksi	75449087	972521789175	972612328080	61.085	0.000093097	0.000093093

Multiple Imputation Parameter Estimates

Variable	Mean	Std Error	95% Confidence Limits		DF
Produksi	1913276	986211	-58719.4	3885271	61.085

Multiple Imputation Parameter Estimates

Variable	Minimum	Maximum	Mu0	t for H0:	
				Mean=Mu0	Pr > t
Produksi	1901548	1922820	0	.94	0.0570

LAMPIRAN 12

Output Imputasi Menggunakan Metode PMM pada KIP 19201

The MI Procedure

Model Information

Data Set	WORK.KIP19201
Method	Monotone
Number of Imputations	5
Seed for random number generator	12304

Monotone Model Specification

Method	Imputed Variables
Regression(PMM)	Produksi

Missing Data Patterns

Group	Tenaga Kerja	Bahan Baku	Produksi	Freq	Percent	-----Group Means-----		
						TenagaKerja	BahanBaku	Produksi
1	X	X	X	54	84.38	48.351852	3458895	2232941
2	X	X	.	10	15.63	24.600000	494854	.

Regression Models for Monotone Predicted Mean Matching Method

Imputed Variable	Effect	Obs Data
Produksi	Intercept	-0.03844
Produksi	TenagaKerja	-0.16083
Produksi	BahanBaku	1.05678

Regression Models for Monotone Predicted Mean Matching Method

Imputed Variable	Effect	-----Imputation-----				
		1	2	3	4	5
Produksi	Intercept	-0.030957	-0.044142	-0.030219	-0.038436	-0.036939
Produksi	TenagaKerja	-0.155125	-0.167484	-0.166063	-0.152132	-0.157532
Produksi	BahanBaku	1.056089	1.068716	1.067404	1.060134	1.057835

The MI Procedure

Multiple Imputation Variance Information

Variable	-----Variance-----			DF	Relative Increase in Variance	Fraction Missing Information
	Between	Within	Total			
Produksi	24063598	972097832390	972126708707	61.089	0.000029705	0.000029705

Multiple Imputation Parameter Estimates

Variable	Mean	Std Error	95% Confidence Limits		DF
Produksi	1914521	985965	-56979.3	3886022	61.089

Multiple Imputation Parameter Estimates

Variable	Minimum	Maximum	Mu0	t for H0:	
				Mean=Mu0	Pr > t
Produksi	1909278	1920349	0	1.94	0.0568

LAMPIRAN 13

Makro Imputasi Berganda Metode Regresi menggunakan SAS 9,0 (Contoh untuk KIP 17115)

```
data KIP17115;
  input TenagaKerja BahanBakar BahanBaku Produksi @@;
  datalines;
    20    1000    123550      .
    21    1200    150870      .
    24     720    133430      .
    30    1200    233256      .
    30    1230    223550      .
    32     936     22360      .
    40    1100    539755      .
    90    1404    813456      .
    19     468     71604    196560
    20     780    111644    265200
    22     804    119810    273360
    23     780    119210    276250
    24     650    118320    287300
    24     840    125488    289800
    36     900    122200    309000
    34     936    137832    331500
    :       :       :       :
    113   2376   1338688   2410200
  ;

proc mi data=KIP17115
  seed=2009 out=outex3 nimpute=5;
  monotone reg (Produksi/details);
  var TenagaKerja BahanBakar BahanBaku Produksi;
run;

proc print data=outex3(obs=185);
  title '185 Observations of the Imputed Data Set';
run;

proc export data=outex3
  outfile='d:\thesis\fokus\KIP17115REG.xls'
  dbms=excel
  replace;
run;
```

LAMPIRAN 14

Makro Imputasi Berganda Metode PMM menggunakan SAS 9,0 (Contoh untuk KIP 17115)

```
data KIP17115;
  input TenagaKerja BahanBakar BahanBaku Produksi @@;
  datalines;
    20    1000    123550      .
    21    1200    150870      .
    24     720    133430      .
    30    1200    233256      .
    30    1230    223550      .
    32     936     22360      .
    40    1100    539755      .
    90    1404    813456      .
    19     468     71604    196560
    20     780    111644    265200
    22     804    119810    273360
    23     780    119210    276250
    24     650    118320    287300
    24     840    125488    289800
    36     900    122200    309000
    34     936    137832    331500
    :      :      :      :
    113   2376  1338688  2410200
  ;
proc mi data=KIP17115
  seed=2009 out=outex3 nimpute=5;
  monotone regpmm (Produksi/details);
  var TenagaKerja BahanBakar BahanBaku Produksi;
run;

proc print data=outex3(obs=185);
  title '185 Observations of the Imputed Data Set';
run;

proc export data=outex3
  outfile='d:\thesis\fokus\KIP17115PMM.xls'
  dbms=excel
  replace;
run;
```

LAMPIRAN 15

Hasil Penghitungan *Mean Square Error* Simulasi (MSE_s) dari ke-5 Skenario Skenario-1

Data Awal

No	x_1	x_2	y	No	x_1	x_2	y
1	28	199.807	200.102	41	38	200.279	200.195
2	27	199.968	200.357	42	29	199.912	199.807
3	29	199.792	200.036	43	21	200.305	200.622
4	32	200.068	199.549	44	26	199.240	199.478
5	38	199.726	199.507	45	19	199.950	200.467
6	42	200.272	200.157	46	42	200.292	200.440
7	31	200.591	200.471	47	20	200.677	200.899
8	31	200.025	200.624	43	36	199.857	199.701
9	32	199.908	200.282	49	24	199.758	199.749
10	24	199.822	199.487	50	23	199.946	200.458
11	27	199.449	198.709	51	25	200.158	200.279
12	27	200.106	200.054	52	33	199.774	200.279
13	24	199.467	199.453	53	33	200.573	201.091
14	31	200.344	200.823	54	42	200.123	200.375
15	29	200.028	200.363	55	30	199.872	199.932
16	25	200.476	200.629	56	34	199.414	199.052
17	27	200.098	200.349	57	28	199.525	199.885
18	32	200.143	200.733	58	30	200.610	200.833
19	25	199.911	200.091	59	29	199.933	199.692
20	32	199.852	199.711	60	29	199.790	200.253
21	34	199.295	199.611	61	31	199.857	200.179
22	38	200.074	200.147	62	26	199.843	199.940
23	32	200.371	200.631	63	32	199.618	199.936
24	28	200.281	200.241	64	40	200.122	200.434
25	31	200.244	200.189	65	31	199.895	199.587
26	26	199.774	199.543	66	39	200.002	200.072
27	30	200.432	200.652	67	32	200.158	199.915
28	37	200.053	200.148	68	31	199.703	199.680
29	28	199.516	199.563	69	33	200.106	199.788
30	25	200.763	200.693	70	31	199.957	199.981
31	27	199.964	199.471	71	27	200.251	200.390
32	27	200.000	199.720	72	26	199.719	199.251
33	33	199.803	200.023	73	35	199.206	199.309
34	28	199.783	199.703	74	28	200.188	200.230
35	22	199.813	199.542	75	28	199.855	199.449
36	33	199.896	199.821	76	27	199.948	200.118
37	32	200.226	200.364	77	36	199.652	199.228
38	28	199.914	199.749	78	29	200.047	200.319
39	44	200.210	199.775	79	29	199.941	199.594
40	36	199.798	199.427	80	26	200.158	200.391

Nilai y yang dimissingkan, hasil imputasi dengan kedua metode dan MSE_s

y dimissingkan 40 buah (50 %)					y dimissingkan 30 buah (37,5 %)				
y _{miss}	Regresi		PMM		y _{miss}	Regresi		PMM	
	y _{imp}	error ²	y _{imp}	error ²		y _{imp}	error ²	y _{imp}	error ²
199.453	199.450	9	199.469	262	200.102	199.641	212.090	199.845	66.152
200.629	200.821	36.747	200.728	9.841	199.549	200.047	248.081	200.155	367.721
200.349	200.289	3.551	200.225	15.376	199.507	199.895	150.552	199.640	17.636
200.091	199.717	140.161	199.992	9.801	200.282	199.940	116.632	199.933	121.661
199.611	198.989	387.077	199.228	146.689	200.054	200.085	933	200.240	34.670
200.147	200.101	2.091	200.039	11.578	199.453	199.509	3.172	199.492	1.537
200.241	200.485	59.407	200.424	33.416	200.823	200.385	191.832	200.449	140.026
199.543	199.802	66.919	199.788	60.025	200.363	199.945	175.017	200.134	52.258
200.652	200.728	5.763	200.585	4.543	200.349	200.336	157	200.235	12.996
200.148	200.022	15.852	200.056	12.589	200.091	200.114	510	200.025	4.330
199.563	199.446	13.786	199.489	5.476	200.241	200.338	9.495	200.417	31.117
200.693	201.021	107.448	201.014	103.169	200.652	200.483	28.515	200.615	1.354
199.720	200.077	127.684	200.092	138.682	200.693	201.063	136.614	201.014	103.169
200.023	199.999	555	199.770	64.110	200.023	199.907	13.375	199.796	51.529
199.703	199.916	45.359	199.785	6.790	200.364	200.171	37.286	200.309	2.981
199.542	199.969	182.404	199.928	149.150	199.749	199.964	46.101	199.984	55.131
199.821	199.952	17.173	199.915	8.836	199.807	199.826	356	199.967	25.728
200.195	200.275	6.412	200.311	13.549	200.622	200.201	177.266	200.579	1.815
199.807	199.706	10.186	199.954	21.609	200.440	200.459	370	200.286	23.654
200.622	200.655	1.089	200.535	7.569	199.701	199.904	41.181	199.805	10.858
200.467	200.188	77.636	200.121	119.854	200.375	200.257	13.889	200.069	93.391
200.279	200.228	2.562	200.317	1.429	199.885	199.069	665.211	199.512	139.129
199.052	199.239	35.094	199.308	65.331	199.692	200.152	212.058	199.992	90.000
199.885	199.667	47.709	199.489	156.816	200.253	199.786	217.688	199.827	181.817
200.833	200.782	2.590	200.817	243	199.587	199.991	163.304	199.932	118.749
200.253	199.570	467.089	199.785	218.650	200.072	199.964	11.716	199.951	14.738
200.179	200.061	13.929	199.864	99.099	200.390	200.221	28.515	200.402	144
199.940	199.741	39.542	199.919	433	199.251	199.936	468.983	199.781	280.900
200.434	200.074	129.648	200.078	126.878	199.309	199.107	40.669	199.087	49.195
199.587	199.813	50.851	199.922	112.091	200.118	200.000	13.815	200.047	5.013
200.072	199.922	22.563	199.933	19.432		Total	3.425.382	Total	2.099.400
199.680	199.881	40.527	199.700	384		MSE	114.179	MSE	69.980
200.390	200.894	253.902	200.411	433					
199.251	199.730	229.037	199.749	248.403					
199.309	198.901	166.478	199.020	83.290					
200.230	200.284	2.875	200.312	6.724					
200.118	200.000	13.840	200.029	7.992					
200.319	200.117	40.981	200.124	38.025					
199.594	199.943	122.009	199.992	158.404					
200.391	200.277	13.084	200.310	6.529					
	Total	3.001.620	Total	2.293.500					
	MSE _s	75.041	MSE _s	57.338					

Nilai y yang dimissingkan, hasil imputasi dengan kedua metode dan MSE_s

y dimissingkan 20 buah (25 %)					y dimissingkan 10 buah (12,5 %)				
y_{miss}	Regresi		PMM		y_{miss}	Regresi		PMM	
	y_{imp}	error ²	y_{imp}	error ²		y_{imp}	error ²	y_{imp}	error ²
200.35	199.93	179.577	200.08	73.333	200.375	200.00	126.95	200.03	104.976
200.03	199.73	93.419	199.84	37.172	199.885	199.92	129.79	199.90	144.400
199.48	199.79	96.477	199.95	221.841	199.692	200.43	154.79	200.42	155.236
198.70	199.42	506.001	199.51	651.572	200.253	199.83	65.857	199.97	13.549
200.34	200.34	27	200.22	14.787	199.587	200.29	10.337	200.32	17.161
200.09	199.97	13.542	200.03	3.295	200.072	199.61	2.666	199.48	5.595
200.14	200.00	19.834	200.08	4.462	200.390	199.77	249.90	199.74	284.089
200.18	200.29	12.186	200.34	23.470	199.251	199.84	285.15	200.07	92.659
200.62	200.46	25.920	200.49	16.180	199.309	199.77	232.31	199.80	199.809
200.89	200.90	6	200.98	7.534	200.118	199.95	27.579	200.00	12.188
200.27	199.68	355.231	199.80	229.633		Total	1.285.3	Total	1.029.6
199.05	199.23	32.818	199.43	148.379		MSE_s	128.53	MSE_s	102.966
199.88	199.50	147.224	199.57	98.596					
200.43	200.09	118.238	200.11	102.400					
199.58	200.04	207.115	199.96	140.925					
199.91	200.19	76.980	200.24	107.322					
200.39	200.35	1.248	200.38	9					
199.30	198.99	98.169	199.25	3.364					
199.44	199.80	125.571	199.95	254.419					
199.22	199.80	331.407	199.63	166.464					
	Total	2.440.989	Total	2.305.15					
	MSE_s	122.049	MSE_s	115.258					

Skenario-2

Data Awal

No	x ₁	x ₂	x ₃	y	No	x ₁	x ₂	x ₃	y
1	29	25.121	250.075	275.086	41	34	24.934	249.854	274.664
2	34	25.159	250.144	275.022	42	31	25.063	249.736	274.329
3	33	25.044	250.120	274.795	43	27	25.073	249.614	275.306
4	29	25.040	249.835	275.159	44	34	24.909	249.983	274.901
5	27	24.989	250.221	275.191	45	27	25.069	249.974	275.013
6	25	24.985	250.706	275.866	46	24	24.950	250.018	274.978
7	29	24.948	250.442	275.441	47	28	25.291	250.056	274.881
8	30	24.943	250.434	275.470	48	34	24.717	250.227	275.300
9	28	24.901	250.338	275.239	49	31	24.918	249.992	275.072
10	31	25.093	249.482	274.741	50	34	25.254	250.094	275.258
11	32	24.918	249.620	274.466	51	21	25.011	250.121	275.446
12	26	24.944	249.611	274.704	52	39	25.108	249.787	274.876
13	30	24.919	249.656	274.410	53	34	25.094	249.337	274.416
14	24	25.021	250.425	275.735	54	26	24.985	249.769	275.340
15	36	25.017	250.073	275.478	55	35	25.026	250.076	275.549
16	30	25.093	249.772	275.247	56	22	25.123	250.277	275.139
17	27	25.027	249.540	274.501	57	28	24.938	249.931	274.976
18	28	24.928	249.915	274.565	58	33	24.936	249.644	275.068
19	22	25.064	250.326	275.954	59	29	24.937	249.937	274.641
20	28	24.902	249.556	274.653	60	32	24.784	249.847	274.344
21	29	25.057	250.119	275.428	61	32	25.345	249.765	275.432
22	29	25.186	249.657	275.351	62	34	24.912	250.006	275.422
23	27	24.927	250.273	274.434	63	29	25.019	250.206	275.325
24	24	24.984	249.793	274.878	64	26	25.070	250.244	275.418
25	32	24.945	250.108	275.337	65	32	24.756	250.061	274.978
26	22	25.110	250.074	275.066	66	28	24.705	249.488	273.995
27	33	25.135	249.777	274.705	67	27	24.846	249.843	275.271
28	22	25.235	250.166	275.248	68	35	25.216	249.933	275.556
29	28	25.004	249.615	275.020	69	29	24.855	250.084	275.119
30	35	24.776	250.288	275.229	70	38	25.075	250.066	275.186
31	32	24.824	250.181	275.072	71	28	25.159	249.663	275.079
32	29	24.888	250.550	275.278	72	27	24.956	250.403	274.846
33	25	25.144	250.243	276.065	73	22	25.049	250.240	275.290
34	24	24.912	250.414	275.804	74	32	25.135	250.075	275.499
35	28	25.039	250.088	274.980	75	21	24.975	249.884	274.677
36	27	24.867	250.217	275.383	76	34	25.167	250.133	275.019
37	31	25.103	250.144	275.153	77	28	24.870	250.239	275.179
38	35	24.994	249.934	274.600	78	28	25.005	249.616	274.463
39	28	25.038	249.733	274.795	79	34	24.861	249.644	274.528
40	34	25.023	250.321	275.119	80	33	24.783	249.837	274.637

Nilai y yang dimissingkan, hasil imputasi dengan kedua metode dan MSE_s

y dimissingkan 40 buah (50 %)					y dimissingkan 30 buah (37,5 %)				
y_{miss}	Regresi		PMM		y_{miss}	Regresi		PMM	
	y_{imp}	error ²	y_{imp}	error ²		y_{imp}	error ²	y_{imp}	error ²
274.410	274.635	50.674	274.686	75.955	274.410	274.629	48.091	274.715	93.269
275.247	274.943	92.452	274.790	209.215	275.735	275.409	106.107	275.480	65.025
274.501	274.641	19.545	274.572	5.041	275.478	275.265	45.435	275.193	81.111
275.954	275.039	837.317	275.266	473.619	274.501	274.365	18.393	274.666	27.159
275.428	275.129	89.214	275.103	105.495	275.954	275.376	334.086	275.404	302.940
275.351	274.588	581.493	274.663	473.619	274.878	274.695	33.581	274.839	1.537
274.878	274.695	33.586	274.782	9.178	274.705	274.812	11.420	274.966	68.121
275.066	275.245	31.896	275.025	1.714	275.729	275.264	1.225	275.231	6
274.705	274.954	62.146	274.787	6.757	276.065	275.288	603.797	275.389	456.706
275.248	275.344	9.251	275.106	20.221	275.153	275.433	78.393	275.294	19.768
275.020	274.982	1.428	274.634	148.842	274.500	275.399	637.857	275.041	194.834
275.229	274.938	84.840	275.323	8.761	274.329	274.694	133.442	274.883	306.916
275.278	275.103	30.649	275.506	51.893	275.306	274.416	792.601	274.763	295.284
276.065	275.426	408.522	275.202	745.460	274.978	275.232	64.548	275.042	4.122
275.804	275.337	217.950	275.357	200.167	275.300	275.176	15.336	275.129	29.241
274.980	275.078	9.691	275.078	9.526	275.340	274.949	152.502	274.844	246.413
275.383	275.250	17.667	275.200	33.562	274.976	275.137	25.971	274.964	139
274.664	275.044	144.339	274.878	45.710	274.641	275.068	181.962	274.987	119.439
274.329	274.863	284.792	274.753	179.946	274.344	274.679	112.418	274.807	214.184
275.306	274.760	297.894	274.627	460.769	274.978	274.926	2.693	274.987	85
275.013	274.953	3.649	274.963	2.540	273.995	274.321	106.261	274.420	180.285
275.446	274.981	216.314	275.087	129.025	275.079	274.914	27.374	274.854	50.805
275.139	275.325	34.578	275.220	6.593	274.846	275.539	480.741	275.429	339.423
274.976	274.950	662	274.947	829	275.290	275.335	2.000	275.302	139
275.068	274.762	93.343	274.687	145.313	275.019	275.320	90.548	275.337	100.997
274.344	274.882	289.218	274.885	293.114	275.248	275.344	9.251	275.106	20.221
275.432	274.775	431.487	274.773	434.017	275.020	274.982	1.428	274.634	148.842
275.422	275.034	150.244	275.046	141.526	275.229	274.938	84.840	275.323	8.761
275.418	275.153	70.338	275.202	46.656	275.278	275.103	30.649	275.506	51.893
274.978	275.064	7.336	275.095	13.783	276.065	275.426	408.522	275.202	745.460
273.995	274.424	184.098	274.554	312.481		Total	4.106.782	Total	3.197.947
275.556	275.070	236.512	274.950	367.236		MSE	164.271	MSE	127.918
275.079	274.736	117.904	274.670	167.445					
274.846	275.298	204.284	275.357	260.712					
275.290	275.060	53.115	275.185	10.983					
275.499	275.032	218.426	275.068	186.106					
275.019	275.135	13.439	275.134	13.133					
274.463	274.687	50.181	274.634	29.309					
274.528	274.693	27.331	274.710	33.270					
274.637	274.747	12.088	274.893	65.638					
	Total	5.719.894	Total	5.925.1					
	MSE	142.997	MSE	148.129					

Nilai y yang dimissingkan, hasil imputasi dengan kedua metode dan MSE_s

y dimissingkan 20 buah (25 %)					y dimissingkan 10 buah (12,5 %)				
y_{miss}	Regresi		PMM		y_{miss}	Regresi		PMM	
	y_{imp}	error ²	y_{imp}	error ²		y_{imp}	error ²	y_{imp}	error ²
275.02	275.68	435.367	275.34	103.298	275.022	275.16	20.428	275.32	93.881
274.79	275.18	153.394	275.20	170.734	275.239	275.28	2.374	275.28	1.971
274.74	274.50	56.241	274.58	24.649	275.735	275.49	60.254	275.45	78.736
274.46	274.52	3.407	274.56	9.722	275.954	275.26	476.03	275.37	336.632
274.50	274.70	41.095	274.58	7.430	275.337	275.11	48.743	275.11	50.535
275.95	275.39	313.307	275.46	242.458	275.020	274.59	178.56	274.66	124.045
275.35	274.78	323.791	274.85	242.458	274.876	274.93	3.778	274.94	5.184
275.33	275.08	62.144	275.10	54.010	275.340	274.90	192.95	274.79	297.679
275.30	274.63	451.533	274.70	359.280	274.344	274.54	40.309	274.72	146.536
274.88	275.44	320.068	275.38	251.001	275.019	275.47	205.23	275.32	95.728
274.87	274.98	11.393	274.91	1.490		Total	1.228.6	Total	1.230.9
275.13	275.35	44.918	275.47	109.826		MSE_s	122.86	MSE_s	123.093
274.97	274.88	7.995	274.89	6.336					
275.41	275.35	4.478	275.38	1.282					
274.97	274.76	44.697	274.85	14.256					
275.27	274.81	204.944	274.74	281.324					
275.07	274.89	33.814	274.84	52.716					
275.29	275.58	86.942	275.35	4.624					
274.67	274.82	21.124	274.91	54.569					
275.17	275.19	356	275.16	104					
	Total	2.621.010	Total	1.991.56					
	MSE_s	131.051	MSE_s	99.578					

Skenario-3

Data Awal

No	x ₁	x ₂	y	No	x ₁	x ₂	y
1	42	1.230.212	1.141.929	41	37	1.483.382	1.502.115
2	23	1.170.048	1.511.791	42	31	1.341.745	1.172.038
3	24	1.244.955	1.039.585	43	45	1.048.488	748.580
4	31	1.104.853	1.328.451	44	24	1.079.564	1.573.067
5	22	967.609	1.564.768	45	36	1.458.315	1.871.673
6	44	988.225	1.322.996	46	29	1.111.083	1.578.553
7	39	978.136	760.229	47	29	1.249.141	1.524.306
8	40	1.162.415	1.481.266	48	43	1.168.771	1.177.805
9	32	1.425.794	1.144.389	49	35	1.455.377	1.950.976
10	22	1.450.491	1.794.485	50	29	1.305.768	1.769.342
11	40	1.393.096	1.459.335	51	28	1.421.747	1.458.801
12	40	1.236.347	1.139.613	52	25	1.043.332	1.203.960
13	41	1.134.799	1.220.800	53	23	1.360.895	1.970.286
14	36	1.337.131	1.683.504	54	33	1.455.177	2.050.180
15	29	1.156.780	1.414.472	55	24	1.397.723	1.939.128
16	29	893.600	1.213.236	56	41	1.316.356	1.567.109
17	35	1.066.163	1.664.174	57	39	994.739	1.170.500
18	23	1.170.186	1.444.769	58	35	1.031.298	1.313.449
19	30	1.199.100	1.431.979	59	36	1.080.813	1.573.430
20	32	1.476.386	1.359.047	60	29	1.475.982	1.586.245
21	41	1.214.254	1.172.209	61	39	1.005.182	1.333.281
22	21	1.426.936	1.824.349	62	26	840.194	858.456
23	37	869.936	1.163.858	63	35	788.158	997.866
24	30	848.055	984.951	64	41	1.168.713	1.351.914
25	23	1.093.027	910.889	65	42	1.435.853	1.472.090
26	44	1.306.898	1.859.933	66	26	1.118.906	1.328.181
27	31	1.474.791	1.322.568	67	25	1.079.756	1.465.709
28	25	1.057.911	1.301.038	68	44	1.415.698	1.796.954
29	40	1.281.975	1.153.666	69	28	905.417	1.456.967
30	29	1.492.053	1.350.523	70	40	1.121.303	1.320.459
31	28	1.023.216	952.369	71	32	975.454	1.374.609
32	43	886.812	1.162.626	72	24	988.599	1.157.922
33	41	1.143.475	1.394.644	73	30	1.125.221	1.554.289
34	24	1.209.082	1.805.990	74	26	755.301	516.447
35	39	1.279.227	1.480.028	75	44	1.230.892	1.220.522
36	33	1.075.465	1.106.851	76	44	1.490.733	2.006.801
37	39	1.163.670	1.229.691	77	24	1.015.581	1.622.649
38	20	1.389.940	1.329.419	78	32	833.535	623.621
39	29	1.410.409	2.069.089	79	41	1.012.586	1.540.324
40	21	1.045.559	1.397.564	80	33	1.278.644	1.079.765

Nilai y yang dimissingkan, hasil imputasi dengan kedua metode dan MSE_s

y dimissingkan 40 buah (50 %)				
y_{miss}	Regresi		PMM	
	y_{imp}	error ²	y_{imp}	error ²
1.220.800	1.559.421	114.664.410.271	1.335.894	13.246.674.874
1.213.236	1.267.718	2.968.287.460	1.208.345	23.923.837
1.664.174	1.004.922	434.613.834.064	1.299.587	132.923.680.569
1.431.979	1.533.782	10.363.809.536	1.444.829	165.117.360
1.172.209	1.288.763	13.584.939.323	1.377.835	42.282.216.377
1.824.349	1.841.532	295.271.514	1.676.330	21.909.742.776
984.951	1.374.031	151.383.619.375	1.175.644	36.363.896.526
1.859.933	1.324.626	286.553.177.665	1.439.950	176.386.056.276
1.322.568	1.422.927	10.072.028.565	1.646.991	105.250.282.929
1.301.038	1.317.815	281.468.212	1.364.192	3.988.452.978
1.153.666	1.394.127	57.821.471.900	1.439.751	81.844.856.093
1.350.523	1.726.916	141.671.820.935	1.688.501	114.228.858.102
1.162.626	1.138.070	603.017.349	1.100.104	3.909.000.484
1.394.644	1.474.324	6.348.869.141	1.343.598	2.605.734.953
1.805.990	1.690.979	13.227.539.692	1.493.622	97.573.642.477
1.480.028	1.438.367	1.735.674.988	1.441.026	1.521.140.403
1.106.851	1.349.075	58.672.390.882	1.335.894	52.460.787.466
1.502.115	1.680.055	31.662.523.027	1.634.820	17.610.617.025
1.172.038	1.532.106	129.648.717.256	1.543.248	137.796.715.616
748.580	987.164	56.922.170.462	1.219.155	221.440.454.165
1.871.673	1.780.823	8.253.637.634	1.617.182	64.765.872.674
1.458.801	1.526.001	4.515.812.019	1.626.001	27.955.773.120
1.567.109	1.341.197	51.036.403.335	1.463.556	10.723.265.230
1.170.500	1.293.995	15.250.920.767	1.234.867	4.143.059.196
1.313.449	1.329.878	269.913.684	1.269.582	1.924.296.142
1.586.245	1.742.311	24.356.570.866	1.676.330	8.115.235.157
1.333.281	1.268.837	4.153.075.999	1.240.098	8.683.146.036
858.456	1.050.709	36.961.115.946	1.175.644	100.608.354.219
1.351.914	1.193.563	25.074.991.660	1.343.598	69.162.509
1.472.090	1.452.592	380.156.178	1.550.316	6.119.244.495
1.328.181	1.430.282	10.424.599.671	1.408.978	6.528.155.209
1.796.954	1.532.106	70.144.607.000	1.519.300	77.091.743.716
1.374.609	1.175.902	39.484.595.258	1.253.872	14.577.423.169
1.157.922	1.438.291	78.606.827.258	1.319.518	26.113.396.493
1.554.289	1.578.121	567.080.430	1.394.980	25.379.421.205
516.447	1.089.891	328.837.620.021	1.100.104	340.655.493.649
2.006.801	1.491.761	265.265.912.645	1.582.421	180.098.554.152
623.621	1.173.855	302.756.931.973	1.151.971	279.153.722.500
1.540.324	1.356.035	33.962.471.878	1.213.924	106.537.221.120
1.079.765	1.481.172	161.127.583.000	1.485.505	164.624.785.304
	Total	2.984.527.068.841	Total	2.717.399.176.581
	MSE_s	74.613.176.721	MSE_s	67.934.979.415

Nilai y yang dimissingkan, hasil imputasi dengan kedua metode dan MSE_s

y dimissingkan 30 buah (37,5 %)				
y_{miss}	Regresi		PMM	
	y_{imp}	error ²	y_{imp}	error ²
1.141.929	1.536.9	156.064.317.967	1.463.290	103.272.892.321
1.328.451	1.515.6	35.035.684.671	1.328.126	105.755
1.564.768	1.255.8	95.411.099.311	1.180.805	147.427.892.540
1.144.389	1.700.0	308.707.154.089	1.681.326	288.301.556.744
1.139.613	1.406.1	71.030.105.496	1.466.344	106.753.146.361
1.220.800	1.528.5	94.731.963.100	1.349.234	16.495.240.982
1.683.504	1.483.2	40.111.234.239	1.591.708	8.426.468.898
1.414.472	1.579.8	27.343.352.641	1.391.598	523.210.726
1.664.174	1.270.4	154.997.327.786	1.260.497	162.955.281.800
1.431.979	1.480.9	2.396.560.506	1.432.939	921.216
984.951	996.381	130.638.611	1.041.460	3.193.267.081
1.322.568	1.877.2	307.680.068.903	1.774.371	204.125.589.367
1.350.523	1.700.3	122.345.011.954	1.785.372	189.093.652.801
1.394.644	1.325.8	4.733.084.444	1.362.721	1.019.090.698
1.229.691	1.210.7	359.923.953	1.383.895	23.778.811.934
1.329.419	1.843.4	264.189.069.554	1.681.326	123.838.677.412
1.172.038	1.676.9	254.887.629.653	1.591.708	176.123.076.768
748.580	1.235.3	236.964.689.470	1.251.898	253.329.210.451
1.578.553	1.365.3	45.437.643.086	1.340.876	56.490.451.400
1.177.805	1.322.7	21.007.810.677	1.383.895	42.473.005.664
2.050.180	1.683.0	134.814.221.808	1.769.342	78.869.982.244
1.170.500	1.184.2	189.590.977	1.187.826	300.176.415
1.573.430	1.484.1	7.966.197.243	1.313.935	67.337.551.227
1.586.245	1.595.5	85.708.288	1.774.864	35.577.278.056
1.472.090	1.784.8	97.843.947.667	1.681.326	43.779.787.390
1.328.181	1.394.9	4.457.824.147	1.352.478	590.324.772
1.374.609	1.215.5	25.308.618.434	1.180.408	37.713.873.040
1.157.922	1.291.6	17.873.415.394	1.212.317	2.958.859.541
1.554.289	1.495.9	3.402.802.445	1.352.960	40.533.205.178
2.006.801	1.615.3	153.221.603.903	1.769.342	56.386.776.681
	Total	2.688.728.300.418	Total	2.271.669.365.465
	MSE_s	89.624.276.681	MSE_s	75.722.312.182

Nilai y yang *dimissingkan*, hasil imputasi dengan kedua metode dan MSE_s

y <i>dimissingkan</i> 20 buah (25 %)				
y_{miss}	Regresi		PMM	
	y_{imp}	error ²	y_{imp}	error ²
1.511.791	1.568.519	3.218.061.213	1.393.275	14.046.042.256
1.039.585	1.651.401	374.318.564.459	1.459.918	176.679.662.756
1.794.485	1.568.480	51.078.146.669	1.625.149	28.674.816.365
1.459.335	1.600.006	19.788.205.673	1.576.751	13.786.517.056
1.664.174	1.317.786	119.984.922.264	1.313.163	123.208.862.525
1.431.979	1.378.490	2.861.049.377	1.415.040	286.936.497
1.824.349	1.641.301	33.506.549.148	1.623.610	40.296.065.825
910.889	1.452.456	293.294.672.806	1.332.656	177.887.739.703
748.580	1.346.680	357.723.209.653	1.278.948	281.290.003.277
1.524.306	1.606.935	6.827.602.631	1.460.037	4.130.504.361
1.567.109	1.506.271	3.701.304.325	1.509.757	3.289.274.845
1.170.500	1.272.810	10.467.278.140	1.230.377	3.585.255.129
1.351.914	1.320.295	999.772.110	1.377.251	641.983.839
1.472.090	1.750.810	77.684.632.395	1.586.245	13.031.364.025
1.465.709	1.267.824	39.158.451.817	1.321.571	20.775.878.355
1.374.609	1.390.336	247.351.982	1.223.291	22.897.197.651
1.554.289	1.140.827	170.951.033.297	1.350.625	41.479.106.362
1.220.522	1.759.873	290.899.014.488	1.429.397	43.628.849.175
1.622.649	1.368.317	64.684.846.230	1.260.758	130.965.385.394
516.447	1.001.027	234.818.169.279	1.030.877	264.638.019.128
	Total	2.156.212.837.958	Total	1.405.219.464.523
	MSE_s	107.810.641.898	MSE_s	70.260.973.226

Nilai y yang *dimissingkan*, hasil imputasi dengan kedua metode dan MSE_s

y <i>dimissingkan</i> 10 buah (12,5 %)				
y_{miss}	Regresi		PMM	
	y_{imp}	error ²	y_{imp}	error ²
1.511.79	1.546.023	1.171.803.860	1.434.149	6.028.311.221
1.144.38	1.721.596	333.167.568.15	1.622.135	228.241.049.418
1.683.50	1.746.259	3.938.231.082	1.515.430	28.249.003.935
1.431.97	1.456.978	624.971.699	1.431.708	73.333
910.889	1.333.966	178.994.491.20	1.360.881	202.492.800.064
1.153.66	1.452.372	89.225.165.623	1.453.146	89.688.509.984
1.203.96	1.292.720	7.878.520.970	1.318.068	13.020.726.951
2.050.18	1.632.518	174.441.345.96	1.639.259	168.856.068.241
1.586.24	1.761.863	30.841.508.262	1.664.174	6.072.929.041
2.006.80	1.603.170	162.917.395.05	1.622.649	147.572.759.104
	Total	983.201.301.86	Total	890.222.231.291
	MSE_s	98.320.130.186	MSE_s	89.022.223.129

Skenario-4

Data Awal

No	x ₁	x ₂	x ₃	y	No	x ₁	x ₂	x ₃	y
1	52	443.833	1.807	376.411	41	29	159.080	1.692	125.114
2	32	414.564	1.614	361.432	42	49	342.773	1.258	290.134
3	28	188.245	1.959	126.435	43	38	208.110	766	228.503
4	35	460.945	1.656	473.864	44	48	185.176	737	122.563
5	50	200.535	1.768	192.443	45	52	384.132	1.493	378.421
6	31	321.522	1.716	268.556	46	31	391.511	1.599	330.608
7	43	108.526	1.935	45.893	47	55	328.474	720	314.991
8	51	425.817	615	452.611	48	39	403.824	1.667	413.176
9	42	186.191	713	212.223	49	52	444.193	1.130	447.259
10	58	307.127	1.405	301.693	50	43	246.796	717	241.541
11	56	491.619	1.220	429.754	51	47	158.587	1.874	172.201
12	35	312.784	1.037	321.014	52	36	232.587	1.863	252.511
13	49	171.478	695	181.669	53	57	459.906	834	427.900
14	35	126.571	1.508	121.506	54	35	325.854	1.981	345.529
15	49	432.960	1.527	423.383	55	42	306.121	1.759	310.602
16	32	371.765	984	348.750	56	56	286.754	1.139	274.184
17	38	411.433	1.834	405.787	57	36	392.432	549	334.792
18	31	233.365	1.243	250.626	58	35	415.675	551	444.075
19	53	139.957	722	127.132	59	33	386.589	1.053	362.554
20	31	471.956	1.582	430.631	60	36	329.400	1.946	360.357
21	38	124.918	1.018	125.745	61	52	330.453	1.864	330.273
22	33	208.779	1.949	216.634	62	58	334.815	1.697	306.372
23	30	207.983	1.775	168.941	63	42	298.024	1.186	285.539
24	39	381.699	899	397.637	64	39	271.305	1.580	217.618
25	54	338.544	1.504	342.355	65	35	477.331	1.118	480.747
26	58	239.778	1.557	192.586	66	48	274.406	1.174	280.795
27	49	395.640	1.052	336.919	67	53	452.962	631	388.636
28	57	112.633	795	133.030	68	55	158.062	1.802	118.193
29	44	168.258	1.380	107.055	69	56	120.334	517	149.109
30	28	126.032	1.388	130.825	70	52	493.697	618	449.337
31	50	318.626	1.486	294.799	71	27	176.163	736	108.916
32	30	112.184	1.606	135.753	72	39	145.502	730	94.342
33	53	220.426	1.075	203.968	73	51	127.508	571	78.087
34	53	123.291	602	77.789	74	36	168.925	738	106.199
35	48	145.845	794	128.366	75	46	497.236	1.008	516.234
36	54	444.555	1.952	426.378	76	41	363.987	1.952	360.286
37	56	263.577	1.983	277.679	77	53	126.508	1.374	130.319
38	54	127.073	861	67.307	78	25	422.975	795	449.077
39	38	347.303	1.377	292.024	79	40	468.230	1.766	466.333
40	26	302.838	984	251.836	80	58	399.765	1.739	415.617

Nilai y yang dimissingkan, hasil imputasi dengan kedua metode dan MSE_s

y dimissingkan 40 buah (50 %)				
y_{miss}	Regresi		PMM	
	y_{imp}	error ²	y_{imp}	error ²
181.669	145.003	1.344.404.663	164.975	278.702.991
348.750	325.713	530.699.027	354.898	37.792.986
405.787	388.786	289.017.482	383.746	485.805.681
127.132	142.310	230.378.991	126.228	816.493
125.745	121.432	18.591.317	122.703	9.252.547
216.634	210.481	37.855.555	191.699	621.774.173
397.637	364.562	1.073.982.109	364.467	1.100.222.364
192.586	206.584	195.957.403	214.131	464.169.789
336.919	388.744	2.685.806.284	373.640	1.348.402.464
133.030	104.963	787.781.082	99.826	1.102.479.053
107.055	149.083	1.766.338.284	153.075	2.117.877.216
130.825	102.102	824.999.127	124.254	43.172.784
135.753	131.327	19.592.857	99.826	1.290.720.588
203.968	211.716	60.030.404	200.355	13.053.769
77.789	89.727	142.517.098	121.717	1.929.704.327
128.366	114.330	196.996.697	132.314	15.589.863
426.378	426.931	305.778	411.759	213.715.161
125.114	125.862	559.659	144.008	356.990.794
290.134	309.008	356.224.381	321.014	953.574.400
228.503	204.166	592.301.460	195.655	1.079.017.383
378.421	366.421	144.011.875	358.700	388.909.953
172.201	156.376	250.435.686	141.593	936.849.664
274.184	266.630	57.062.204	267.172	49.173.754
334.792	384.626	2.483.393.226	375.125	1.626.718.623
444.075	423.752	413.014.418	398.048	2.118.447.908
360.357	310.754	2.460.495.901	307.187	2.827.048.900
330.273	293.868	1.325.360.324	303.623	710.201.180
306.372	322.260	252.413.152	308.065	2.865.572
217.618	253.106	1.259.366.428	253.149	1.262.451.961
480.747	448.523	1.038.416.473	458.971	474.176.755
280.795	255.542	637.738.942	252.376	807.639.561
118.193	124.995	46.265.629	136.072	319.672.944
108.916	182.694	5.443.203.108	166.651	3.333.330.225
94.342	143.460	2.412.542.126	139.830	2.069.194.535
78.087	111.115	1.090.827.711	122.703	1.990.605.302
106.199	165.928	3.567.499.493	157.250	2.606.225.021
360.286	341.279	361.269.611	336.576	562.145.132
449.077	406.160	1.841.836.688	402.443	2.174.748.610
466.333	424.944	1.713.089.644	441.023	640.575.852
415.617	425.410	95.906.711	373.089	1.808.647.795
	Total	38.068.489.005	Total	40.172.464.073
	MSE_s	951.712.225	MSE_s	1.004.311.602

Nilai y yang dimissingkan, hasil imputasi dengan kedua metode dan MSE_s

y dimissingkan 30 buah (37,5 %)				
y_{miss}	Regresi		PMM	
	y_{imp}	error ²	y_{imp}	error ²
376.411	412.500	1.302.408.580	420.597	1.952.367.247
473.864	427.810	2.120.963.360	441.405	1.053.586.681
192.443	195.675	10.443.032	175.626	282.811.489
212.223	165.899	2.145.887.846	167.583	1.992.765.312
321.014	302.498	342.849.409	297.021	575.654.452
181.669	198.175	272.447.928	157.042	606.498.980
121.506	106.361	229.364.686	111.168	106.878.379
423.383	419.142	17.989.544	415.617	60.310.756
405.787	379.844	673.025.749	386.593	368.409.636
127.132	139.293	147.899.063	123.043	16.716.650
397.637	345.243	2.745.111.808	365.691	1.020.534.138
336.919	360.714	566.225.737	378.421	1.722.416.004
130.825	100.370	927.490.274	114.269	274.087.891
203.968	210.255	39.527.690	202.205	3.107.464
277.679	246.761	955.922.882	238.573	1.529.247.951
67.307	105.726	1.476.021.287	110.997	1.908.781.148
290.134	310.428	411.835.716	321.104	959.128.512
228.503	178.293	2.521.015.236	188.509	1.599.520.036
330.608	365.249	1.199.964.552	370.271	1.573.121.839
413.176	406.712	41.779.551	382.507	940.587.561
345.529	298.826	2.181.157.911	304.290	1.700.622.130
334.792	380.326	2.073.342.735	379.109	1.964.014.216
362.554	388.264	661.020.583	370.271	59.545.916
360.357	306.872	2.860.674.920	307.473	2.796.696.302
480.747	457.937	520.284.061	457.445	542.983.204
280.795	226.900	2.904.678.497	251.999	829.209.616
108.916	188.339	6.308.077.453	159.641	2.573.025.625
94.342	125.671	981.509.836	131.154	1.355.138.069
78.087	110.915	1.077.648.388	111.339	1.105.695.504
360.286	342.996	298.953.448	342.497	316.462.752
	Total	38.015.521.762	Total	31.789.925.461
	MSE_s	1.267.184.059	MSE_s	1.059.664.182

Nilai y yang *dimissingkan*, hasil imputasi dengan kedua metode dan MSE_s

y <i>dimissingkan</i> 20 buah (25 %)				
Y_{miss}	Regresi		PMM	
	Y_{imp}	error ²	Y_{imp}	error ²
361.432	418.406	3.246.060.017	406.382	2.020.520.480
126.435	154.626	794.712.229	177.602	2.618.020.956
301.693	306.828	26.369.302	281.194	420.225.400
429.754	474.249	1.979.807.716	469.345	1.567.478.954
405.787	406.711	854.432	400.745	25.423.781
127.132	137.154	100.442.463	116.506	112.907.626
216.634	194.032	510.838.149	198.883	315.105.101
342.355	330.182	148.188.305	318.932	548.655.668
228.503	196.563	1.020.149.055	197.232	977.887.949
314.991	311.233	14.120.823	309.300	32.382.928
252.511	184.108	4.679.024.431	216.436	1.301.434.485
274.184	275.951	3.121.194	265.212	80.496.784
334.792	376.635	1.750.850.084	386.777	2.702.461.019
217.618	244.456	720.268.943	258.282	1.653.560.896
480.747	485.223	20.035.766	469.345	129.996.483
388.636	440.182	2.656.560.331	433.774	2.037.439.044
108.916	142.031	1.096.588.995	170.718	3.819.462.483
78.087	88.362	105.580.775	108.625	932.581.659
516.234	499.828	269.146.130	473.864	1.795.216.900
130.319	88.259	1.769.076.263	103.999	692.752.928
	Total	20.912.195.402	Total	23.784.011.524
	MSE_s	1.045.609.770	MSE_s	1.189.200.576

Nilai y yang *dimissingkan*, hasil imputasi dengan kedua metode dan MSE_s

y <i>dimissingkan</i> 10 buah (12,5 %)				
Y_{miss}	Regresi		PMM	
	Y_{imp}	error ²	Y_{imp}	error ²
361.432	393.666	1.039.044.873	400.575	1.532.143.135
212.223	157.456	2.999.429.774	164.975	2.232.411.303
121.506	85.706	1.281.647.000	107.313	201.435.572
127.132	121.702	29.479.636	114.375	162.746.152
342.355	332.222	102.668.011	316.522	667.323.223
107.055	153.132	2.123.098.173	149.109	1.768.538.916
252.511	219.798	1.070.109.503	216.671	1.284.476.928
345.529	311.000	1.192.269.028	309.556	1.294.085.508
360.357	302.649	3.330.199.467	312.658	2.275.175.521
360.286	366.651	40.514.766	346.384	193.271.165
	Total	13.208.460.233	Total	11.611.607.422
	MSE_s	1.320.846.023	MSE_s	1.161.160.742

Skenario-5

Data Awal

No	x ₁	x ₂	y	No	x ₁	x ₂	y
1	33	457.292	342.326	41	37	442.467	352.937
2	50	217.609	94.643	42	40	146.777	126.698
3	38	396.180	242.365	43	58	485.593	341.210
4	35	201.066	120.875	44	35	386.286	296.182
5	36	153.558	124.289	45	25	307.112	166.255
6	59	139.231	130.375	46	34	325.078	248.352
7	36	171.036	70.551	47	49	443.461	357.045
8	48	498.680	330.197	48	30	441.273	348.278
9	30	482.960	339.316	49	42	179.608	75.929
10	46	480.081	338.472	50	53	116.346	39.195
11	51	470.142	338.767	51	49	266.895	166.583
12	58	296.760	173.541	52	40	110.565	91.943
13	59	373.244	250.216	53	38	151.488	54.979
14	33	179.744	78.517	54	47	449.794	304.952
15	26	313.563	167.722	55	54	398.009	235.769
16	34	214.712	148.352	56	57	451.133	347.736
17	57	422.101	304.084	57	27	348.298	221.756
18	51	100.875	36.780	58	37	426.896	255.053
19	36	432.970	342.806	59	59	336.660	251.890
20	31	114.211	17.613	60	52	464.704	326.564
21	49	189.496	139.762	61	42	382.247	258.644
22	46	390.719	299.223	62	40	475.026	333.216
23	32	389.102	251.198	63	42	254.361	132.670
24	54	357.947	219.375	64	55	450.310	344.106
25	46	416.025	336.307	65	38	226.185	122.671
26	60	447.698	355.513	66	36	327.808	273.686
27	34	189.458	140.723	67	60	279.118	171.627
28	59	142.556	132.930	68	27	302.062	221.074
29	40	474.201	358.230	69	58	143.267	84.736
30	41	167.144	63.525	70	49	364.607	287.473
31	37	239.201	113.294	71	58	431.040	320.035
32	34	462.491	370.919	72	28	463.041	373.437
33	56	114.521	76.804	73	47	239.425	175.953
34	47	288.557	168.464	74	42	477.671	380.045
35	35	196.536	105.355	75	38	392.682	264.503
36	42	246.186	214.341	76	42	348.042	282.577
37	54	336.200	210.737	77	41	441.950	338.808
38	43	166.144	65.241	78	58	495.532	353.808
39	29	378.107	273.405	79	49	359.717	265.041
40	31	199.283	129.231	80	54	181.683	99.035

Nilai y yang *dimissingkan*, hasil imputasi dengan kedua metode dan MSE_s

y <i>dimissingkan</i> 40 buah (50 %)				
y_{miss}	Regresi		PMM	
	y_{imp}	error ²	y_{imp}	error ²
342.326	338.387	15.518.882	340.142	4.768.109
94.643	125.204	47.142.081.404	139.387	41.184.237.721
242.365	280.576	3.813.097.128	288.256	2.923.608.156
120.875	124.176	47.589.367.045	130.670	44.798.177.674
124.289	76.500	70.663.254.981	99.141	59.138.944.225
130.375	66.062	76.321.548.441	67.980	75.265.617.978
70.551	122.253	48.432.157.166	112.604	52.772.381.062
330.197	367.830	650.469.731	370.159	774.653.623
339.316	361.565	370.143.935	362.780	418.374.298
338.472	354.066	137.831.643	352.354	100.552.762
338.767	353.482	124.449.394	342.134	37.018
173.541	186.742	24.206.389.229	185.114	24.715.738.714
250.216	250.229	8.481.784.032	261.388	6.550.992.219
78.517	112.966	52.605.968.097	119.944	49.453.931.830
167.722	245.701	9.336.375.205	227.225	13.248.286.241
148.352	138.486	41.550.807.930	139.387	41.184.237.721
304.084	262.318	6.401.301.264	299.690	1.817.828.496
36.780	62.865	78.098.409.094	46.542	87.488.292.970
342.806	312.718	876.639.178	322.647	387.278.784
17.613	48.205	86.507.538.044	70.344	73.974.425.910
139.762	93.153	62.087.052.284	119.944	49.453.931.830
299.223	268.637	5.430.012.741	281.838	3.658.798.144
251.198	296.412	2.108.072.236	288.079	2.942.737.009
219.375	264.871	5.999.318.845	251.690	8.214.920.750
336.307	293.430	2.390.814.793	299.690	1.817.828.496
355.513	328.937	179.255.385	323.977	336.685.801
140.723	117.210	50.677.349.216	122.360	48.385.129.142
132.930	98.820	59.295.101.385	80.550	68.526.569.466
358.230	347.517	26.943.793	350.273	63.157.988
63.525	100.801	58.334.369.349	103.149	57.205.541.658
113.294	166.231	31.009.408.799	168.195	30.321.744.466
370.919	332.938	88.140.681	343.557	1.514.869
76.804	54.234	82.997.216.505	59.215	80.151.838.321
168.464	168.459	30.229.665.541	184.977	24.758.644.861
105.355	141.267	40.424.905.086	130.670	44.798.177.674
214.341	158.887	33.649.686.045	168.326	30.276.069.600
210.737	216.380	15.862.390.649	238.286	10.824.404.832
65.241	107.667	55.064.629.780	100.902	58.285.354.637
273.405	281.072	3.752.106.739	276.513	4.331.403.620
129.231	143.888	39.377.643.089	130.670	44.798.177.674
	Total	1.182.309.014.762	Total	1.145.354.996.346
	MSE_s	29.557.725.369	MSE_s	28.633.874.909

Nilai y yang *dimissingkan*, hasil imputasi dengan kedua metode dan MSE_s

y <i>dimissingkan</i> 30 buah (37,5 %)				
y_{miss}	Regresi		PMM	
	y_{imp}	error ²	y_{imp}	error ²
342.326	336.834	30.162.385	339.809	6.336.296
94.643	145.526	38.730.401.639	138.699	41.463.873.678
242.365	262.424	6.384.400.310	286.514	3.115.001.669
120.875	110.381	53.798.485.060	130.670	44.798.177.674
124.289	92.458	62.434.208.131	91.598	62.864.730.567
130.375	69.494	74.437.466.344	72.000	73.076.254.406
70.551	131.131	44.603.475.784	106.136	55.785.810.576
330.197	378.664	1.320.453.712	368.144	666.579.451
339.316	349.428	50.432.915	361.947	384.991.489
338.472	359.746	303.468.209	353.349	121.515.348
338.767	369.331	729.268.036	342.722	157.133
173.541	205.141	18.819.727.931	212.804	16.775.844.867
250.216	239.981	10.474.585.741	260.216	6.742.052.100
78.517	108.867	54.502.968.323	116.874	50.828.694.485
167.722	213.712	16.541.577.330	225.878	13.560.090.125
148.352	125.584	46.976.933.959	138.699	41.463.873.678
304.084	309.458	1.080.281.777	302.707	1.569.696.856
36.780	54.412	82.894.373.974	44.404	88.757.398.915
342.806	296.461	2.103.629.254	322.647	387.278.784
17.613	50.874	84.944.269.343	62.169	78.487.720.524
139.762	99.709	58.862.868.343	115.581	51.413.113.629
299.223	285.860	3.188.391.023	281.838	3.658.798.144
251.198	293.935	2.341.708.119	284.595	3.332.822.176
219.375	273.656	4.715.616.243	251.582	8.234.437.238
336.307	310.528	1.011.112.890	301.444	1.671.337.924
355.513	332.550	95.560.689	323.977	336.685.801
140.723	107.005	55.376.143.876	122.818	48.183.586.458
132.930	66.120	76.289.722.766	75.401	71.248.742.085
358.230	363.953	467.720.738	350.490	66.650.896
63.525	73.350	72.347.917.155	100.468	58.495.098.678
	Total	875.857.331.999	Total	827.497.351.650
	MSE_s	29.195.244.400	MSE_s	27.583.245.055

Nilai y yang *dimissingkan*, hasil imputasi dengan kedua metode dan MSE_s

y <i>dimissingkan</i> 20 buah (25 %)				
y_{miss}	Regresi		PMM	
	y_{imp}	$error^2$	y_{imp}	$error^2$
342.326	327.080	232.440.353	338.190	17.104.842
94.643	156.530	34.520.185.608	143.886	39.378.274.848
242.365	276.871	4.284.394.669	286.514	3.115.001.669
120.875	140.085	40.901.305.459	132.385	44.075.307.457
124.289	92.252	62.536.758.687	91.598	62.864.730.567
130.375	91.211	63.058.680.223	81.989	67.775.249.434
70.551	115.998	51.224.352.717	107.724	55.038.286.086
330.197	386.116	1.917.524.943	370.919	817.559.649
339.316	335.058	52.817.623	361.467	366.362.568
338.472	346.031	13.730.065	354.622	151.196.534
338.767	360.222	320.250.594	346.019	13.638.249
173.541	199.145	20.500.880.508	212.941	16.740.529.979
250.216	250.701	8.395.080.480	265.902	5.840.658.346
78.517	121.303	48.851.257.655	114.193	52.044.665.689
167.722	228.820	12.883.597.934	225.157	13.728.480.826
148.352	119.677	49.572.436.679	138.425	41.575.780.922
304.084	312.068	915.558.126	306.823	1.260.477.210
36.780	49.628	85.672.213.522	49.699	85.630.795.231
342.806	300.910	1.715.312.273	320.035	496.888.681
17.613	59.495	79.993.587.889	61.826	78.680.137.800
	Total	567.562.366.008	Total	569.611.126.587
	MSE_s	28.378.118.300	MSE_s	28.480.556.329

Nilai y yang *dimissingkan*, hasil imputasi dengan kedua metode dan MSE_s

y <i>dimissingkan</i> 10 buah (12,5 %)				
y_{miss}	Regresi		PMM	
	y_{imp}	$error^2$	y_{imp}	$error^2$
342.326	338.575	14.069.739	336.414	34.956.474
94.643	156.174	34.652.546.607	138.747	41.444.327.809
242.365	281.388	3.713.379.923	285.555	3.222.991.858
120.875	107.036	55.361.496.323	126.277	46.677.256.821
124.289	67.330	75.622.550.461	88.717	64.317.423.437
130.375	95.952	60.700.278.473	74.137	71.925.125.170
70.551	101.281	58.102.784.484	104.260	56.675.229.903
330.197	362.404	403.131.564	371.423	846.612.132
339.316	362.799	419.126.717	355.744	180.031.990
338.472	366.845	601.159.848	354.963	159.698.824
	Total	289.590.524.14	Total	285.483.654.418
	MSE_s	28.959.052.414	MSE_s	28.548.365.442