



---

TESIS - SS142501

**INTERVAL KONFIDENSI UNTUK PARAMETER  
MODEL REGRESI NONPARAMETRIK SPLINE  
*TRUNCATED* MULTIVARIABEL**

(Studi Kasus Pada Data Indeks Pembangunan Gender di  
Provinsi  
Jawa Timur)

RIFANI NUR SINDY S.  
NRP. 1315 201 208

DOSEN PEMBIMBING  
Prof. Dr. Drs. I Nyoman Budiantara, M.Si  
Dr. Vita Ratnasari, S.Si., M.Si.

PROGRAM MAGISTER  
JURUSAN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA  
2017



---

THESIS - SS142501

# **CONFIDENCE INTERVALS FOR PARAMETERS OF NONPARAMETRIC SPLINE TRUNCATED REGRESSION**

(Case Study On Development Gender Index In East Java)

RIFANI NUR SINDY S.  
NRP. 1315 201 208

SUPERVISOR  
Prof. Dr. Drs. I Nyoman Budiantara, M.Si  
Dr. Vita Ratnasari, S.Si., M.Si.

PROGRAM OF MAGISTER  
DEPARTMENT OF STATISTICS  
FACULTY OF MATHEMATICS AND NATURAL SCIENCES  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA  
2017

**INTERVAL KONFIDENSI UNTUK PARAMETER MODEL REGRESI  
NONPARAMETRIK SPLINE *TRUNCATED* MULTIVARIABEL  
(STUDI KASUS PADA DATA INDEKS PEMBANGUNAN GENDER DI  
JAWA TIMUR)**

Disusun untuk memenuhi syarat memperoleh gelar Magister Sains (M.Si)  
di  
Institut Teknologi Sepuluh Nopember

Oleh :  
**RIFANI NUR SINDY SETIAWAN**  
**NRP. 1315 2012 08**

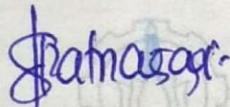
Tanggal Ujian : 20 Juni 2017  
Periode Wisuda : September 2017

Disetujui Oleh :



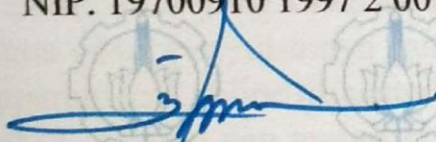
1. Prof. Dr. I Nyoman Budiantara, M.Si  
NIP. 19650603 198903 1 003

(Pembimbing I)



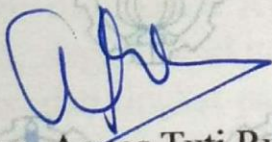
2. Dr. Vita Ratnasari, M.Si  
NIP. 19700910 1997 2 001

(Pembimbing II)



3. Dr. Dra. Ismaini Zain, M.Si  
NIP. 19600525 198803 2 001

(Penguji I)

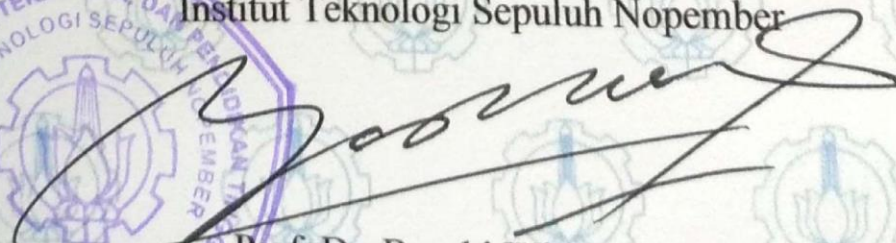
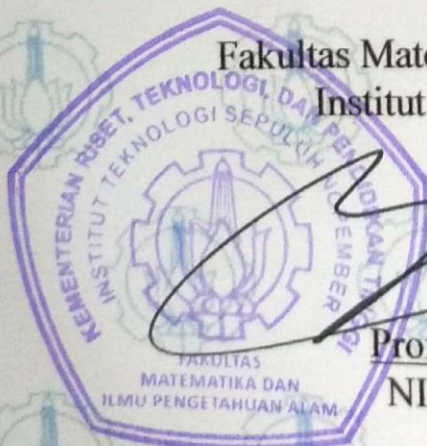


4. Dr. Dra. Agnes Tuti Rumiati, M.Sc  
NIP. 19570724 198503 2 002

(Penguji II)

Dekan

Fakultas Matematika dan Ilmu Pengetahuan Alam  
Institut Teknologi Sepuluh Nopember



Prof. Dr. Basuki Widodo, M.Sc  
NIP. 19650605 198903 1 002

## SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini, mahasiswa Departemen Statistika FMIPA ITS:

Nama : Rifani Nur Sindy Setiawan

NRP : 1315 2012 08

Program Studi : Magister Statistika

menyatakan bahwa data yang digunakan dalam Tesis ini merupakan data sekunder yang diambil dari publikasi BPS yaitu:

Sumber : BPS Provinsi Jawa Timur

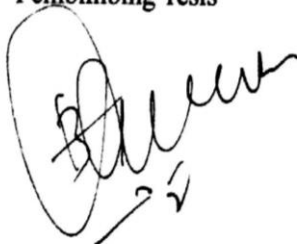
Keterangan : Pembangunan Manusia Berbasis Gender Provinsi Jawa Timur Tahun 2015, Profil Gender Provinsi Jawa Timur Tahun 2015, Laporan Eksekutif Kesehatan Provinsi Jawa Timur 2015, Hasil Survey Sosial Ekonomi Nasional Tahun 2014 Provinsi Jawa Timur.

Surat pernyataan ini dibuat dengan sebenarnya. Apabila terdapat pemalsuan data maka saya siap menerima sanksi sesuai aturan yang berlaku.

Surabaya, Juli 2017

**Mengetahui :**

Pembimbing Tesis



Prof. Dr. Drs. I Nyoman Budiantara

NIP. 19650603 198903 1 003

Mahasiswa



Rifani Nur Sindy Setiawan

NRP. 1315 2012 08

**INTERVAL KONFIDENSI UNTUK PARAMETER MODEL  
REGRESI NONPARAMETRIK  
SPLINE *TRUNCATED* MULTIVARIABEL  
(Studi Kasus Pada Data Indeks Pembangunan Gender di Provinsi  
Jawa Timur)**

**Nama** : Rifani Nur Sindy S.  
**NRP** : 1315201208  
**Jurusan** : Statistika FMIPA-ITS  
**Dosen Pembimbing** : Prof. Dr. Drs. I Nyoman Budiantara, M.Si  
Dr. Vita Ratnasari, S.Si., M.Si.

**ABSTRAK**

Salah satu bagian terpenting dari inferensi statistik adalah interval konfidensi. Banyak upaya telah dilakukan untuk mendapatkan interval konfidensi untuk regresi nonparametrik berdasarkan metode estimasi yang digunakan. Interval konfidensi untuk parameter dalam regresi nonparametrik dapat digunakan untuk menentukan variabel prediktor yang berpengaruh secara signifikan terhadap variabel respon. Dalam penelitian ini dikembangkan interval konfidensi untuk parameter model dan menggunakan beberapa variabel prediktor. Tetapi khusus *spline truncated* yang mempunyai sifat serupa dengan pendekatan regresi parametrik, namun didekati dengan fungsi *spline truncated*. Dari hasil penelitian didapatkan interval konfidensi saat  $\sigma^2$  diketahui dan interval konfidensi saat  $\sigma^2$  tidak diketahui. Interval konfidensi untuk parameter model regresi nonparametrik yang dihasilkan serupa dengan interval konfidensi regresi klasik, namun unsur penyusunnya yang berbeda. Pemodelan dengan menggunakan regresi nonparametrik *spline truncated* menghasilkan model terbaik dengan kombinasi titik knot (3, 2, 2, 1) dan memiliki nilai  $R^2$  87,48. Penerapan interval konfidensi untuk parameter model regresi nonparametrik *spline truncated* multivariabel pada penelitian ini diterapkan pada data Indeks Pembangunan Gender (IPG) di Provinsi Jawa Timur didapatkan APS Sekolah Lanjutan Tingkat Atas (SLTA) penduduk perempuan, angka kesakitan penduduk perempuan, penolong terakhir kelahiran oleh medis, dan tingkat partisipasi angkatan kerja (TPAK) penduduk perempuan berpengaruh secara signifikan terhadap IPG di provinsi Jawa Timur.

**Kata Kunci:** Regresi Nonparametrik, *Spline Truncated*, Interval Konfidensi, IPG.

*(halaman ini sengaja dikosongkan)*

**CONFIDENCE INTERVALS FOR PARAMETERS OF  
NONPARAMETRIC SPLINE TRUNCATED REGRESSION  
(Case Study On Development Gender Index In East Java)**

**Name** : Rifani Nur Sindy S.  
**NRP** : 1315201208  
**Departement** : Statistika FMIPA-ITS  
**Supervisor** : Prof. Dr. Drs. I Nyoman Budiantara, M.Si  
Dr. Vita Ratnasari, S.Si., M.Si.

**ABSTRACT**

One of the most important parts of statistical inference is the confidence interval. Many attempts have been made to obtain confidence intervals for nonparametric regression based on the estimation method used. Confidence intervals for parameters of nonparametric regression can be used to determine predictor variables that significantly influence response variables. In this research will be developed confidence interval for model parameters and using some predictor variables. Specifically, however the spline truncated has properties similar to the parametric regression approach, but is approximated by a truncated spline function. From the result obtained confidence intervals for parameters of nonparametric spline truncated regression divided into two forms, when  $\sigma^2$  known and  $\sigma^2$  unknown. The result obtained confidence intervals for parameters of nonparametric spline truncated regression similar with confidence intervals for parametric regression, but the constituent elements are different. Modelling using nonparametric spline truncated regression obtained the best model with combination knot (3, 2, 2, 1) and  $R^2$  equal to 87,48. Determination of predictor variables that significantly influence GDI by using confidence interval obtained high school enrollment rate of female population, morbidity of female population, percentage of last aid of birth by medical, and female labor-force participation rate have significantly influenced to GDI in East Java.

**Keywords:** Nonparametric Regression, Spline *Truncated*, Confidence Interval, Gender Development Index (GDI).

*(halaman ini sengaja dikosongkan)*



## KATA PENGANTAR

Puji dan syukur penulis hadirkan kepada Allah SWT, karena atas segala rahmat dan ridho-Nya sehingga tesis yang diberi judul “Interval Konfidensi Untuk Parameter Model regresi Nonparametrik Spline *Truncated* Multivariabel (Studi Kasus Pada Data Indeks Pembangunan Gender di Provinsi Jawa Timur)” ini bisa terselesaikan. Tesis ini merupakan salah satu syarat untuk menyelesaikan pendidikan di Program Magister S2 Statistika ITS. Ada banyak pihak yang telah membantu dalam penulisan tesis ini, sehingga penulis ingin menyampaikan ucapan terima kasih kepada

1. Bapak Bapak Prof. Dr. Drs. I Nyoman Budiantara, M.Si dan ibu Dr. Vita Ratnasari, S.Si., M.Si. selaku dosen pembimbing, yang telah bersedia meluangkan waktu untuk memberikan bimbingan, saran, dan ilmu yang sangat bermanfaat dalam penyelesaian tesis ini.
2. Ibu Dr. Dra. Ismaini Zain, M.Si. dan ibu Dr. Dra. Agnes Tuti Rumiati, M.Sc. selaku dosen penguji yang telah memberikan banyak saran dan masukan agar tesis ini menjadi lebih baik.
3. Bapak Dr. Suhartono, M.Sc. selaku Ketua Departemen Statistika ITS dan Bapak Dr. rer. pol. Heri Kuswanto, M.Si. selaku Kaprodi Pascasarjana Statistika FMIPA ITS.
4. Bapak /Ibu dosen pengajar di Jurusan Statistika ITS, terima kasih atas semua ilmu berharga yang telah diberikan.
5. Bapak/Ibu staf dan karyawan di Jurusan Statistika ITS, terima kasih atas segala bantuan selama masa perkuliahan penulis
6. Kedua orang tua yang sangat penulis sayangi dan hormati, Ibu Eni Suryanti dan Bapak Alex Setiawan yang tidak pernah lelah mendoakan yang terbaik untuk penulis serta selalu memberi motivasi untuk tidak pernah menyerah.
7. Semua teman-teman seperjuangan S2 Statistika ITS, terima kasih atas bantuan dan kebersamaan selama ini (Nisa, Ce Ribka, Ayak, Chusnul, Halwa, Uni, Mbak Titis, Mas Rama, Jaya, Sem, Luthfi, dan (alm.) Mail).

8. Teman-teman Iprit (Eva, Irma, Husnul, Fuji), FFS (Firstca dan Sindy), dan Nilam yang selalu memberi semangat pada penulis saat penulis mengalami kesulitan.
9. Serta, semua pihak yang telah membantu penulis, namun tidak dapat penulis sebutkan satu per satu.

Penulis menyadari bahwa tesis ini masih jauh dari sempurna, sehingga kritik dan saran sangat diharapkan. Semoga tesis ini dapat memberikan manfaat guna memperluas wawasan keilmuan pembacanya.

Surabaya, July 2017

Penulis

## DAFTAR ISI

<b>ABSTRAK</b> .....	iii
<b>ABSTRACT</b> .....	v
<b>KATA PENGANTAR</b> .....	vii
<b>DAFTAR ISI</b> .....	ix
<b>DAFTAR TABEL</b> .....	xi
<b>DAFTAR GAMBAR</b> .....	xii
<b>BAB 1 PENDAHULUAN</b> .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	5
1.3 Tujuan Penelitian .....	5
1.4 Manfaat Penelitian .....	5
1.5 Batasan Masalah .....	6
<b>BAB 2 TINJAUAN PUSTAKA</b> .....	7
2.1 Analisis Regresi .....	7
2.2 Regresi Nonparametrik .....	7
2.3 Regresi Nonparametrik Spline <i>Truncated</i> .....	8
2.4 Pemilihan Titik Knot Optimal .....	9
2.5 Interval Konfidensi Untuk Parameter Regresi .....	9
2.6 Pemeriksaan Asumsi Residual .....	12
2.6.1 Asumsi Independensi Residual .....	12
2.6.2 Asumsi Residual Identik .....	12
2.6.3 Asumsi Normalitas Residual .....	13
2.7 Koefisien Determinasi $R^2$ .....	13
2.8 <i>Pivotal Quantity</i> .....	14
2.9 Indeks Pembangunan Gender .....	14
2.10 Faktor- Faktor Yang Mempengaruhi IPG .....	15
<b>BAB 3 METODE PENELITIAN</b> .....	19
3.1 Sumber Data .....	19
3.2 Variabel Penelitian .....	19

3.3 Struktur Data Penelitian .....	20
3.4 Tahapan Penelitian.....	21
<b>BAB 4 PEMBAHASAN</b> .....	25
4.1 Interval Konfidensi Untuk Parameter Regresi Nonparametrik Spline <i>Truncated</i> Multivariabel.....	25
4.2 Aplikasi Interval Konfidensi Untuk Parameter Regresi Nonparametrik Spline <i>Truncated</i> Multivariabel Pada Data IPG di Jawa Timur.....	39
4.2.1 Deskripsi IPG dan Faktor-Faktor yang Diduga Mempengaruhi .	39
4.2.2 Pemodelan IPG di Provinsi Jawa Timur dengan Regresi Nonparametrik Spline Linier Multivariabel.....	42
4.2.2.1 Pemilihan Titik Knot Optimal dengan Satu Titik Knot.....	42
4.2.2.2 Pemilihan Titik Knot Optimal dengan Dua Titik Knot.....	43
4.2.2.3 Pemilihan Titik Knot Optimal dengan Tiga Titik Knot.....	44
4.2.2.4 Pemilihan Titik Knot Optimal dengan Kombinasi Titik Knot....	45
4.2.2.5 Pemodelan dengan Titik Knot Optimal .....	45
4.2.2.6 Interval Konfidensi Untuk Parameter Model Dalam Regresi Nonparametrik Spline Linier Multivariabel .....	46
4.2.2.7 Pemeriksaan Asumsi Residual .....	47
(i) Pemeriksaan Asumsi Independensi residual .....	47
(ii) Pemeriksaan Asumsi Identik Residual .....	48
(iii) Pemeriksaan Asumsi Normalitas Residual .....	49
4.2.3 Interpretasi Model Regresi Nonparametrik Spline <i>Truncated</i> Multivariabel Pada Data IPG di Jawa Timur .....	49
<b>BAB 5 PEMBAHASAN</b> .....	53
5.1 Kesimpulan .....	53
5.2 Saran .....	54
<b>DAFTAR PUSTAKA</b> .....	53
<b>LAMPIRAN</b> .....	59
<b>BIOGRAFI PENULIS</b> .....	83

## DAFTAR TABEL

<b>Tabel 3.1</b> Variabel Penelitian .....	19
<b>Tabel 3.2</b> Struktur Data Penelitian .....	21
<b>Tabel 4.1</b> Deskripsi IPG dan Faktor Yang Diduga Mempengaruhi .....	40
<b>Tabel 4.2</b> Nilai GCV untuk Spline Linier Satu Knot .....	43
<b>Tabel 4.3</b> Nilai GCV untuk Spline Linier Dua Knot.....	44
<b>Tabel 4.4</b> Nilai GCV untuk Spline Linier Tiga Knot.....	45
<b>Tabel 4.5</b> Nilai GCV dengan Kombinasi Titik Knot.....	45
<b>Tabel 4.6</b> Perbandingan Nilai GCV Minimum.....	46
<b>Tabel 4.7</b> Pengujian Parameter Menggunakan Interval Konfidensi.....	47
<b>Tabel 4.7</b> ANOVA dari Uji Glejser.....	49

*(halaman ini sengaja dikosongkan)*

## DAFTAR GAMBAR

<b>Gambar 4.1</b> Distribusi Normal .....	31
<b>Gambar 4.2</b> Distribusi <i>student-t</i> .....	38
<b>Gambar 4.3</b> Persebaran IPG di Jawa Timur Tahun 2014 .....	41
<b>Gambar 4.4</b> <i>Scatterplot</i> Antara IPG dengan Faktor Yang Diduga Mempengaruhi .....	42
<b>Gambar 4.3</b> Plot ACF Residual .....	48

*(halaman ini sengaja dikosongkan)*



# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang

Analisis regresi adalah salah satu metode yang digunakan untuk mengetahui hubungan antar satu variabel respon dengan satu atau lebih variabel prediktor. Pendekatan regresi dapat dilakukan dengan tiga pendekatan yaitu regresi parametrik, regresi semiparametrik, dan regresi nonparametrik. Pendekatan regresi parametrik membutuhkan asumsi- asumsi diantaranya bentuk kurva harus diketahui (mengikuti suatu pola tertentu misal linier, kuadratik, kubik), *error* berdistribusi normal, dan varians yang homogen. Suatu pendekatan regresi nonparametrik merupakan metode statistika yang digunakan untuk mengetahui hubungan antara variabel respon dengan variabel prediktor yang tidak diketahui bentuk fungsinya, hanya diasumsikan *smooth* (mulus) dalam arti termuat dalam suatu ruang fungsi tertentu. Sedangkan pendekatan regresi semiparametrik yaitu regresi yang memuat komponen parametrik dan komponen nonparametrik (Budiantara, 2009).

Pendekatan regresi nonparametrik merupakan regresi yang sangat fleksible dalam memodelkan pola data (Eubank, 1988). Pendekatan regresi nonparametrik digunakan untuk mengatasi bentuk kurva regresi yang tidak memiliki pola. Berbeda dengan statistik nonparametrik merupakan metode statistik yang digunakan dengan mengabaikan segala asumsi, terutama berkaitan dengan suatu distribusi tertentu. Pendekatan regresi nonparametrik telah banyak dikembangkan antara lain menggunakan Kernel, Spline, Polinomial lokal, dan deret fourier. Analisis regresi spline mampu menangani karakter data/fungsi yang bersifat mulus dan perilakunya yang berubah- ubah pada sub interval tertentu. Spline memiliki beberapa kelebihan diantaranya memiliki intepretasi statistik yang sederhana dan baik serta penyajian visual yang baik. Analisis regresi nonparametrik spline univariabel adalah analisis regresi nonparametrik jika terdapat satu variabel respon dan satu variabel prediktor. Apabila terdapat satu variabel respon dan lebih dari satu variabel prediktor maka disebut dengan regresi nonparametrik Spline multivariabel (Budiantara, 2004).

Penelitian mengenai regresi nonparametrik Spline *Truncated* multivariabel yang telah dilakukan, diantaranya Merdekawati (2013) meneliti mengenai pemodelan regresi spline *truncated* multivaribel diterapkan pada data kemiskinan di kabupaten/kota Provinsi Jawa Tengah. Wasono (2014) meneliti tentang regresi nonparametrik multivariabel heteroskedastisitas spline untuk mengatasi pelanggaran asumsi variansi *error* yang tidak konstan untuk setiap pengamatan (heteroskedastisitas). Kemudian Pratiwi (2015) mengkaji tentang pengujian hipotesis model spline *truncated* dalam regresi nonparametrik multivariabel.

Salah satu bagian terpenting dari inferensi statistik adalah interval konfidensi. Banyak upaya telah dilakukan untuk mendapatkan interval konfidensi untuk regresi nonparametrik berdasarkan metode estimasi yang digunakan. Beberapa penelitian mengenai interval konfidensi diantaranya dilakukan oleh Wahba (1983) interval konfidensi dengan Bayesian untuk *cross* validasi *smoothing* spline, Nychka (1988) interval konfidensi dengan Bayesian dengan *smoothing* spline, Mao dan Zhao (2003) mengkaji tentang konfidensi interval untuk kurva regresi dengan pendekatan polinomial spline, Intansari (2016) meneliti tentang estimasi titik dan interval konfidensi dalam kurva regresi dalam regresi Spline kuadrat. Namun dari beberapa penelitian tersebut hanya terbatas pada regresi spline univariabel dan dalam konteks regresi nonparametrik belum dikembangkan interval konfidensi untuk parameter model regresi nonparametrik spline multivariabel dengan pendekatan *pivotal quantity*.

Interval konfidensi untuk parameter dalam regresi nonparametrik dapat digunakan untuk menentukan variabel prediktor yang berpengaruh secara signifikan terhadap variabel respon. Pengambilan kesimpulan tersebut dilakukan dengan melihat apakah interval konfidensi parameter memuat nilai nol. Jika interval konfidensi memuat nilai nol, maka variabel prediktor tidak berpengaruh secara signifikan terhadap variabel respon. Dalam penelitian ini dikembangkan interval konfidensi untuk parameter model dan menggunakan beberapa variabel prediktor. Tetapi khusus spline *truncated* yang mempunyai sifat serupa dengan pendekatan regresi parametrik, namun didekati dengan fungsi spline *truncated*. Penerapan interval konfidensi untuk parameter model regresi nonparametrik spline *truncated*

multivariabel pada penelitian ini akan diterapkan pada data Indeks Pembangunan Gender (IPG) di Provinsi Jawa Timur.

Hakekat pembangunan nasional adalah pembangunan manusia seutuhnya dan pembangunan masyarakat Indonesia secara adil dan merata. Namun untuk dapat menciptakan kondisi tersebut ada beberapa permasalahan yang dihadapi, antara lain masih adanya kesenjangan pencapaian pembangunan antara perempuan dan laki-laki serta masih rendahnya kualitas hidup dan peran perempuan dalam pembangunan. Kesenjangan gender di berbagai bidang pembangunan ditandai pula oleh masih rendahnya peluang yang dimiliki perempuan untuk bekerja dan berusaha, serta rendahnya akses mereka terhadap sumber daya ekonomi seperti teknologi, informasi, pasar, kredit, dan modal kerja. Perbedaan peran gender yang ada di Indonesia merupakan masalah ketidakadilan sosial yang menempatkan perempuan sebagai korban utamanya. Bentuk-bentuk ketidaksetaraan gender dan keadilan gender dikenal dengan kesenjangan gender (*Gender Gap*) yang pada gilirannya menimbulkan permasalahan gender (BPS, 2015).

Untuk meningkatkan kesetaraan dan keadilan gender, maka kebutuhan dasar perempuan seperti kesehatan, pendidikan, serta partisipasi kerja harus mendapat perhatian. Kebutuhan dasar tersebut mencerminkan kualitas dari sumber daya manusia. Pemerintah telah berupaya untuk mewujudkan kesetaraan dan keadilan gender dalam kehidupan bermasyarakat dan bernegara melalui beberapa kebijakan dan program-program. Namun dalam prakteknya masih banyak menemui kendala dan tantangan (BPS, 2015).

Untuk mengevaluasi sejauh mana pembangunan sudah responsif gender dapat digunakan ukuran indikator terkait, seperti Indeks Pembangunan Gender (IPG). IPG diperkenalkan oleh *United Nation Development Programs* (UNDP) dalam laporan Pembangunan Manusia tahun 1995. Dari angka IPG ini diharapkan mampu memberikan informasi mengenai perkembangan capaian pembangunan yang sudah mengakomodasi aspek-aspek gender (BPS, 2014). Berdasarkan data yang dikeluarkan oleh UNDP dalam publikasi BPS menyebutkan bahwa IPG Indonesia masih menempati posisi rendah dibandingkan negara-negara ASEAN (tidak termasuk Vietnam dan Myanmar) yaitu pada posisi ketiga terendah setelah Timor Leste dan Kamboja.

Pada tahun 2014 capaian IPG di Indonesia masih belum merata terlihat dari disparitas angka IPG antarprovinsi. Dalam skala nasional capaian IPG dari 34 provinsi, Jawa Timur menempati urutan nomor 16. Di wilayah pulau jawa, Jawa Timur menempati posisi terendah kedua setelah Jawa Barat. Jika dilihat perkembangan nilai IPG Jawa Timur dari tahun 2010 hingga tahun 2014 sudah mengalami peningkatan. Namun posisi capaian IPG Jawa Timur tahun 2014 memisahkan sebanyak 20 kabupaten/kota dibawah capaian IPG provinsi sebanyak 18 kabupaten/kota diatas capaian IPG provinsi. Kondisi ini memberikan gambaran masih banyak kabupaten/kota yang perlu peningkatan dalam program-program yang menuju pengarusutamaan gender.

Hafizh (2013) meneliti tentang disparitas gender di Jawa Timur dengan menggunakan pendekatan model regresi probit ordinal menyimpulkan bahwa faktor-faktor yang mempengaruhi disparitas gender adalah angka partisipasi sekolah (APS) tingkat SMP penduduk perempuan, persentase penduduk perempuan dengan pendidikan terakhir yang ditamatkan setingkat SMP, dan persentase penduduk perempuan yang bekerja di sektor formal. Hakim (2013) juga pernah meneliti tentang komponen Indeks Pembangunan Gender dengan Geographically Weighted Multivariate Regression Model di Provinsi Kalimantan Timur dan Kalimantan Selatan tahun 2011. Dari penelitian tersebut didapatkan hasil bahwa faktor-faktor yang mempengaruhi komponen IPG di provinsi Kalimantan Timur dan Kalimantan Selatan untuk jenis kelamin laki-laki adalah kepadatan penduduk, rasio fasilitas kesehatan per 1000 penduduk, persentase penduduk yang berpendidikan diatas SLTP, dan TPT. Sedangkan faktor yang mempengaruhi IPG jenis kelamin perempuan adalah persentase penduduk yang berpendidikan diatas SLTP.

Berdasarkan uraian tersebut, maka penelitian ini ingin mengkaji interval konfidensi untuk parameter model regresi nonparametrik Spline *truncated* multivariabel. Metode tersebut diterapkan pada data Indeks Pembangunan Gender (IPG) di Provinsi Jawa Timur, karena kurva masing-masing variabel yang diduga mempengaruhi IPG dengan IPG tidak membentuk suatu pola tertentu. Sehingga, dari penelitian ini akan didapatkan variabel yang berpengaruh secara signifikan terhadap angka IPG di Jawa Timur.

## **1.2 Rumusan Masalah**

Berdasarkan uraian latar belakang tersebut, adapun rumusan masalah dalam penelitian ini adalah sebagai berikut.

1. Bagaimana kajian interval kepercayaan untuk parameter model dalam regresi nonparametrik Spline multivariabel?
2. Bagaimana penerapan interval kepercayaan untuk parameter model dalam regresi nonparametrik pada data IPG di Jawa Timur?

## **1.3 Tujuan**

Berdasarkan rumusan masalah tersebut, maka tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut.

1. Mengkaji interval kepercayaan untuk parameter model dalam regresi nonparametrik Spline multivariabel.
2. Menerapkan interval kepercayaan untuk parameter model dalam regresi nonparametrik pada data IPG di Jawa Timur.

## **1.4 Manfaat Penelitian**

Manfaat yang ingin dicapai dari penelitian ini adalah sebagai berikut

1. Dapat memberikan wawasan keilmuan yang lebih khusus kepada penulis tentang interval kepercayaan untuk parameter model dalam regresi nonparametrik Spline *truncated* multivariabel.
2. Dengan menerapkan interval kepercayaan untuk parameter model dalam regresi nonparametrik pada data IPG di Jawa Timur diharapkan dapat memberi masukan kepada pemerintah Jawa Timur untuk meningkatkan upaya pembangunan berbasis gender.

## **1.5 Batasan Masalah**

Berdasarkan rumusan masalah, maka yang menjadi batasan masalah dalam penelitian ini sebagai berikut.

1. Mendapatkan interval kepercayaan untuk parameter regresi Spline multivariabel menggunakan model Spline linier.

2. Data yang digunakan dalam penelitian ini adalah data sekunder yang diambil dari publikasi BPS tahun 2015.
3. Titik knot pada Spline linier yang digunakan adalah satu, dua, tiga, serta kombinasi titik knot.
4. Pemilihan titik knot optimal menggunakan metode GCV.

## BAB 2 TINJAUAN PUSTAKA

### 2.1 Analisis Regresi

Analisis regresi adalah salah satu metode yang digunakan untuk mengetahui hubungan antar satu variabel respon dengan satu atau lebih variabel prediktor. Persamaan regresi adalah persamaan matematik yang memungkinkan untuk meramalkan nilai-nilai suatu peubah tak bebas (variabel respon) dari nilai-nilai satu atau lebih peubah bebas/variabel prediktor (Walpole, 1995). Jika diberikan data berpasangan  $(x_{1i}, x_{2i}, \dots, x_{pi}, y_i)$ ,  $i = 1, 2, \dots, n$  maka bentuk regresi linier berganda yang terbentuk menurut Gujarati (2004) adalah:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} + \varepsilon_i \quad (2.1)$$

dengan  $y_i$  merupakan variabel respon  $(x_{1i}, x_{2i}, \dots, x_{pi})$  merupakan variabel prediktor sebanyak  $p$ , dan  $\varepsilon_i$  merupakan residual yang diasumsikan  $N(0, \sigma^2 I)$ .

### 2.2 Regresi Nonparametrik

Regresi nonparametrik merupakan suatu metode statistika yang digunakan untuk mengetahui hubungan antara variabel respon dengan variabel prediktor yang tidak diketahui bentuk fungsinya, hanya diasumsikan *smooth* (mulus). Kurva Regresi hanya diasumsikan *smooth* (mulus) dalam arti termuat dalam suatu ruang fungsi tertentu. Regresi nonparametrik merupakan regresi yang sangat fleksibel dalam memodelkan pola data. (Eubank, 1988)

Misalkan  $x$  adalah variabel prediktor dan  $y$  adalah variabel respon untuk  $n$  buah pengamatan, maka model regresi nonparametrik secara umum adalah sebagai berikut

$$y_i = f(x_i) + \varepsilon_i, i = 1, 2, \dots, n \quad (2.2)$$

dengan  $f(x_i)$  merupakan fungsi regresi nonparametrik serta  $\varepsilon_i$  merupakan *error* yang berdistribusi normal, independen dengan mean 0 dan varians  $\sigma^2$  (Mao dan Zhao, 2003).

### 2.3 Regresi Nonparametrik Spline *Truncated*

Spline merupakan potongan polinomial, yaitu polinomial yang memiliki sifat tersegmen kontinu. Spline mempunyai sifat fleksibilitas yang tinggi dan mempunyai kemampuan mengestimasi perilaku data yang cenderung berbeda pada interval yang berlainan (Eubank, 1998). Kemampuan tersebut ditunjukkan dengan fungsi *truncated* (potongan-potongan), dimana potongan-potongan tersebut merupakan titik knot. Titik perpaduan bersama dari potongan-potongan tersebut menunjukkan terjadinya perubahan pola perilaku fungsi spline pada selang yang berbeda disebut titik knot (Hardle, 1990).

Analisis regresi nonparametrik spline univariabel adalah analisis regresi nonparametrik jika terdapat satu variabel respon dan satu variabel prediktor. Apabila jika terdapat satu variabel respon dan lebih dari satu variabel prediktor maka disebut dengan regresi nonparametrik Spline multivariabel (Budiantara, 2004). Jika diberikan data  $(x_{1i}, x_{2i}, \dots, x_{pi}, y_i)$  dan hubungan antara  $(x_{1i}, x_{2i}, \dots, x_{pi})$  dan  $y_i$  mengikuti model regresi nonparametrik multivariabel

$$y_i = f(x_{1i}, x_{2i}, \dots, x_{pi}) + \varepsilon_i, i = 1, 2, \dots, n \quad (2.3)$$

dengan  $f$  adalah kurva regresi yang tidak diketahui bentuknya. Apabila kurva regresi  $f$  diasumsikan bersifat aditif dan dihipotesiskan dengan fungsi spline linier maka diperoleh model regresi

$$\begin{aligned} y_i &= f(x_{1i}) + f(x_{2i}) + \dots + f(x_{pi}) + \varepsilon_i \\ &= \sum_{j=1}^p f(x_{ji}) + \varepsilon_i, i = 1, 2, \dots, n \end{aligned} \quad (2.4)$$

dimana

$$f(x_{ji}) = \beta_0 + \beta_{j1}x_{ji} + \sum_{u=1}^r \beta_{j(1+u)}(x_{ji} - K_{ju})_+^1 \quad (2.5)$$

dengan

$$(x_{ji} - K_{ju})_+^1 = \begin{cases} (x_{ji} - K_{ju})_+^1, & x_{ji} \geq K_{ju} \\ 0 & , x_{ji} < K_{ju} \end{cases}$$

dengan  $K_{j1}, K_{j2}, \dots, K_{jr}$  adalah titik- titik knot yang memperlihatkan pola perubahan perilaku dari fungsi pada sub-sub interval yang berbeda.



## 2.4 Pemilihan Titik Knot Optimal

Salah satu langkah penting dalam pendekatan Spline adalah memilih titik knot yang optimal. Titik knot merupakan titik perpaduan bersama dimana terdapat perubahan perilaku fungsi pada interval yang berlainan (Budiantara, 2006). Pemilihan titik knot optimal dilakukan untuk menentukan model terbaik yang terbentuk. Salah satu metode yang digunakan untuk memilih titik knot optimal adalah dengan menggunakan metode GCV (*Generalized Cross Validation*) (Budiantara, 2000). Model Spline yang terbaik didapat dari nilai GCV terkecil.

$$GCV(K_{11}, K_{12}, \dots, K_{1r}, \dots, K_{p1}, \dots, K_{pr}) = \frac{MSE(K_{11}, K_{12}, \dots, K_{1r}, \dots, K_{p1}, \dots, K_{pr})}{\left(n^{-1} \text{tr} \left[ I - A(K_{11}, K_{12}, \dots, K_{1r}, \dots, K_{p1}, \dots, K_{pr}) \right] \right)^2} \quad (2.6)$$

dimana  $MSE(K_{11}, K_{12}, \dots, K_{1r}, \dots, K_{p1}, \dots, K_{pr}) = n^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2$ ,

$K_{11}, K_{12}, \dots, K_{1r}, \dots, K_{p1}, \dots, K_{pr}$

adalah titik knot, dan matriks  $A(K_{11}, K_{12}, \dots, K_{1r}, \dots, K_{p1}, \dots, K_{pr}) = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$  (Budiantara, Ratna, Zain, dan Wibowo, 2012).

## 2.5 Interval Konfidensi Untuk Parameter Regresi Parametrik

Diberikan model regresi linier berganda 2.1 untuk mendapatkan interval konfidensi untuk koefisien regresi  $\beta_j$ , diasumsikan *error*  $\varepsilon_i$  berdistribusi normal independen dengan mean nol dan varians  $\sigma^2$ . Oleh karena itu, observasi  $y_i$  berdistribusi normal independen dengan mean  $\beta_0 + \sum_{j=1}^p \beta_j x_{ji}$  dan varians  $\sigma^2$

$$E(y_i) = E(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} + \varepsilon_i)$$

$$= \beta_0 + \sum_{j=1}^p \beta_j x_{ji} + E(\varepsilon_i)$$

$$= \beta_0 + \sum_{j=1}^p \beta_j x_{ji}$$

$$\text{Var}(y_i) = \text{Var}(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} + \varepsilon_i)$$

$$= \sigma^2$$

Apabila persamaan (2.1) dinyatakan dalam bentuk matriks  $\underline{y} = \mathbf{X}\underline{\beta} + \underline{\varepsilon}$ ,

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{21} & \dots & x_{p1} \\ 1 & x_{12} & x_{22} & \dots & x_{p2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} & \dots & x_{pn} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

dengan

$$\tilde{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{21} & \dots & x_{p1} \\ 1 & x_{12} & x_{22} & \dots & x_{p2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} & \dots & x_{pn} \end{pmatrix}, \tilde{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}, \tilde{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Salah satu metode yang digunakan untuk mendapatkan  $\hat{\tilde{\beta}}$  adalah dengan menggunakan metode MLE. Telah disebutkan bahwa diasumsikan *error*  $\varepsilon$  berdistribusi normal independen dengan mean  $\mathbf{0}$  dan varians  $\sigma^2\mathbf{I}$ , sehingga didapatkan distribusi probabilitas dari  $\varepsilon$  adalah

$$h(\tilde{\varepsilon}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} \tilde{\varepsilon}'\tilde{\varepsilon}\right)$$

$$h(\tilde{y}, \tilde{\beta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} (\tilde{y} - \mathbf{X}\tilde{\beta})' (\tilde{y} - \mathbf{X}\tilde{\beta})\right) \quad (2.7)$$

Berdasarkan persamaan (4.6) diperoleh fungsi likelihood

$$L(\tilde{\beta}) = \prod_{i=1}^n h(y_i, \tilde{\beta})$$

$$= \left(\sqrt{2\pi\sigma^2}\right)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} (\tilde{y} - \mathbf{X}\tilde{\beta})' (\tilde{y} - \mathbf{X}\tilde{\beta})\right) \quad (2.8)$$

Apabila persamaan (2.8) dibuat transformasi logaritma, akan didapatkan

$$\ell(\tilde{\beta}) = \log L(\tilde{\beta})$$

$$= \log\left(\left(\sqrt{2\pi\sigma^2}\right)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} (\tilde{y} - \mathbf{X}\tilde{\beta})' (\tilde{y} - \mathbf{X}\tilde{\beta})\right)\right)$$

$$= -\frac{n}{2} \log\left(\sqrt{2\pi\sigma^2}\right) - \frac{1}{2\sigma^2} (\tilde{y} - \mathbf{X}\tilde{\beta})' (\tilde{y} - \mathbf{X}\tilde{\beta}) \quad (2.9)$$

Dengan menggunakan derivatif parsial terhadap  $\tilde{\beta}$  diperoleh:

$$\begin{aligned}\frac{\partial \ell(\underline{\beta})}{\partial \underline{\beta}} &= \frac{\partial \left( -\frac{n}{2} \log(\sqrt{2\pi}\sigma^2) - \frac{1}{2\sigma^2} (\underline{y} - \mathbf{X}\underline{\beta})' (\underline{y} - \mathbf{X}\underline{\beta}) \right)}{\partial \underline{\beta}} \\ &= \frac{\partial \left( -\frac{n}{2} \log(\sqrt{2\pi}\sigma^2) - \frac{1}{2\sigma^2} \underline{y}' \underline{y} - 2\underline{\beta}' \mathbf{X}' \underline{y} + \underline{\beta}' \mathbf{X}' \mathbf{X} \underline{\beta} \right)}{\partial \underline{\beta}}\end{aligned}$$

$$0 = -\frac{1}{2\sigma^2} (-2\mathbf{X}' \underline{y} + 2\mathbf{X}' \mathbf{X} \hat{\underline{\beta}})$$

$$\hat{\underline{\beta}} = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \underline{y}$$

Karena estimator least-square  $\hat{\underline{\beta}}$  merupakan kombinasi linier dari observasi, maka  $\hat{\underline{\beta}}$  berdistribusi normal dengan vektor mean  $\underline{\beta}$  dan matriks covarians  $\sigma^2(\mathbf{X}' \mathbf{X})^{-1}$ .

$$\begin{aligned}E(\hat{\underline{\beta}}) &= E((\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \underline{y}) \\ &= (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' E(\underline{y}) \\ &= (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{X} \underline{\beta} \\ &= \underline{\beta}\end{aligned}$$

$$\begin{aligned}Var(\hat{\underline{\beta}}) &= Var((\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \underline{y}) \\ &= (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' Var(\underline{y}) (\mathbf{X}' \mathbf{X})^{-1} \\ &= (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \sigma^2 \mathbf{I} \mathbf{X}' (\mathbf{X}' \mathbf{X})^{-1} \\ &= \sigma^2 (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{X}' (\mathbf{X}' \mathbf{X})^{-1} \\ &= \sigma^2 (\mathbf{X}' \mathbf{X})^{-1}\end{aligned}$$

Hal ini mengimplikasikan bahwa distribusi marginal dari setiap koefisien regresi  $\hat{\beta}_j$  adalah normal dengan mean  $\beta_j$  dan varians  $\sigma^2 C_{jj}$ , dimana  $C_{jj}$  merupakan elemen diagonal ke- $j$  dari matriks  $(\mathbf{X}' \mathbf{X})^{-1}$ . Akibatnya, statistik

$$t_j = \frac{\hat{\beta}_j - \beta_j}{\sqrt{\sigma^2 C_{jj}}}, j = 0, 1, \dots, p \quad (2.10)$$

berdistribusi  $t$  dengan derajat bebas  $n-p$ , dimana  $\sigma^2$  merupakan estimasi dari varians *error*. Berdasarkan persamaan (2.7), maka dapat didefinisikan interval konfidensi  $(1 - \alpha) \times 100\%$  untuk koefisien regresi  $\beta_j, j = 1, 2, \dots, p$  adalah sebagai berikut

$$\hat{\beta}_j - t_{\frac{\alpha}{2}, n-p} \sqrt{\sigma^2 C_{jj}} \leq \beta_j \leq \hat{\beta}_j + t_{\frac{\alpha}{2}, n-p} \sqrt{\sigma^2 C_{jj}} \quad (2.11)$$

(Montgomery, Peck, dan Vining, 2012)

## 2.6 Pemeriksaan Asumsi Residual

Pemeriksaan asumsi residual yang dilakukan meliputi asumsi independensi residual, asumsi residual identik, dan asumsi normalitas residual.

### 2.6.1 Asumsi Independensi Residual.

Pemeriksaan asumsi independensi residual digunakan untuk mendeteksi apakah ada ketergantungan (korelasi) antara residual. Asumsi independen pada residual ditunjukkan oleh nilai kovarian antara  $\varepsilon_i$  dan  $\varepsilon_j$  sama dengan nol. Untuk menguji asumsi ini dapat dilihat dari plot berada pada batas daerah signifikan yaitu  $\pm Z_{\alpha/2} / \sqrt{n}$  maka dikatakan tidak terdapat kasus autokorelasi (Intansari, 2016).

### 2.6.2 Asumsi Residual Identik

Pemeriksaan asumsi identik digunakan untuk melihat homogenitas dari varians residual. Uji Glejser adalah satu metode yang dapat digunakan untuk mendeteksi adanya heterogenitas varians residual (Setiawan & Kusriani, 2010). Hipotesis yang digunakan adalah

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 = \sigma^2 \text{ (residual identik)}$$

$$H_1: \text{minimal ada satu } \sigma_i^2 \neq \sigma^2, i = 1, 2, \dots, n \text{ (residual tidak identik)}$$

dengan statistik uji

$$F_{hitung} = \frac{\sum_{i=1}^n (|\hat{e}_i| - |\bar{e}_i|)^2 / k - 1}{\sum_{i=1}^n (|e_i| - |\hat{e}_i|)^2 / n - k} \quad (2.12)$$

Daerah penolakan  $H_0$  yaitu tolak  $H_0$  jika  $F_{hitung} > F_{tabel}(F_{\alpha; (k-1, n-k)})$

### 2.6.3 Asumsi Normalitas Residual

Pengujian asumsi normalitas residual dilakukan untuk memeriksa apakah residual mengikuti distribusi normal atau tidak. Hipotesis yang digunakan untuk pengujian residual adalah sebagai berikut.

$H_0$ : residual berdistribusi normal

$H_1$ : residual tidak berdistribusi normal

dengan statistik uji yang digunakan adalah *Kolmogorov Smirnov*

$$z_{hitung} = \text{Sup}_x |F_n(x) - F_0(x)| \quad (2.13)$$

Daerah penolakan  $H_0$  yaitu tolak  $H_0$  jika  $z_{hitung} > z_\alpha$ .

## 2.7 Koefisien Determinasi $R^2$

Salah satu tujuan analisis regresi adalah mendapatkan model terbaik yang mampu menjelaskan hubungan antara variabel prediktor dan variabel respon. Kriteria yang dapat digunakan dalam pemilihan model terbaik salah satunya adalah dengan menggunakan koefisien determinasi  $R^2$ . Secara umum semakin besar nilai  $R^2$ , maka semakin baik pula model yang didapatkan. Koefisien determinasi didefinisikan sebagai berikut :

$$R^2 = \frac{SSR}{SST} \quad (2.11)$$

dimana  $SSR = (\hat{y} - \bar{y})' (\hat{y} - \bar{y})$  dan  $SST = (y - \bar{y})' (y - \bar{y})$ . Besaran nilai  $R^2$  tidak pernah negatif dan batasannya adalah  $0 \leq R^2 \leq 1$  (Gujarati dan Porter, 2015).

## 2.8 Pivotal Quantity

Dalam statistik, sebuah *pivotal quantity* atau *pivot* merupakan fungsi dari pengamatan dan parameter tidak teramati sehingga distribusi fungsi probabilitas tidak tergantung pada parameter yang tidak diketahui (termasuk parameter *nuisance*). Sebuah *pivot quantity* tidak perlu menjadi statistik-fungsi dan nilainya dapat bergantung pada parameter model, namun distribusinya tidak boleh. Jika itu adalah statistik, maka itu dikenal sebagai statistik tambahan.

Secara lebih formal, misalkan  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  adalah sebuah sampel random dari distribusi yang bergantung pada parameter (atau vektor parameter)  $\theta$ . Misalkan  $g(\mathbf{X}, \theta)$  adalah variabel random yang distribusinya sama untuk semua  $\theta$ . Maka  $g$  disebut sebagai *pivotal quantity* (atau *pivot*).

*Pivotal quantity* biasanya digunakan untuk normalisasi agar data dari kumpulan data yang berbeda dapat dibandingkan. *Pivotal quantity* sangat penting

untuk kontruksi statistik uji, karena memungkinkan statistik untuk tidak bergantung pada parameter-misalnya statistic *t-student* adalah untuk distribusi normal dengan varians yang tidak yang tidak diketahui (dan mean). *Pivotal quantity* juga menyediakan satu metode untuk membangun interval konfidensi. Dalam bentuk statistik tambahan, *pivotal quantity* dapat digunakan untuk membangun interval prediksi *frequentist*/interval kepercayaan prediktif (Shao, 2003).

## 2.9 Indeks Pembangunan Gender (IPG)

Tingkat keberhasilan capaian pembangunan yang sudah mengakomodasi persoalan gender saat ini telah dapat diukur, salah satunya adalah dengan IPG (Indeks Pembangunan Gender). Indeks Pembangunan Gender (IPG) diperkenalkan pertama kali oleh UNDP pada tahun 1995, lima tahun setelah UNDP memperkenalkan Indeks Pembangunan Manusia (IPM). Pada metode lama, IPG tidak mengukur langsung ketimpangan antargender yang terjadi, namun hanya disparitas dari masing-masing komponen IPM untuk setiap jenis kelamin. Selain itu, angka IPG pada metode lama tidak bisa diinterpretasikan terpisah dari IPM.

Pada tahun 2014, UNDP kembali melakukan penghitungan IPG dengan menggunakan metode baru. IPG metode baru ini merupakan pengukuran langsung terhadap ketimpangan antargender dalam pencapaian IPM. Pada metode baru ini digunakan rasio IPM perempuan dengan IPM laki-laki, sehingga bisa terlihat pencapaian pembangunan manusia antara perempuan dengan laki-laki

$$IPG = \frac{IPM_P}{IPM_L}$$

Angka ini menunjukkan rasio antara pembangunan perempuan dan pembangunan laki-laki. Ketika angka IPG semakin mendekati angka 100, maka pembangunan gender semakin seimbang atau merata. Namun jika semakin menjauhi angka 100, maka pembangunan gender makin timpang antar jenis kelamin. Dalam metode baru ini, dimensi yang digunakan masih sama dengan metode lama yaitu umur panjang dan hidup sehat, pengetahuan, serta standar hidup layak (BPS, 2015).

## **2.10 Faktor- faktor Yang Mempengaruhi Indeks Pembangunan Gender (IPG)**

Indeks Pembangunan Gender (IPG) merupakan salah satu alat untuk mengukur tingkat keberhasilan capaian pembangunan yang sudah mengakomodasi persoalan gender. Dalam publikasi BPS (2014) menyebutkan bahwa IPG merupakan pengukuran langsung terhadap ketimpangan antargender dalam pencapaian IPM. Oleh karena itu strategi untuk meningkatkan angka IPG sama dengan meningkatkan angka Indeks Pembangunan Manusia (IPM), tetapi lebih menekankan secara khusus terhadap komposisi jenis kelamin. Dimensi yang digunakan untuk pendekatan dalam mengukur kualitas hidup dalam IPG adalah umur panjang dan hidup sehat, pengetahuan, serta standar hidup layak (BPS, 2015).

Beberapa penelitian yang membahas mengenai pembangunan gender diantaranya adalah Hafizh (2013) meneliti tentang disparitas gender di Jawa Timur dengan menggunakan pendekatan model regresi probit ordinal menyimpulkan bahwa faktor-faktor yang mempengaruhi disparitas gender adalah angka partisipasi sekolah (APS) tingkat SMP penduduk perempuan, persentase penduduk perempuan dengan pendidikan terakhir yang ditamatkan setingkat SMP, dan persentase penduduk perempuan yang bekerja di sektor formal. Kemudian Hakim (2013) meneliti tentang komponen Indeks Pembangunan Gender di Provinsi Kalimantan Timur dan Kalimantan Selatan tahun 2011. Dari penelitian tersebut didapatkan hasil bahwa faktor- faktor yang mempengaruhi komponen IPG di provinsi Kalimantan Timur dan Kalimantan Selatan untuk jenis kelamin laki- laki adalah kepadatan penduduk, rasio fasilitas kesehatan per 1000 penduduk, persentase penduduk yang berpendidikan diatas Sekolah Lanjutan Tingkat Pertama (SLTP), dan Tingkat Pengangguran Terbuka (TPT). Sedangkan faktor yang mempengaruhi IPG jenis kelamin perempuan adalah persentase penduduk yang berpendidikan diatas SLTP.

Membangun kesetaraan dan keadilan gender tidak dapat dilakukan dalam waktu yang relatif singkat. Kesetaraan dan keadilan gender pada hakikatnya merujuk pada tidak adanya perbedaan hak dan kewajiban antara perempuan dan laki- laki yang dijamin oleh perundang- undangan yang dihasilkan oleh negara maupun lingkungan bermasyarakat. Jaminan tidak adanya perbedaan dalam status dan kedudukan perempuan dan laki-laki dalam berbangsa dan bernegara meliputi partisipasi dalam program pembangunan terutama dalam peningkatan kualitas

hidup melalui program peningkatan kapabilitas atau kemampuan dasar. Program tersebut mencakup berbagai layanan dasar kesehatan, pendidikan, dan kemudahan akses ekonomi yang diberikan pemerintah.

Dalam publikasi BPS Laporan kesehatan Provinsi Jawa Timur 2015 menyebutkan bahwa tingkat kesehatan masyarakat merupakan salah satu indikator untuk melihat tingkat kesejahteraan umum masyarakat pada suatu wilayah. Semakin baik/tinggi tingkat kesehatan masyarakat, maka dapat dikatakan semakin baik pula tingkat kesejahteraan masyarakat tersebut. Hakim (2014) menyebutkan bahwa semakin banyak fasilitas kesehatan yang tersedia maka semakin besar peluang masyarakat untuk berobat ke tenaga medis yang akan berdampak terhadap peningkatan kesehatan.

Berdasarkan teori L. Blum (Notoadmojo, 2007) dalam publikasi BPS (2016) menyatakan bahwa derajat kesehatan penduduk dapat diukur dari angka kematian (mortalitas) dan angka kesakitan (morbiditas). Angka kesakitan adalah persentase penduduk yang mempunyai keluhan kesehatan yang menyebabkan terganggunya kegiatan sehari-hari. Angka ini merupakan salah satu indikator yang menentukan derajat kesehatan penduduk (BPS, 2015).

Pendidikan merupakan salah satu cara untuk meningkatkan kualitas SDM. Pendidikan juga merupakan salah satu aspek yang dapat digunakan untuk melihat kesetaraan gender (BPS, 2014). Untuk meningkatkan kapabilitas dasar penduduk di bidang pendidikan, ada berbagai macam upaya yang dapat dilakukan, misalnya dengan memperluas cakupan pendidikan formal. Pendidikan dasar merupakan hak setiap warga negara, sehingga pemerintah perlu menjamin bahwa warga negaranya minimal menikmati pendidikan dasar hingga 9 tahun. Angka partisipasi Sekolah (APS) dapat digunakan untuk mengukur pemerataan akses terhadap pendidikan (BPS, 2015).

Hakim (2014) dalam penelitiannya menyebutkan bahwa tingkat pendidikan berpengaruh terhadap kesehatan. Tingkat pendidikan yang tinggi akan membuat pola pikir seseorang untuk menjaga kesehatannya semakin tinggi. Selain itu, semakin tinggi pendidikan seseorang akan membuat seseorang lebih mudah mencari pekerjaan karena lapangan pekerjaan yang tersedia semakin banyak. Salah satu indikator untuk mengukur tingkat kesejahteraan rakyat adalah indikator



ketenagakerjaan. Dua indikator ketenagakerjaan yang kerap dikenal masyarakat luas adalah indikator Tingkat Partisipasi Angkatan Kerja (TPAK) dan Tingkat Pengangguran Terbuka (TPT) (Aryanto, 2014). TPAK dapat didefinisikan sebagai perbandingan antara angkatan kerja dengan penduduk usia kerja (biasanya berusia 15 tahun keatas). Angkatan kerja mencakup penduduk yang bekerja dan yang mencari pekerjaan. Sedangkan TPT adalah perbandingan antara penduduk yang mencari pekerjaan dengan angkatan kerja (BPS, 2015).

*(halaman ini sengaja dikosongkan)*

## BAB 3 METODOLOGI PENELITIAN

### 3.1 Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari publikasi Badan Pusat Statistik (BPS) provinsi Jawa Timur. Publikasi yang dipakai dalam penelitian ini antara lain:

1. Pembangunan Manusia Berbasis Gender Provinsi Jawa Timur Tahun 2015
2. Profil Gender Provinsi Jawa Timur Tahun 2015
3. Laporan Eksekutif Kesehatan Provinsi Jawa Timur 2015
4. Hasil Survey Sosial Ekonomi Nasional Tahun 2014 Provinsi Jawa Timur

Unit observasi dalam penelitian ini meliputi 29 kabupaten dan 9 kota di provinsi Jawa Timur tahun 2014.

### 3.2 Variabel Penelitian

Variabel penelitian dalam penelitian ini terdiri dari satu variabel respon ( $y$ ) dan variabel prediktor ( $x$ ) terdiri dari 4 variabel dengan rincian pada Tabel 3.1.

**Tabel 3.1** Variabel Penelitian

Variabel	Keterangan	Skala Data
$y$	Indeks Pembangunan Gender	Ratio
$x_1$	Angka Partisipasi Sekolah Sekolah Lanjutan Tingkat Atas (SLTA) Penduduk Perempuan	Ratio
$x_2$	Angka Kesakitan Penduduk Perempuan	Ratio
$x_3$	Penolong Terakhir Kelahiran Oleh Medis	Ratio
$x_4$	Tingkat Partisipasi Angkatan Kerja (TPAK) Penduduk Perempuan	Ratio

Adapun penjelasan dari variabel- variabel dalam penelitian ini adalah sebagai berikut.

a. Variabel Respon

Dalam penelitian ini variabel respon yang digunakan adalah IPG di Provinsi Jawa Timur tahun 2014. IPG menunjukkan rasio antara pembangunan perempuan terhadap pembangunan laki-laki, dan angka IPG dalam penelitian ini berupa persentase.

b. Variabel Prediktor

Variabel prediktor dalam penelitian ini terdiri dari 4 variabel yang diduga akan mempengaruhi angka IPG. Keempat variabel prediktor tersebut adalah sebagai berikut:

1. Angka Partisipasi Sekolah (APS) SLTA Penduduk Perempuan merupakan angka yang menyatakan APS khusus untuk jenis kelamin perempuan. Angka Partisipasi Sekolah (APS) Sekolah Lanjutan Tingkat Atas (SLTA) adalah angka yang mengukur proporsi anak yang bersekolah pada kelompok umur 16-18 tahun pada jenjang pendidikan SLTA (BPS, 2015).
2. Angka Kesakitan penduduk perempuan merupakan angka yang menyatakan angka kesakitan untuk jenis kelamin perempuan. Angka kesakitan adalah persentase penduduk yang mempunyai keluhan kesehatan yang menyebabkan terganggunya kegiatan sehari-hari (BPS, 2015).
3. Penolong Terakhir Kelahiran Oleh Medis adalah persentase penolong persalinan yang menangani proses kelahiran bayi hingga pemotongan tali pusar oleh tenaga medis: dokter, bidan, dan tenaga medis lain (BPS, 2015).
4. Tingkat Partisipasi Angkatan Kerja (TPAK) penduduk perempuan merupakan angka yang menyatakan TPAK untuk jenis kelamin perempuan. TPAK didefinisikan sebagai perbandingan antara angkatan kerja dengan penduduk usia kerja (BPS, 2015).

### **3.3 Struktur Data Penelitian**

Dalam penelitian ini unit observasi yang digunakan adalah sebanyak 38 kabupaten/kota di Provinsi Jawa Timur dan banyaknya variabel prediktor terdiri

dari 4 variabel. Sehingga, struktur data penelitian yang digunakan adalah sebagai berikut.

**Tabel 3.2** Struktur Data Penelitian

Kabupaten/ Kota	$y$	$x_1$	$x_2$	$x_3$	$x_4$
<b>1</b>	$y_1$	$x_{11}$	$x_{21}$	$x_{31}$	$x_{41}$
<b>2</b>	$y_2$	$x_{12}$	$x_{22}$	$x_{32}$	$x_{42}$
<b>3</b>	$y_3$	$x_{13}$	$x_{23}$	$x_{33}$	$x_{43}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
<b>38</b>	$y_{38}$	$x_{1(38)}$	$x_{2(38)}$	$x_{3(38)}$	$x_{4(38)}$

### 3.4 Tahapan Penelitian

Langkah- langkah yang dilakukan untuk menjawab tujuan dalam penelitian ini adalah sebagai berikut:

1. Mendapatkan interval konfidensi untuk parameter model regresi nonparametrik spline *truncated* multivariabel dengan langkah- langkah:

a. Diberikan data  $(x_{1i}, x_{2i}, \dots, x_{pi}, y_i)$ ,  $i = 1, 2, \dots, n$  dan  $p$  menunjukkan banyaknya variabel prediktor, mengikuti model regresi nonparametrik multivariabel

$$y_i = f(x_{1i}, x_{2i}, \dots, x_{pi}) + \varepsilon_i, i = 1, 2, \dots, n$$

b. Model regresi nonparametrik multivariabel (a) diasumsikan bersifat aditif

$$f(x_{1i}, x_{2i}, \dots, x_{pi}) = \sum_{j=1}^p f_j(x_{ji}), i = 1, 2, \dots, n$$

c. Menghampiri kurva regresi  $f_j(x_{ji})$  dengan menggunakan model Spline linier dengan  $r$  knot:

$$f_j(x_{ji}) = \beta_0 + \beta_{j1}x_{ji} + \sum_{u=1}^r \beta_{j(1+u)}(x_{ji} - K_{ju})_+^1$$

d. Membentuk model nonparametrik Spline linier dengan  $r$  knot dengan model

$$y_i = \beta_0 + \sum_{j=1}^p \left( \beta_{j1}x_{ji} + \sum_{u=1}^r \beta_{j(1+u)}(x_{ji} - K_{ju})_+^1 \right) + \varepsilon_i, i = 1, 2, \dots, n$$

- e. Menyajikan model regresi pada (d) dalam bentuk matriks  $\underline{y} = \mathbf{X}(\mathbf{K})\underline{\beta} + \underline{\varepsilon}$  dengan  $\underline{\varepsilon} \sim N(0, \sigma^2 \mathbf{I})$ .
- f. Mencari estimasi untuk parameter  $\underline{\beta}$  menggunakan metode MLE dengan langkah- langkah sebagai berikut:
- i. Membentuk fungsi likelihood  $L(\underline{\beta})$ .
  - ii. Membuat transformasi  $l(\underline{\beta}) = \log(L(\underline{\beta}))$ .
  - iii. Dengan menggunakan derivatif, menurunkan parsial  $l(\underline{\beta})$  terhadap  $\underline{\beta}$  dan disamakan dengan 0, sehingga didapatkan  $\hat{\underline{\beta}}$ .
- g. Mencari distribusi dari  $\hat{\underline{\beta}}$ .
- h. Setelah didapatkan distribusi dari  $\hat{\underline{\beta}}$  kemudian dicari pivotal quantity untuk  $\underline{\beta}$ .
- i. Misalkan diberikan  $T_{ju}(x_1, x_2, \dots, x_n, y)$  adalah pivotal quantity untuk  $\beta_{ju}$  dengan  $j = 1, 2, \dots, p$  dan  $u = 1, 2, \dots, 1 + r$ .
- j. Menyelesaikan persamaan dalam probabilitas  $P(u \leq T_{ju}(x_1, x_2, \dots, x_n) \leq v) = 1 - \alpha$
- k. Menghitung panjang interval konfidensi  $1 - \alpha = \ell(u, v)$
- l. Membentuk fungsi lagrange  $F(u, v, \lambda) = \ell(u, v) + \lambda(\Phi(b) - \Phi(a)) - (1 - \alpha)$  dengan:  
 $\Phi$  : fungsi distribusi kumulatif  $N(0,1)$   
 $\lambda$  : konstanta lagrange
- m. Menghitung derivatif parsial:
- $$\frac{\partial F(u, v, \lambda)}{\partial u} = 0$$
- $$\frac{\partial F(u, v, \lambda)}{\partial v} = 0$$
- $$\frac{\partial F(u, v, \lambda)}{\partial \lambda} = 0$$

- n. Menentukan nilai  $(u, v, \lambda)$ .  
Menghitung interval konfidensi terpendek untuk  $\beta_{ju}$  dengan menyelesaikan persamaan  $P(u \leq T_{ju}(x_1, x_2, \dots, x_n) \leq v) = 1 - \alpha$
  - o. Dari langkah (n) didapatkan CI  $(1 - \alpha) \times 100\%$  untuk parameter  $\beta_{ju}$ .
2. Menerapkan interval konfidensi untuk parameter regresi nonparametrik Spline *truncated* multivariabel pada data IPG di Jawa Timur dengan langkah-langkah:
    - a. Memodelkan IPG dengan variabel yang diduga mempengaruhinya dengan langkah-langkah
      - i. Membuat *scatterplot* masing-masing variabel prediktor dengan variabel respon untuk mengetahui pola hubungannya.
      - ii. Menentukan lokasi titik knot dengan cara mencoba-coba pada tempat yang terlihat adanya perubahan perilaku data pada selang tertentu. Titik knot yang digunakan adalah satu titik knot, dua titik knot, tiga titik knot, dan kombinasi titik knot.
    - b. Menentukan model terbaik yang terbentuk dengan langkah-langkah
      - i. Menghitung nilai GCV.
      - ii. Memilih titik knot optimal yang dilihat dari nilai GCV paling minimum.
    - c. Menghitung nilai koefisien determinasi  $R^2$ .
    - d. Melakukan pengujian asumsi residual meliputi beberapa uji berikut
      - i. Uji independensi dengan menggunakan uji Durbin Watson.
      - ii. Uji homogenitas dengan menggunakan uji Glejser.
      - iii. Uji normalitas dengan menggunakan uji kolmogorov-smirnov.
    - e. Mendapatkan interval konfidensi untuk parameter model.
    - f. Mengambil kesimpulan yaitu menentukan variabel prediktor yang berpengaruh secara signifikan terhadap variabel respon.

*(halaman ini sengaja dikosongkan)*



## BAB 4

### HASIL DAN PEMBAHASAN

Pada sub bab ini akan dibahas tentang estimasi parameter regresi dan interval konfidensi untuk parameter regresi nonparametrik spline *truncated* multivariabel. Kemudian memodelkan data IPG di provinsi Jawa Timur dengan menggunakan model regresi nonparametrik spline *truncated* multivariabel dengan fungsi Spline linier satu knot, dua knot, tiga knot, dan kombinasi knot. Model terbaik ditentukan dengan menggunakan nilai GCV minimum. Selanjutnya diterapkan interval konfidensi untuk parameter model regresi nonparametrik spline *truncated* multivariabel pada data IPG di provinsi Jawa Timur.

#### 4.1 Interval Konfidensi Untuk Parameter Regresi Nonparametrik Spline *Truncated* Multivariabel

Salah satu bagian terpenting dari inferensi statistik adalah interval konfidensi. Untuk mendapatkan interval konfidensi umumnya digunakan pendekatan Bayesian (Wahba, 1983 ; 1990 dan Wang, 1998). Dalam penelitian ini menggunakan pendekatan *Pivotal Quantity*. *Pivotal Quantity* adalah sebuah statistik yang distribusinya tidak memuat parameter. Langkah pertama yang dilakukan untuk mendapatkan interval konfidensi untuk parameter regresi nonparametrik spline *truncated* multivariabel adalah mencari distribusi dari  $\hat{\beta}$ .

Jika diberikan data berpasangan  $(x_{1i}, x_{2i}, \dots, x_{pi}, y_i)$ ,  $i = 1, 2, \dots, n$  dan  $p$  menunjukkan banyaknya variabel prediktor, mengikuti model regresi nonparametrik multivariabel pada persamaan 2.3

$$y_i = f(x_{1i}, x_{2i}, \dots, x_{pi}) + \varepsilon_i, i = 1, 2, \dots, n$$

Apabila model pada persamaan (2.3) diasumsikan bersifat aditif maka didapatkan

$$f(x_{1i}, x_{2i}, \dots, x_{pi}) = \sum_{j=1}^p f_j(x_{ji}), i = 1, 2, \dots, n \quad (4.1)$$

Jika kurva regresi  $f_j(x_{ji})$  dihipotesiskan dengan menggunakan model Spline linier dengan  $r$  knot

$$f_j(x_{ji}) = \beta_0 + \beta_{j1}x_{ji} + \sum_{u=1}^r \beta_{j(1+u)}(x_{ji} - K_{ju})_+^1 \quad (4.2)$$

Sehingga didapatkan model regresi nonparametrik spline linier multivariabel dengan r knot

$$y_i = \beta_0 + \sum_{j=1}^p \left( \beta_{j1}x_{ji} + \sum_{u=1}^r \beta_{j(1+u)}(x_{ji} - K_{ju})_+^1 \right) + \varepsilon_i, i = 1, 2, \dots, n \quad (4.3)$$

Apabila persamaan (4.3) dinyatakan dalam bentuk matriks  $\underline{y} = \mathbf{X}(\mathbf{K})\underline{\beta} + \underline{\varepsilon}$ ,

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{21} & \dots & x_{p1} & (x_{11} - K_{11})_+^1 & \dots & (x_{11} - K_{1r})_+^1 & \dots & (x_{p1} - K_{11})_+^1 & \dots & (x_{p1} - K_{pr})_+^1 \\ 1 & x_{12} & x_{22} & \dots & x_{p2} & (x_{12} - K_{11})_+^1 & \dots & (x_{12} - K_{1r})_+^1 & \dots & (x_{p2} - K_{11})_+^1 & \dots & (x_{p2} - K_{pr})_+^1 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \dots & x_{pn} & (x_{1n} - K_{11})_+^1 & \dots & (x_{1n} - K_{1r})_+^1 & \dots & (x_{pn} - K_{11})_+^1 & \dots & (x_{pn} - K_{pr})_+^1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_{11} \\ \beta_{21} \\ \vdots \\ \beta_{p1} \\ \beta_{12} \\ \beta_{13} \\ \vdots \\ \beta_{1(1+r)} \\ \vdots \\ \beta_{p2} \\ \beta_{p3} \\ \vdots \\ \beta_{p(1+r)} \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

dengan

$$\underline{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix},$$

$$\mathbf{X}(\mathbf{K}) = \begin{pmatrix} 1 & x_{11} & x_{21} & \dots & x_{p1} & (x_{11} - K_{11})_+^1 & \dots & (x_{11} - K_{1r})_+^1 & \dots & (x_{p1} - K_{11})_+^1 & \dots & (x_{p1} - K_{pr})_+^1 \\ 1 & x_{12} & x_{22} & \dots & x_{p2} & (x_{12} - K_{11})_+^1 & \dots & (x_{12} - K_{1r})_+^1 & \dots & (x_{p2} - K_{11})_+^1 & \dots & (x_{p2} - K_{pr})_+^1 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \dots & x_{pn} & (x_{1n} - K_{11})_+^1 & \dots & (x_{1n} - K_{1r})_+^1 & \dots & (x_{pn} - K_{11})_+^1 & \dots & (x_{pn} - K_{pr})_+^1 \end{pmatrix},$$

$$\underline{\beta}' = (\beta_0 \quad \beta_{11} \quad \beta_{21} \quad \dots \quad \beta_{p1} \quad \beta_{12} \quad \beta_{13} \quad \dots \quad \beta_{1(1+r)} \quad \dots \quad \beta_{p2} \quad \beta_{p3} \quad \dots \quad \beta_{p(1+r)}) \quad ,$$

$$\underline{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Apabila diasumsikan  $\underline{\varepsilon} \sim N(0, \sigma^2 \mathbf{I})$ , karena  $\underline{y}$  merupakan kombinasi linier dari  $\underline{\varepsilon}$  maka  $\underline{y}$  juga berdistribusi normal  $\underline{y} \sim N(E(\underline{y}), Var(\underline{y}))$

$$\begin{aligned} E(\underline{y}) &= E(\mathbf{X}(\mathbf{K})\underline{\beta} + \underline{\varepsilon}) \\ &= \mathbf{X}(\mathbf{K})\underline{\beta} + E(\underline{\varepsilon}) \\ &= \mathbf{X}(\mathbf{K})\underline{\beta} \end{aligned} \quad (4.4)$$

$$\begin{aligned} Var(\underline{y}) &= Var(\mathbf{X}(\mathbf{K})\underline{\beta} + \underline{\varepsilon}) \\ &= 0 + Var(\underline{\varepsilon}) \\ &= \sigma^2 \mathbf{I} \end{aligned} \quad (4.5)$$

Berdasarkan (4.4) dan (4.5) didapatkan  $\underline{y}$  berdistribusi normal dengan mean  $\mathbf{X}(\mathbf{K})\underline{\beta}$  dan varians  $\sigma^2 \mathbf{I}$ . Salah satu metode yang dapat digunakan untuk mendapatkan estimasi titik dari  $\underline{\beta}$  adalah dengan menggunakan metode MLE.

Distribusi probabilitas dari  $\underline{\varepsilon}$  adalah

$$\begin{aligned} g(\underline{\varepsilon}) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{1}{2\sigma^2} \underline{\varepsilon}'\underline{\varepsilon}\right) \\ g(\underline{y}, \underline{\beta}) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{1}{2\sigma^2} (\underline{y} - \mathbf{X}(\mathbf{K})\underline{\beta})' (\underline{y} - \mathbf{X}(\mathbf{K})\underline{\beta})\right) \end{aligned} \quad (4.6)$$

Berdasarkan persamaan (4.6) diperoleh fungsi likelihood

$$\begin{aligned} L(\underline{\beta}) &= \prod_{i=1}^n g(y_i, \underline{\beta}) \\ &= \left(\sqrt{2\pi\sigma^2}\right)^{-\frac{n}{2}} \exp\left(\frac{1}{2\sigma^2} (\underline{y} - \mathbf{X}(\mathbf{K})\underline{\beta})' (\underline{y} - \mathbf{X}(\mathbf{K})\underline{\beta})\right) \end{aligned} \quad (4.7)$$

Apabila persamaan (4.7) dibuat transformasi logaritma, akan didapatkan

$$\begin{aligned} \ell(\underline{\beta}) &= \log L(\underline{\beta}) \\ &= \log\left(\left(\sqrt{2\pi\sigma^2}\right)^{-\frac{n}{2}} \exp\left(\frac{1}{2\sigma^2} (\underline{y} - \mathbf{X}(\mathbf{K})\underline{\beta})' (\underline{y} - \mathbf{X}(\mathbf{K})\underline{\beta})\right)\right) \\ &= -\frac{n}{2} \log\left(\sqrt{2\pi\sigma^2}\right) - \frac{1}{2\sigma^2} (\underline{y} - \mathbf{X}(\mathbf{K})\underline{\beta})' (\underline{y} - \mathbf{X}(\mathbf{K})\underline{\beta}) \end{aligned} \quad (4.8)$$

Dengan menggunakan derivatif parsial terhadap  $\underline{\beta}$  diperoleh:

$$\begin{aligned}
\frac{\partial \ell(\underline{\hat{\beta}})}{\partial \underline{\hat{\beta}}} &= \frac{\partial \left( -\frac{n}{2} \log(\sqrt{2\Pi\sigma^2}) - \frac{1}{2\sigma^2} (\underline{y} - \mathbf{X}(\mathbf{K})\underline{\hat{\beta}})' (\underline{y} - \mathbf{X}(\mathbf{K})\underline{\hat{\beta}}) \right)}{\partial \underline{\hat{\beta}}} \\
&= \frac{\partial \left( -\frac{n}{2} \log(\sqrt{2\Pi\sigma^2}) - \frac{1}{2\sigma^2} \underline{y}'\underline{y} - 2\underline{\hat{\beta}}'\mathbf{X}(\mathbf{K})'\underline{y} + \underline{\hat{\beta}}'\mathbf{X}(\mathbf{K})'\mathbf{X}(\mathbf{K})\underline{\hat{\beta}} \right)}{\partial \underline{\hat{\beta}}} \\
0 &= -\frac{1}{2\sigma^2} \left( -2\mathbf{X}(\mathbf{K})'\underline{y} + 2\mathbf{X}(\mathbf{K})'\mathbf{X}(\mathbf{K})\underline{\hat{\beta}} \right) \\
\underline{\hat{\beta}} &= (\mathbf{X}(\mathbf{K})'\mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})'\underline{y} \tag{4.9}
\end{aligned}$$

Ekspektasi dan variansi  $\underline{\hat{\beta}}$ , berturut-turut diberikan oleh:

$$\begin{aligned}
E(\underline{\hat{\beta}}) &= E\left( (\mathbf{X}(\mathbf{K})'\mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})'\underline{y} \right) \\
&= (\mathbf{X}(\mathbf{K})'\mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})'E(\underline{y}) \\
&= (\mathbf{X}(\mathbf{K})'\mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})'\mathbf{X}(\mathbf{K})\underline{\beta} \\
&= \underline{\beta} \tag{4.10}
\end{aligned}$$

$$\begin{aligned}
Var(\underline{\hat{\beta}}) &= Var\left( (\mathbf{X}(\mathbf{K})'\mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})'\underline{y} \right) \\
&= (\mathbf{X}(\mathbf{K})'\mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})'Var(\underline{y})\left( (\mathbf{X}(\mathbf{K})'\mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})' \right)' \\
&= (\mathbf{X}(\mathbf{K})'\mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})'\sigma^2\mathbf{I}\mathbf{X}(\mathbf{K})(\mathbf{X}(\mathbf{K})'\mathbf{X}(\mathbf{K}))^{-1} \\
&= \sigma^2 (\mathbf{X}(\mathbf{K})'\mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})'\mathbf{X}(\mathbf{K})(\mathbf{X}(\mathbf{K})'\mathbf{X}(\mathbf{K}))^{-1} \\
&= \sigma^2 (\mathbf{X}(\mathbf{K})'\mathbf{X}(\mathbf{K}))^{-1} \tag{4.11}
\end{aligned}$$

Karena sifat linieritas dari distribusi normal, maka

$$\underline{\hat{\beta}} \sim N(\underline{\beta}, \sigma^2 (\mathbf{X}(\mathbf{K})'\mathbf{X}(\mathbf{K}))^{-1})$$

Penelitian yang pernah dilakukan oleh Pratiwi (2015) telah mendapatkan estimasi dari  $\underline{\hat{\beta}}$ , seperti yang telah dijabarkan pada persamaan (4.1) hingga (4.11).

Selanjutnya akan dirancang interval konfidensi  $(1-\alpha)100\%$  untuk  $\underline{\beta}_{ju}$  untuk kasus variansi  $\sigma^2$  diketahui dan variansi  $\sigma^2$  tidak diketahui.

#### 4.2.1 Interval Konfidensi Saat Variansi $\sigma^2$ Diketahui.

Pada subbab ini dirancang interval konfidensi terpendek  $(1-\alpha)100\%$  untuk  $\beta_{ju}$  dimana  $\sigma^2$  diketahui. Diambil sebuah transformasi:

$$Z_{ju}(x_1, x_2, \dots, x_n) = \frac{\hat{\beta}_{ju} - \beta_{ju}}{\sqrt{\sigma^2 a_{jj}}} \quad (4.12)$$

dengan  $a_{jj}$  merupakan elemen diagonal ke  $jj$  dari  $(\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))^{-1}$ . Variabel random  $Z_{ju}$  berdistribusi  $N(0,1)$ .

$$\begin{aligned} E(Z_{ju}) &= E\left(\frac{\hat{\beta}_{ju} - \beta_{ju}}{\sqrt{\sigma^2 a_{jj}}}\right) \\ &= \frac{1}{\sqrt{\sigma^2 a_{jj}}} E(\hat{\beta}_{ju} - \beta_{ju}) \\ &= \frac{1}{\sqrt{\sigma^2 a_{jj}}} (\beta_{ju} - \beta_{ju}) \\ &= 0 \\ \text{Var}(Z_{ju}) &= \text{Var}\left(\frac{\hat{\beta}_{ju} - \beta_{ju}}{\sqrt{\sigma^2 a_{jj}}}\right) \\ &= \left(\frac{1}{\sqrt{\sigma^2 a_{jj}}}\right)^2 \text{Var}(\hat{\beta}_{ju} - \beta_{ju}) \\ &= \left(\frac{1}{\sqrt{\sigma^2 a_{jj}}}\right)^2 \sigma^2 a_{jj} \\ &= 1 \end{aligned}$$

Akibatnya, karena  $Z_{ju}$  merupakan kombinasi linier dari  $\hat{\beta}$  dan  $\hat{\beta} \sim N(\beta, \sigma^2 (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))^{-1})$  maka  $Z_{ju} \sim N(0,1)$ . Karena  $\sigma^2$  diketahui maka  $Z_{ju}(x_1, x_2, \dots, x_n)$  merupakan *pivotal quantity* untuk parameter regresi  $\beta_{ju}$ . Interval konfidensi  $1-\alpha$  dapat diperoleh dengan menyelesaikan persamaan dalam probabilitas

$$P\left(u \leq Z_{ju}(x_1, x_2, \dots, x_n) \leq v\right) = 1 - \alpha \quad (4.13)$$

dengan  $u$  dan  $v$  merupakan elemen bilangan Rill,  $u < v$ . Apabila persamaan (4.12) di substitusi ke dalam persamaan (4.13), maka bentuknya akan menjadi

$$P\left(u \leq \frac{\hat{\beta}_{ju} - \beta_{ju}}{\sqrt{\sigma^2 a_{jj}}} \leq v\right) = 1 - \alpha$$

persamaan diatas ekuivalen dengan persamaan (4.14) dan (4.15) sebagai berikut

$$\beta_{ju} \leq \hat{\beta}_{ju} - u\sqrt{\sigma^2 a_{jj}} \quad (4.14)$$

$$\beta_{ju} \geq \hat{\beta}_{ju} - v\sqrt{\sigma^2 a_{jj}} \quad (4.15)$$

Berdasarkan persamaan (4.14) dan (4.15) didapatkan interval konfidensi untuk parameter regresi nonparametrik spline *truncated* multivariabel adalah

$$P\left(\hat{\beta}_{ju} - v\sqrt{\sigma^2 a_{jj}} \leq \beta_{ju} \leq \hat{\beta}_{ju} - u\sqrt{\sigma^2 a_{jj}}\right) = 1 - \alpha \quad (4.16)$$

Pada penelitian ini menggunakan konsep interval konfidensi terpendek, sehingga harus ditentukan nilai  $u$  dan  $v$  agar panjang interval konfidensi  $\ell(u, v)$  pada persamaan (4.16) menjadi terpendek. Untuk mendapatkan interval konfidensi terpendek, dapat dilakukan dengan menyelesaikan optimasi bersyarat berikut dengan metode *lagrange* sebagai berikut

$$\underset{u, v \in R}{\text{Min}} \{\ell(u, v)\} = \underset{u, v \in R}{\text{Min}} \{(v - u)\sqrt{\sigma^2 a_{jj}}\} \quad (4.17)$$

dengan syarat pada persamaan (4.18)

$$\int_u^v \rho(z) dz = 1 - \alpha, \text{ atau } \Phi(v) - \Phi(u) - (1 - \alpha) = 0 \quad (4.18)$$

Fungsi  $\rho$  merupakan distribusi probabilitas  $N(0,1)$  dan  $\Phi$  merupakan distribusi probabilitas Kumulatif  $N(0,1)$ . Kemudian dibentuk fungsi lagrange

$$F(u, v, \lambda) = (v - u)\sqrt{\sigma^2 a_{jj}} + \lambda(\Phi(v) - \Phi(u) - (1 - \alpha)) \quad (4.19)$$

dimana  $\lambda$  merupakan konstanta *lagrange*. Langkah selanjutnya adalah melakukan derivatif parsial persamaan (4.19) terhadap  $u$ ,  $v$ , dan  $\lambda$

$$\frac{\partial F(u, v, \lambda)}{\partial u} = 0 \Rightarrow -\sqrt{\sigma^2 a_{jj}} - \lambda(\Phi'(u)) = 0 \quad (4.20)$$

$$\frac{\partial F(u, v, \lambda)}{\partial v} = 0 \Rightarrow \sqrt{\sigma^2 a_{jj}} + \lambda(\Phi'(v)) = 0 \quad (4.21)$$

$$\frac{\partial F(u, v, \lambda)}{\partial \lambda} = 0 \Rightarrow (\Phi(v) - \Phi(u) - (1 - \alpha)) = 0 \quad (4.22)$$

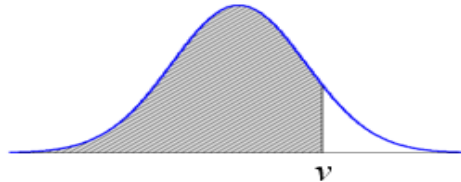
berdasarkan persamaan (4.20) dan (4.21) diperoleh

$$\begin{aligned} -\sqrt{\sigma^2 a_{jj}} - \lambda(\Phi'(u)) &= 0 \\ \frac{\sqrt{\sigma^2 a_{jj}} + \lambda(\Phi'(v))}{\lambda(\Phi'(v) - \Phi'(u))} &+ \\ \Phi'(v) - \Phi'(u) &= 0 \\ \rho(v) &= \rho(u) \end{aligned} \quad (4.23)$$

maka penyelesaian persamaan (4.23) adalah  $u = v$  atau  $u = -v$ . Tetapi yang memenuhi adalah  $u = -v$ , dan jika hasil tersebut di substitusi ke dalam persamaan (4.16) maka akan menghasilkan persamaan (4.24) sebagai berikut.

$$P\left(\hat{\beta}_{ju} - v\sqrt{\sigma^2 a_{jj}} \leq \tilde{\beta}_{ju} \leq \hat{\beta}_{ju} + v\sqrt{\sigma^2 a_{jj}}\right) = 1 - \alpha \quad (4.24)$$

dimana  $v$  diperoleh dari.



**Gambar 4.1** Distribusi Normal

$\int_v^{\infty} \rho(z) dz = \alpha/2$ . Dalam aplikasi  $v$  sering ditulis dengan simbol  $Z_{\alpha/2}$ , dimana nilai

tersebut dapat dilihat dalam tabel distribusi  $N(0,1)$ .

#### 4.2.2 Interval Konfidensi Saat Variansi $\sigma^2$ Tidak Diketahui.

Pada subbab ini akan dibahas tentang interval konfidensi saat  $\sigma^2$  tidak diketahui. Persamaan (4.12) merupakan *pivotal quantity* untuk  $\tilde{\beta}_{ju}$  saat  $\sigma^2$  diketahui, namun dalam aplikasi sering dijumpai kasus dengan  $\sigma^2$  tidak diketahui. Oleh Karena itu  $\sigma^2$  pada persamaan (4.12) akan diduga dengan MSE, sehingga akan didapatkan *pivotal quantity* seperti pada persamaan (4.25).

$$T_{ju}(x_1, x_2, \dots, x_n) = \frac{\hat{\beta}_{ju} - \beta_{ju}}{\sqrt{MSE(\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))_{jj}^{-1}}} \quad (4.25)$$

dimana  $(\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))_{jj}^{-1}$  merupakan elemen diagonal ke  $jj$  dari matriks  $(\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))^{-1}$  dan MSE nya adalah sebagai berikut.

$$\begin{aligned} MSE &= \frac{(\underline{y} - \hat{\underline{y}})' (\underline{y} - \hat{\underline{y}})}{n - p(r+1)} \\ &= \frac{(\underline{y} - \mathbf{X}(\mathbf{K}) \hat{\underline{\beta}})' (\underline{y} - \mathbf{X}(\mathbf{K}) \hat{\underline{\beta}})}{n - p(r+1)} \\ &= \frac{(\underline{y} - \mathbf{X}(\mathbf{K}) (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})' \underline{y})' (\underline{y} - \mathbf{X}(\mathbf{K}) (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})' \underline{y})}{n - p(r+1)} \\ &= \frac{\underline{y}' [I - \mathbf{X}(\mathbf{K}) (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})'] \underline{y}}{n - p(r+1)} \end{aligned} \quad (4.26)$$

Variabel random  $T = \frac{B}{\sqrt{A/a}} \sim t_{(a)}$  jika dan hanya jika memenuhi ketiga syarat yaitu

$B \sim N(0,1)$ ,  $A \sim \chi_{(a)}$ , Adan  $B$  saling independen. Variabel random  $T_{ju}$  pada persamaan (4.25) berdistribusi  $t_{(n-p(r+1))}$  karena telah memenuhi ketiga syarat tersebut. Untuk mempermudah pembuktiannya, akan dilakukan sedikit penjabaran pada persamaan (4.25) dan diperoleh hasil pada persamaan (4.27).

$$\begin{aligned} T_{ju}(x_1, x_2, \dots, x_n) &= \frac{\hat{\beta}_{ju} - \beta_{ju}}{\sqrt{\frac{\underline{y}' [I - \mathbf{X}(\mathbf{K}) (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})'] \underline{y}}{n - p(1+r)} (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))_{jj}^{-1}}} \\ &= \frac{\frac{\hat{\beta}_{ju} - \beta_{ju}}{\sqrt{(\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))_{jj}^{-1}}}}{\sqrt{\frac{\underline{y}' [I - \mathbf{X}(\mathbf{K}) (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})'] \underline{y}}{n - p(1+r)}}} \end{aligned}$$



$$= \frac{\frac{\hat{\beta}_{ju} - \beta_{ju}}{\sqrt{\sigma^2 (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))_{jj}^{-1}}}}{\sqrt{\frac{\tilde{y}' [I - \mathbf{X}(\mathbf{K})(\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})'] \tilde{y}}{\sigma^2}} / n - p(1+r)}} \quad (4.27)$$

Berikut ini adalah pembuktian bahwa variabel random  $T_{ju}$  berdistribusi  $t$  dengan derajat bebas  $n-p(1+r)$ .

i) Misalkan  $Z = \frac{\hat{\beta}_{ju} - \beta_{ju}}{\sqrt{\sigma^2 (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))_{jj}^{-1}}}$ , karena  $Z$  merupakan kombinasi linier dari

$\hat{\beta}_{ju}$  maka  $Z$  berdistribusi normal dengan mean  $E(Z)$  dan varians  $Var(Z)$

$$\begin{aligned} E(Z) &= E \left( \frac{\hat{\beta}_{ju} - \beta_{ju}}{\sqrt{\sigma^2 (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))_{jj}^{-1}}} \right) \\ &= \frac{1}{\sqrt{\sigma^2 (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))_{jj}^{-1}}} E(\hat{\beta}_{ju} - \beta_{ju}) \\ &= \frac{1}{\sqrt{\sigma^2 (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))_{jj}^{-1}}} (\beta_{ju} - \beta_{ju}) \\ &= 0 \end{aligned} \quad (4.28)$$

$$\begin{aligned} Var(Z) &= Var \left( \frac{\hat{\beta}_{ju} - \beta_{ju}}{\sqrt{\sigma^2 (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))_{jj}^{-1}}} \right) \\ &= \left( \frac{1}{\sqrt{\sigma^2 (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))_{jj}^{-1}}} \right)^2 Var(\hat{\beta}_{ju} - \beta_{ju}) \\ &= \left( \frac{1}{\sqrt{\sigma^2 (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))_{jj}^{-1}}} \right)^2 \sigma^2 (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))_{jj}^{-1} \\ &= 1 \end{aligned} \quad (4.29)$$

Berdasarkan persamaan (4.28) dan (4.29) didapatkan  $Z$  berdistribusi Normal dengan mean 0 dan varians 1.

ii) Jika  $\tilde{y} \sim N(\mathbf{X}(\mathbf{K})\tilde{\beta}, \sigma^2\mathbf{I})$  maka  $\tilde{y}'\mathbf{A}\tilde{y}$  akan berdistribusi  $\chi^2_{\left(c, \frac{1}{2\sigma^2}(\mathbf{X}(\mathbf{K})\tilde{\beta})' \mathbf{A}(\mathbf{X}(\mathbf{K})\tilde{\beta})\right)}$

jika dan hanya jika  $\mathbf{A}$  idempoten dengan  $c$  adalah rank dari  $\mathbf{A}$ .

$$\tilde{y}'\mathbf{A}\tilde{y} = \frac{\tilde{y}' \left[ \mathbf{I} - \mathbf{X}(\mathbf{K})(\mathbf{X}(\mathbf{K})'\mathbf{X}(\mathbf{K}))^{-1}\mathbf{X}(\mathbf{K})' \right] \tilde{y}}{\sigma^2}$$

dengan  $\mathbf{A} = \left[ \mathbf{I} - \mathbf{X}(\mathbf{K})(\mathbf{X}(\mathbf{K})'\mathbf{X}(\mathbf{K}))^{-1}\mathbf{X}(\mathbf{K})' \right]$ , berikut ini adalah pembuktian

bahwa  $\mathbf{A}$  idempoten

$$\begin{aligned} \mathbf{A}\mathbf{A} &= \left[ \mathbf{I} - \mathbf{X}(\mathbf{K})(\mathbf{X}(\mathbf{K})'\mathbf{X}(\mathbf{K}))^{-1}\mathbf{X}(\mathbf{K})' \right] \left[ \mathbf{I} - \mathbf{X}(\mathbf{K})(\mathbf{X}(\mathbf{K})'\mathbf{X}(\mathbf{K}))^{-1}\mathbf{X}(\mathbf{K})' \right] \\ &= \mathbf{I} - \mathbf{X}(\mathbf{K})(\mathbf{X}(\mathbf{K})'\mathbf{X}(\mathbf{K}))^{-1}\mathbf{X}(\mathbf{K})' - \mathbf{X}(\mathbf{K})(\mathbf{X}(\mathbf{K})'\mathbf{X}(\mathbf{K}))^{-1}\mathbf{X}(\mathbf{K})' \\ &\quad + \mathbf{X}(\mathbf{K})(\mathbf{X}(\mathbf{K})'\mathbf{X}(\mathbf{K}))^{-1}\mathbf{X}(\mathbf{K})'(\mathbf{X}(\mathbf{K})'\mathbf{X}(\mathbf{K}))^{-1}\mathbf{X}(\mathbf{K})' \\ &= \mathbf{I} - 2\mathbf{X}(\mathbf{K})(\mathbf{X}(\mathbf{K})'\mathbf{X}(\mathbf{K}))^{-1}\mathbf{X}(\mathbf{K})' + \mathbf{X}(\mathbf{K})(\mathbf{X}(\mathbf{K})'\mathbf{X}(\mathbf{K}))^{-1}\mathbf{X}(\mathbf{K})' \\ &= \mathbf{I} - \mathbf{X}(\mathbf{K})(\mathbf{X}(\mathbf{K})'\mathbf{X}(\mathbf{K}))^{-1}\mathbf{X}(\mathbf{K})' \end{aligned} \quad (4.30)$$

berdasarkan hasil pada persamaan (4.30) menunjukkan bahwa  $\mathbf{A}$  idempoten, rank dari  $\mathbf{A}$  adalah sebagai berikut.

$$\begin{aligned} \text{rank}(\mathbf{A}) &= \text{rank}\left(\mathbf{I} - \mathbf{X}(\mathbf{K})(\mathbf{X}(\mathbf{K})'\mathbf{X}(\mathbf{K}))^{-1}\mathbf{X}(\mathbf{K})'\right) \\ &= \text{trace}\left(\mathbf{I} - \mathbf{X}(\mathbf{K})(\mathbf{X}(\mathbf{K})'\mathbf{X}(\mathbf{K}))^{-1}\mathbf{X}(\mathbf{K})'\right) \\ &= \text{trace}(\mathbf{I}_n) - \text{trace}\left(\mathbf{X}(\mathbf{K})(\mathbf{X}(\mathbf{K})'\mathbf{X}(\mathbf{K}))^{-1}\mathbf{X}(\mathbf{K})'\right) \\ &= n - \text{trace}\left(\mathbf{X}(\mathbf{K})'\mathbf{X}(\mathbf{K})(\mathbf{X}(\mathbf{K})'\mathbf{X}(\mathbf{K}))^{-1}\right) \\ &= n - \text{trace}\left(\mathbf{I}_{p(r+1)}\right) \\ &= n - p(r+1) \end{aligned}$$

Setelah terbukti bahwa  $\mathbf{A}$  idempotent, kemudian akan dihitung

$$\begin{aligned} \frac{1}{2\sigma^2}(\mathbf{X}(\mathbf{K})\tilde{\beta})' \mathbf{A}(\mathbf{X}(\mathbf{K})\tilde{\beta}) &= \frac{1}{2\sigma^2}(\mathbf{X}(\mathbf{K})\tilde{\beta})' \mathbf{X}(\mathbf{K})(\mathbf{X}(\mathbf{K})'\mathbf{X}(\mathbf{K}))^{-1}\mathbf{X}(\mathbf{K})'(\mathbf{X}(\mathbf{K})\tilde{\beta}) \\ &= \frac{1}{2\sigma^2}\tilde{\beta}'\mathbf{X}(\mathbf{K})'\mathbf{X}(\mathbf{K})\tilde{\beta} - \tilde{\beta}'\mathbf{X}(\mathbf{K})'\mathbf{X}(\mathbf{K})(\mathbf{X}(\mathbf{K})'\mathbf{X}(\mathbf{K}))^{-1}\mathbf{X}(\mathbf{K})'(\mathbf{X}(\mathbf{K})\tilde{\beta}) \end{aligned}$$

$$= \frac{1}{2\sigma^2} \beta' \mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}) \beta - \beta' \mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}) \beta = 0$$

Berdasarkan uraian tersebut dapat diambil kesimpulan bahwa  $\tilde{y}' \mathbf{A} \tilde{y} \sim \chi_{(n-p(1+r))}$ .

iii) Jika  $\tilde{y} \sim N(\mathbf{X}(\mathbf{K})\beta, \sigma^2 \mathbf{I})$  maka  $\mathbf{B}\tilde{y}$  dan  $\tilde{y}' \mathbf{A} \tilde{y}$  independen jika dan hanya jika

$\mathbf{B}\mathbf{A}=\mathbf{0}$ . Karena  $\hat{\beta}_{ju}$  dapat ditulis menjadi  $\hat{\beta} = (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})' y$  maka

diperoleh  $(\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})' y = \mathbf{B}\tilde{y}$  dengan  $\mathbf{B} = (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})'$ .

Berikut ini hasil dari perkalian matriks  $\mathbf{B}\mathbf{A}$ .

$$\begin{aligned} \mathbf{B}\mathbf{A} &= \left[ (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})' \right] \left[ \mathbf{I} - \mathbf{X}(\mathbf{K}) (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})' \right] \\ &= (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})' - (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}) (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})' \\ &= 0 \end{aligned}$$

Berdasarkan uraian pada poin (i), (ii), dan (iii) dapat ditarik kesimpulan bahwa  $T_{ju}$

berdistribusi  $t$  dengan derajat bebas  $n-p(1+r)$ . Sehingga  $T_{ju}(x_1, x_2, \dots, x_n)$

merupakan *pivotal quantity* untuk parameter regresi  $\beta_{ju}$  saat  $\sigma^2$  tidak diketahui.

Interval konfidensi  $1-\alpha$  dapat diperoleh dengan menyelesaikan persamaan dalam probabilitas

$$P(g \leq T_{ju}(x_1, x_2, \dots, x_n, y) \leq h) = 1 - \alpha \quad (4.31)$$

dengan  $g$  dan  $h$  merupakan elemen bilangan Rill,  $g < h$ . Apabila persamaan (4.25)

di substitusi ke dalam persamaan (4.31), maka bentuknya akan menjadi

$$P \left( g \leq \frac{\hat{\beta}_{ju} - \beta_{ju}}{\sqrt{\frac{y' \left[ \mathbf{I} - \mathbf{X}(\mathbf{K}) (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})' \right] y}{n - p(r+1)} (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))_{jj}^{-1}}} \leq h \right) = 1 - \alpha$$

persamaan tersebut ekuivalen dengan

$$\beta_{ju} \leq \hat{\beta}_{ju} - g \sqrt{\frac{y' \left[ \mathbf{I} - \mathbf{X}(\mathbf{K}) (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})' \right] y}{n - p(r+1)} (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))_{jj}^{-1}} \quad (4.32)$$

$$\beta_{ju} \geq \hat{\beta}_{ju} - h \sqrt{\frac{y' \left[ I - \mathbf{X}(\mathbf{K}) (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})' \right] y}{n - p(r+1)}} (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))_{jj}^{-1}} \quad (4.33)$$

Berdasarkan persamaan (4.32) dan (4.33) didapatkan interval konfidensi untuk parameter regresi nonparametrik spline *truncated* multivariabel saat  $\sigma^2$  tidak diketahui adalah

$$P \left( \hat{\beta}_{ju} - h \sqrt{\frac{y' \left[ I - \mathbf{X}(\mathbf{K}) (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})' \right] y}{n - p(r+1)}} (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))_{jj}^{-1}} \leq \beta_{ju} \leq \hat{\beta}_{ju} + g \sqrt{\frac{y' \left[ I - \mathbf{X}(\mathbf{K}) (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})' \right] y}{n - p(r+1)}} (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))_{jj}^{-1}} \right) = 1 - \alpha \quad (4.34)$$

Pada penelitian ini menggunakan konsep interval konfidensi terpendek, sehingga harus ditentukan nilai  $g$  dan  $h$  agar panjang interval konfidensi  $\ell(g, h)$  pada persamaan (4.34) menjadi terpendek. Untuk mendapatkan interval konfidensi terpendek, dapat dilakukan dengan menyelesaikan optimasi bersyarat berikut dengan metode lagrange

$$\text{Min}_{g, h \in R} \{ \ell(u, v) \} = \text{Min}_{g, h \in R} \left\{ (h - g) \sqrt{\frac{y' \left[ I - \mathbf{X}(\mathbf{K}) (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})' \right] y}{n - p(r+1)}} (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))_{jj}^{-1}} \right\} \quad (4.35)$$

dengan syarat

$$\int_g^h \rho(t) dt = 1 - \alpha, \text{ atau } \Phi(h) - \Phi(g) - (1 - \alpha) = 0 \quad (4.36)$$

Fungsi  $\rho$  merupakan distribusi probabilitas  $t_{(n-p(1+r))}$  dan  $\Phi$  merupakan distribusi probabilitas Kumulatif  $t_{(n-p(1+r))}$ . Kemudian dibentuk fungsi lagrange untuk optimasi

$$F(g, h, \lambda) = (h - g) \sqrt{\frac{y' \left[ I - \mathbf{X}(\mathbf{K}) (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})' \right] y}{n - p(r+1)}} (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))_{jj}^{-1} + \lambda (\Phi(h) - \Phi(g) - (1 - \alpha)) \quad (4.37)$$

dimana  $\lambda$  merupakan konstanta lagrange. Langkah selanjutnya adalah melakukan derivatif parsial persamaan (4.37) terhadap  $g$ ,  $h$ , dan  $\lambda$

$$\frac{\partial F(g, h, \lambda)}{\partial g} = 0 \Rightarrow -\sqrt{\frac{y' \left[ I - \mathbf{X}(\mathbf{K}) (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})' \right] y}{n - p(r+1)}} (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))_{jj}^{-1} - \lambda (\Phi'(g)) = 0 \quad (4.38)$$

$$\frac{\partial F(g, h, \lambda)}{\partial h} = 0 \Rightarrow \sqrt{\frac{y' \left[ I - \mathbf{X}(\mathbf{K}) (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})' \right] y}{n - p(r+1)}} (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))_{jj}^{-1} + \lambda (\Phi'(h)) = 0 \quad (4.39)$$

$$\frac{\partial F(g, h, \lambda)}{\partial \lambda} = 0 \Rightarrow (\Phi(h) - \Phi(g) - (1 - \alpha)) = 0 \quad (4.40)$$

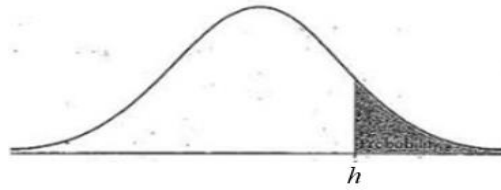
berdasarkan persamaan (4.38) dan (4.39) diperoleh

$$\begin{aligned} -\sqrt{\frac{y' \left[ I - \mathbf{X}(\mathbf{K}) (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})' \right] y}{n - p(r+1)}} (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))_{jj}^{-1} - \lambda (\Phi'(g)) &= 0 \\ \sqrt{\frac{y' \left[ I - \mathbf{X}(\mathbf{K}) (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})' \right] y}{n - p(r+1)}} (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))_{jj}^{-1} + \lambda (\Phi'(h)) &= 0 \\ \lambda (\Phi'(h) - \Phi'(g)) &= 0 \\ \Phi'(h) - \Phi'(g) &= 0 \\ \rho(h) = \rho(g) & \quad (4.41) \end{aligned}$$

maka penyelesaian persamaan (4.41) adalah  $g = h$  atau  $g = -h$ . Tetapi yang memenuhi adalah  $g = -h$ , dan jika hasil tersebut di substitusi ke dalam persamaan (4.34) maka akan menghasilkan

$$P \left( \hat{\beta}_{ju} - h \sqrt{\frac{y' [I - \mathbf{X}(\mathbf{K}) (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})'] y}{n - p(r+1)}} (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))_{jj}^{-1} \leq \beta_{ju} \leq \hat{\beta}_{ju} + h \sqrt{\frac{y' [I - \mathbf{X}(\mathbf{K}) (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})'] y}{n - p(r+1)}} (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))_{jj}^{-1} \right) = 1 - \alpha \quad (4.43)$$

dimana  $h$  diperoleh dari



**Gambar 4.2** Distribusi *student-t*

$\int_h^{\infty} \rho(t) dt = \alpha/2$ . Dalam aplikasi  $h$  sering ditulis dengan symbol  $t_{\alpha/2}$ , dimana nilai tersebut dapat dilihat dalam tabel distribusi  $t_{\alpha/2, (n-p(r+1))}$ .

Berdasarkan uraian yang tersebut didapatkan hasil yaitu interval konfidensi untuk parameter regresi nonparametrik spline *truncated* multivariabel terbagi menjadi dua yaitu saat saat  $\sigma^2$  diketahui dan  $\sigma^2$  tidak diketahui. Interval konfidensi saat  $\sigma^2$  diketahui adalah sebagai berikut.

$$P \left( \hat{\beta}_{ju} - z_{\alpha/2} \sqrt{\sigma^2 a_{jj}} \leq \beta_{ju} \leq \hat{\beta}_{ju} + z_{\alpha/2} \sqrt{\sigma^2 a_{jj}} \right) = 1 - \alpha \quad (4.44)$$

dengan  $a_{jj}$  merupakan elemen diagonal ke  $jj$  dari  $(\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))^{-1}$ . Interval konfidensi saat saat  $\sigma^2$  tidak diketahui adalah

$$P \left( \hat{\beta}_{ju} - t_{\alpha/2, (n-p(r+1))} \sqrt{\frac{y' [I - \mathbf{X}(\mathbf{K}) (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})'] y}{n - p(1+r)}} (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))_{jj}^{-1} \leq \beta_{ju} \leq \hat{\beta}_{ju} + t_{\alpha/2, (n-p(r+1))} \sqrt{\frac{y' [I - \mathbf{X}(\mathbf{K}) (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})'] y}{n - p(1+r)}} (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))_{jj}^{-1} \right) = 1 - \alpha \quad (4.45)$$

dengan  $(\mathbf{X}(\mathbf{K})'\mathbf{X}(\mathbf{K}))_{jj}^{-1}$  merupakan elemen diagonal ke  $jj$  dari  $(\mathbf{X}(\mathbf{K})'\mathbf{X}(\mathbf{K}))^{-1}$ . Setelah didapatkan interval konfidensi selanjutnya akan diterapkan pada data IPG di provinsi Jawa Timur.

## 4.2 Aplikasi Interval Konfidensi Untuk Parameter Regresi Nonparametrik Spline *Truncated* Multivariabel Pada Data IPG di Jawa Timur

Pada bagian ini akan dibahas mengenai Indeks Pembangunan Gender (IPG) di Jawa Timur dan beberapa faktor yang diduga mempengaruhinya menggunakan statistika deskriptif dan pemodelan IPG menggunakan model regresi nonparametrik spline *truncated* multivariabel dengan fungsi Spline linier satu knot, dua knot, tiga knot, dan kombinasi knot.

### 4.2.1 Deskripsi IPG dan Faktor-Faktor yang Diduga Mempengaruhi

Provinsi Jawa Timur memiliki 29 Kabupaten dan 9 Kota atau secara administratif terdapat 38 Kabupaten/Kota. Dalam skala nasional capaian IPG Jawa Timur pada tahun 2014 menempati urutan nomor 16. Dalam penelitian ini digunakan data IPG Jawa Timur pada tahun 2014 beserta beberapa faktor yang diduga mempengaruhinya. Berikut ini hasil analisa deskriptif dari tiap variabel penelitian.

**Tabel 4.1** Deskripsi IPG dan Faktor Yang Diduga Mempengaruhi

Variabel	Minimum	Maximum	Rata-Rata
$y$	76,63	98,23	90,06
$x_1$	40,60	89,45	70,88
$x_2$	9,78	26,40	14,99
$x_3$	69,63	100	94,84
$x_4$	43,58	72,11	54,28

Berdasarkan Tabel 4.1 dapat diketahui bahwa rata-rata IPG ( $y$ ) di Provinsi Jawa Timur pada tahun 2014 sebesar 90,06. Nilai IPG tertinggi di provinsi Jawa Timur sebesar 98,23 yaitu Kota Blitar, sedangkan Kabupaten Sumenep memiliki nilai IPG terendah yaitu sebesar 76,63. Beberapa variabel yang diduga mempengaruhi IPG diantaranya APS Sekolah Lanjutan Tingkat Atas

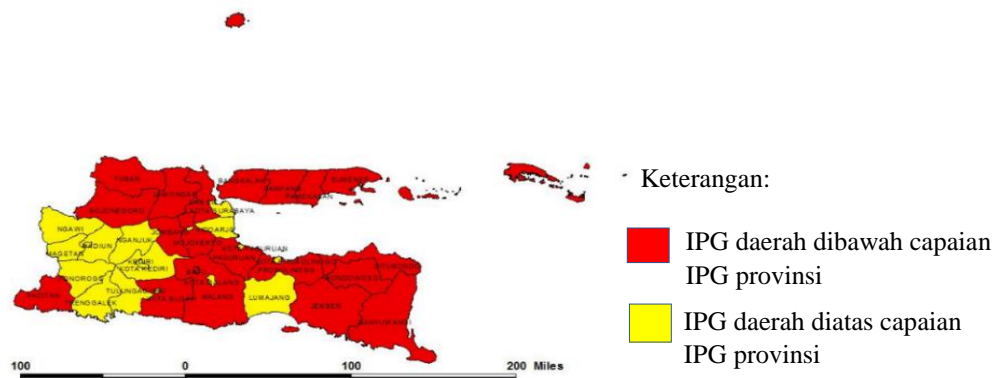
(SLTA) penduduk perempuan ( $x_1$ ), angka kesakitan penduduk perempuan ( $x_2$ ), penolong terakhir kelahiran oleh medis ( $x_3$ ), dan tingkat partisipasi angkatan kerja (TPAK) penduduk perempuan ( $x_4$ ). Berikut ini adalah deskripsi variabel yang diduga mempengaruhi IPG di provinsi Jawa Timur tahun 2014.

Rata-rata APS SLTA penduduk perempuan di provinsi Jawa Timur sebesar 70,88%. APS SLTA penduduk perempuan memiliki nilai tertinggi sebesar 89,45% yaitu Kota Kediri, sedangkan Kabupaten Sampang memiliki APS penduduk perempuan terendah sebesar 40,60%. Hal tersebut menunjukkan bahwa proporsi anak perempuan yang bersekolah pada kelompok umur 16-18 tahun pada jenjang SLTA di Kabupaten Sampang masih rendah, bahkan dibawah 50%. Pada bidang kesehatan, rata-rata angka kesakitan penduduk perempuan sebesar 14,99%. Kabupaten yang memiliki angka kesakitan penduduk perempuan tertinggi yaitu Kabupaten Sampang sebesar 26,4%. Pada variabel penolong terakhir kelahiran oleh medis memiliki nilai maksimum 100%, hal itu menunjukkan pada daerah tersebut ketersediaan tenaga kesehatan sangat baik. Kabupaten yang memiliki persentase penolong terakhir kelahiran oleh medis mencapai 100% adalah Kabupaten Sidoarjo, Magetan, Ngawi, Kota Kediri, Kota Mojokerto, dan Kota Batu.

Pada variabel TPAK penduduk perempuan pada tahun 2014 memiliki rata-rata sebesar 54,28%. Daerah yang memiliki TPAK penduduk perempuan di bawah rata-rata TPAK provinsi Jawa Timur diantaranya adalah Kabupaten Kediri, Malang, Lumajang, Jember, Banyuwangi, Situbondo, Mojokerto, Jombang, Nganjuk, Madiun, Ngawi, Bojonegoro, Tuban, Lamongan, Gresik, Kota Blitar, Malang. Probolinggo, Pasuruan, Mojokerto, Madiun, Surabaya, dan Batu. Masih terdapat lebih dari 50% kabupaten/kota yang nilai TPAK penduduk perempuan dibawah rata-rata TPAK provinsi.

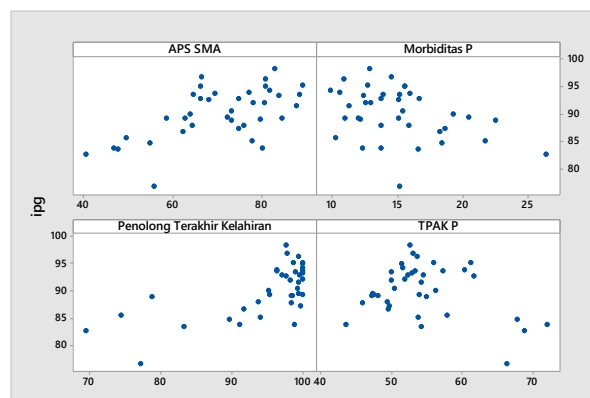
Perkembangan nilai IPG di provinsi Jawa Timur dari tahun 2010 hingga tahun 2014 sudah mengalami peningkatan, namun pada tahun 2014 masih memisahkan sebanyak 20 kabupaten/kota dibawah capaian IPG provinsi sebanyak 18 kabupaten/kota diatas capaian IPG provinsi. Berikut ini akan ditampilkan persebaran nilai IPG di Provinsi Jawa Timur tahun 2014.





**Gambar 4.3** Persebaran IPG di Jawa Timur Tahun 2014

Berdasarkan Gambar 4.3 dapat dilihat bahwa terdapat 20 daerah yang masih memiliki IPG dibawah capaian IPG provinsi, ditunjukkan oleh warna merah. Sedangkan daerah yang berwarna kuning menunjukkan nilai IPG nya sudah berada diatas capaian IPG provinsi. Daerah yang memiliki IPG dibawah capaian IPG provinsi adalah Kabupaten Pacitan, Malang, Lumajang, Jember, Banyuwangi, Bondowoso, Situbondo, Probolinggo, Pasuruan, Mojokerto, Jombang, Bojonegoro, Tuban, Lamongan, Gresik, Bangkalan, Sampang, Pamekasan, Sumenep, dan Kota Batu. Sebelum melakukan pemodelan, terlebih dahulu diselidiki pola hubungan antara IPG dan variabel yang diduga mempengaruhinya. Berikut ini adalah *scatterplot* antara IPG dengan masing-masing variabel yang diduga mempengaruhinya.



**Gambar 4.4** Scatterplot Antara IPG dengan Faktor Yang Diduga Mempengaruhi

Berdasarkan Gambar 4.4 dapat dilihat bahwa pola hubungan IPG dengan beberapa variabel prediktor yaitu angka partisipasi sekolah APS Sekolah Lanjutan Tingkat Atas (SLTA) penduduk perempuan ( $x_1$ ), angka kesakitan penduduk perempuan ( $x_2$ ), penolong terakhir kelahiran oleh medis ( $x_3$ ), dan tingkat partisipasi angkatan kerja (TPAK) penduduk perempuan ( $x_4$ ) cenderung tidak

mengikuti pola tertentu. Oleh Karena itu data IPG dengan faktor-faktor yang diduga mempengaruhinya akan dimodelkan menggunakan regresi nonparametrik *spline truncated* multivariabel.

#### 4.2.2 Pemodelan IPG di Provinsi Jawa Timur dengan Regresi Nonparametrik Spline Linier Multivariabel

Langkah pertama yang dilakukan dalam pemodelan menggunakan regresi nonparametrik spline linier multivariabel adalah melakukan pemilihan titik knot optimal dengan menggunakan satu, dua, tiga, dan kombinasi titik knot. Titik knot merupakan titik perpaduan bersama dimana terdapat perubahan perilaku fungsi pada interval yang berlainan (Budiantara, 2006). Pemilihan titik knot optimal dilakukan untuk menentukan model terbaik yang terbentuk. Salah satu metode yang digunakan untuk memilih titik knot optimal adalah dengan menggunakan metode GCV (*Generalized Cross Validation*) (Budiantara, 2000). Berikut ini dibahas mengenai pemilihan titik knot optimal menggunakan satu, dua, tiga, dan kombinasi titik knot.

##### 4.2.2.1 Pemilihan Titik Knot Optimal dengan Satu Titik Knot

Pada sub bab ini akan dibahas mengenai pemilihan titik knot optimal pada data IPG dan enam variabel prediktor yang diduga mempengaruhinya. Berikut ini adalah model regresi nonparametrik spline multivariabel dengan satu titik knot menggunakan empat variabel prediktor.

$$\hat{y}_i = \hat{\beta}_{01} + \hat{\beta}_{11}x_{1i} + \hat{\beta}_{12}(x_{1i} - K_{11})_+^1 + \hat{\beta}_{21}x_{2i} + \hat{\beta}_{22}(x_{2i} - K_{21})_+^1 + \hat{\beta}_{31}x_{3i} + \hat{\beta}_{32}(x_{3i} - K_{31})_+^1 + \hat{\beta}_{41}x_{4i} + \hat{\beta}_{42}(x_{4i} - K_{41})_+^1$$

Nilai GCV yang dihasilkan oleh regresi nonparametrik dengan satu knot adalah sebagai mana tabel 4.2

**Tabel 4.2** Nilai GCV untuk Spline Linier Satu Knot

Knot				GCV
$x_1$	$x_2$	$x_3$	$x_4$	
67,52	18,94	86,36	57,53	8,30
68,51	19,28	86,98	58,05	8,19
<b>69,51</b>	<b>19,62</b>	<b>87,60</b>	<b>58,57</b>	<b>8,16</b>

**Tabel 4.2** Nilai GCV untuk Spline Linier Satu Knot (Lanjutan)

Knot				GCV
$x_1$	$x_2$	$x_3$	$x_4$	
70,51	19,96	88,22	59,08	8,22
71,51	20,29	88,84	59,60	8,37

Berdasarkan tabel 4.2 didapatkan nilai GCV yang minimum dengan satu titik knot adalah 8,16.

#### 4.2.2.2 Pemilihan Titik Knot Optimal dengan Dua Titik Knot

Langkah selanjutnya setelah mendapatkan *gcv* minimum dengan satu titik knot, selanjutnya adalah melakukan pemilihan titik knot optimal dengan dua titik knot. Berikut ini adalah model regresi nonparametrik spline multivariabel dengan dua titik knot menggunakan empat variabel prediktor.

$$\hat{y}_i = \hat{\beta}_{01} + \hat{\beta}_{11}x_{1i} + \hat{\beta}_{12}(x_{1i} - K_{11})_+^1 + \hat{\beta}_{13}(x_{1i} - K_{12})_+^1 + \hat{\beta}_{21}x_{2i} + \hat{\beta}_{22}(x_{2i} - K_{21})_+^1 + \hat{\beta}_{23}(x_{2i} - K_{22})_+^1 + \hat{\beta}_{31}x_{3i} + \hat{\beta}_{32}(x_{3i} - K_{31})_+^1 + \hat{\beta}_{33}(x_{3i} - K_{32})_+^1 + \hat{\beta}_{41}x_{4i} + \hat{\beta}_{42}(x_{4i} - K_{41})_+^1 + \hat{\beta}_{43}(x_{4i} - K_{42})_+^1$$

Nilai GCV yang dihasilkan oleh regresi nonparametrik dengan dua knot adalah sebagai berikut,

**Tabel 4.3** Nilai GCV untuk Spline Linier Dua Knot

Knot				GCV
$x_1$	$x_2$	$x_3$	$x_4$	
72,50	20,63	89,46	62,21	8,84
73,50	20,97	90,08	62,79	
72,50	20,63	89,46	62,21	8,53
74,50	21,31	90,70	63,38	
<b>72,50</b>	<b>20,63</b>	<b>89,46</b>	<b>62,21</b>	<b>8,49</b>
<b>75,49</b>	<b>21,65</b>	<b>91,32</b>	<b>63,96</b>	
72,50	20,63	89,46	62,21	8,67
76,49	21,99	91,94	64,54	
72,50	20,63	89,46	62,21	9,23
77,49	22,33	92,56	65,12	

Berdasarkan tabel 4.3 didapatkan nilai GCV yang minimum dengan dua titik knot adalah 8,49.

#### 4.2.2.3 Pemilihan Titik Knot Optimal dengan Tiga Titik Knot

Setelah didapatkan nilai gcv minimum dengan dua titik knot, selanjutnya adalah melakukan pemilihan titik knot optimal dengan tiga titik knot. Berikut ini adalah model regresi nonparametrik spline multivariabel dengan tiga titik knot menggunakan empat variabel prediktor.

$$\hat{y}_i = \hat{\beta}_{01} + \hat{\beta}_{11}x_{1i} + \hat{\beta}_{12}(x_{1i} - K_{11})_+^1 + \hat{\beta}_{13}(x_{1i} - K_{12})_+^1 + \hat{\beta}_{14}(x_{1i} - K_{13})_+^1 + \hat{\beta}_{21}x_{2i} + \hat{\beta}_{22}(x_{2i} - K_{21})_+^1 + \hat{\beta}_{23}(x_{2i} - K_{22})_+^1 + \hat{\beta}_{24}(x_{2i} - K_{23})_+^1 + \hat{\beta}_{31}x_{3i} + \hat{\beta}_{32}(x_{3i} - K_{31})_+^1 + \hat{\beta}_{33}(x_{3i} - K_{32})_+^1 + \hat{\beta}_{34}(x_{3i} - K_{33})_+^1 + \hat{\beta}_{41}x_{4i} + \hat{\beta}_{42}(x_{4i} - K_{41})_+^1 + \hat{\beta}_{43}(x_{4i} - K_{42})_+^1 + \hat{\beta}_{44}(x_{4i} - K_{43})_+^1$$

Nilai GCV yang dihasilkan oleh regresi nonparametrik dengan tiga titik knot adalah sebagai mana tabel 4.4.

**Tabel 4.4** Nilai GCV untuk Spline Linier Tiga Knot

Knot				GCV
$x_1$	$x_2$	$x_3$	$x_4$	
42,59	10,46	70,87	44,74	10,22
47,58	12,15	73,97	47,66	
73,50	20,97	90,08	62,79	
42,59	10,46	70,87	44,74	10,32
47,58	12,15	73,97	47,66	
74,50	21,31	90,70	63,38	
<b>42,59</b>	<b>10,46</b>	<b>70,87</b>	<b>44,74</b>	<b>7,19</b>
<b>47,58</b>	<b>12,15</b>	<b>73,97</b>	<b>47,66</b>	
<b>75,49</b>	<b>21,65</b>	<b>91,32</b>	<b>63,96</b>	
42,59	10,46	70,87	44,74	7,50
47,58	12,15	73,97	47,66	
76,49	21,99	91,94	64,54	
42,59	10,46	70,87	44,74	8,14
47,58	12,15	73,97	47,66	
77,49	22,33	92,56	64,54	

Berdasarkan tabel 4.4 didapatkan nilai GCV yang minimum dengan tiga titik knot adalah 7,19.

#### 4.2.2.4 Pemilihan Titik Knot Optimal dengan Kombinasi Titik Knot

Setelah didapatkan nilai gcv minimum dengan satu, dua, dan tiga titik knot, selanjutnya adalah melakukan pemilihan titik knot optimal dengan kombinasi titik knot. Berikut ini adalah nilai GCV minimum yang dihasilkan oleh regresi nonparametrik dengan kombinasi titik knot.

**Tabel 4.5** Nilai GCV dengan Kombinasi Titik Knot

Knot				GCV
$x_1$	$x_2$	$x_3$	$x_4$	
<b>69,51</b>	<b>20,63</b>	<b>89,46</b>	<b>58,56</b>	<b>6,13</b>
<b>79,48</b>	<b>21,65</b>	<b>91,32</b>		
<b>80,47</b>				
69,51	20,64	89,46	62,21	7,57
79,48	21,65	91,32	63,95	
80,48				
69,51	20,63	89,46	60,46	6,76
79,48	21,65	91,32	66,29	
80,47			66,87	

Berdasarkan tabel 4.5 didapatkan nilai GCV yang minimum dengan kombinasi titik knot adalah 6,13.

#### 4.2.2.5 Pemodelan dengan Titik Knot Optimal

Setelah didapatkan nilai GCV minimum dengan menggunakan satu, dua, tiga, dan kombinasi titik knot, selanjutnya adalah melakukan pemilihan model terbaik dengan membandingkan nilai GCV terkecil dari masing- masing knot yang ditunjukkan pada tabel 4.6 berikut.

**Tabel 4.6** Perbandingan Nilai GCV Minimum

Banyaknya Titik Knot	Nilai GCV Minimum
1	8,15
2	8,48
3	7,18
Kombinasi Titik Knot	6,13

Berdasarkan tabel 4.6 didapatkan nilai GCV minimum adalah 6,13, yaitu pada kombinasi titik knot. Hasil ini akan digunakan dalam pemodelan IPG di Provinsi Jawa Timur. Model regresi nonparametrik Spline linier multivariabel yang terbentuk adalah

$$\hat{y}_i = -18,05 + 0,31x_{1i} - 0,81(x_{1i} - 69,51)_+^1 + 4,75(x_{1i} - 79,48)_+^1 - 4,68(x_{1i} - 80,47)_+^1 - 0,54x_{2i} - 0,17(x_{2i} - 20,63)_+^1 + 5,42(x_{2i} - 21,65)_+^1 + 0,81x_{3i} - 3,18(x_{3i} - 89,46)_+^1 + 2,62(x_{3i} - 91,32)_+^1 + 0,54x_{4i} - 1,58(x_{4i} - 58,56)_+^1$$

Model regresi nonparametrik spline linier multivariabel dengan kombinasi titik knot 3, 2, 2, 1 ini memiliki  $R^2$  sebesar 87,48. Nilai  $R^2$  sebesar 87,48 dapat diartikan bahwa model ini dapat menjelaskan IPG sebesar 87,48%.

#### 4.2.2.6 Interval Konfidensi Untuk Parameter Model Dalam Regresi Nonparametrik Spline Linier Multivariabel

Setelah didapatkan model spline linier multivariable terbaik untuk IPG di provinsi Jawa Timur, selanjutnya akan dibangun interval konfidensi untuk parameter model spline linier multivariabel. Berdasarkan hasil yang telah didapatkan pada subbab interval konfidensi untuk parameter regresi nonparametrik spline truncated multivariabel akan dibangun interval konfidensi 95% dengan persamaan sebagai berikut

$$P \left( \hat{\beta}_{ju} - t_{\alpha/2, (n-p(r+1))} \sqrt{\frac{y' \left[ I - \mathbf{X}(\mathbf{K}) (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})' \right] y}{n-p(1+r)} (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))_{jj}^{-1}} \leq \beta_{ju} \leq \hat{\beta}_{ju} + t_{\alpha/2, (n-p(r+1))} \sqrt{\frac{y' \left[ I - \mathbf{X}(\mathbf{K}) (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})' \right] y}{n-p(1+r)} (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))_{jj}^{-1}} \right) = 1 - \alpha$$

Interval konfidensi untuk parameter dapat digunakan untuk mengetahui signifikansi parameter terhadap model. Pengambilan kesimpulan tersebut dilakukan dengan melihat apakah interval konfidensi parameter memuat nilai nol. Jika interval konfidensi memuat nilai nol, maka parameter tidak berpengaruh secara signifikan

terhadap model. Berikut ini adalah hasil interval konfidensi untuk parameter model regresi nonparametrik spline linier multivariabel.

**Tabel 4.7** Interval Konfidensi Parameter Model 95%

Variabel	Parameter	Estimasi Parameter	Batas Bawah	Batas Atas
-	$\hat{\beta}_{01}$	-18,05	-57,68	21,58
$x_1$	$\hat{\beta}_{11}$	0,31	0,07	0,54
	$\hat{\beta}_{12}$	-0,81	-1,29	-0,33
	$\hat{\beta}_{13}$	4,75	1,29	8,20
	$\hat{\beta}_{14}$	-4,68	-8,12	-1,23
$x_2$	$\hat{\beta}_{21}$	-0,55	-0,88	-0,20
	$\hat{\beta}_{22}$	-0,18	-5,41	5,05
	$\hat{\beta}_{23}$	5,42	-1,14	11,98
$x_3$	$\hat{\beta}_{31}$	0,82	0,37	1,26
	$\hat{\beta}_{32}$	-3,19	-6,48	0,10
	$\hat{\beta}_{33}$	2,62	-0,45	5,70
$x_4$	$\hat{\beta}_{41}$	0,54	0,28	0,79
	$\hat{\beta}_{42}$	-1,59	-2,09	-1,07

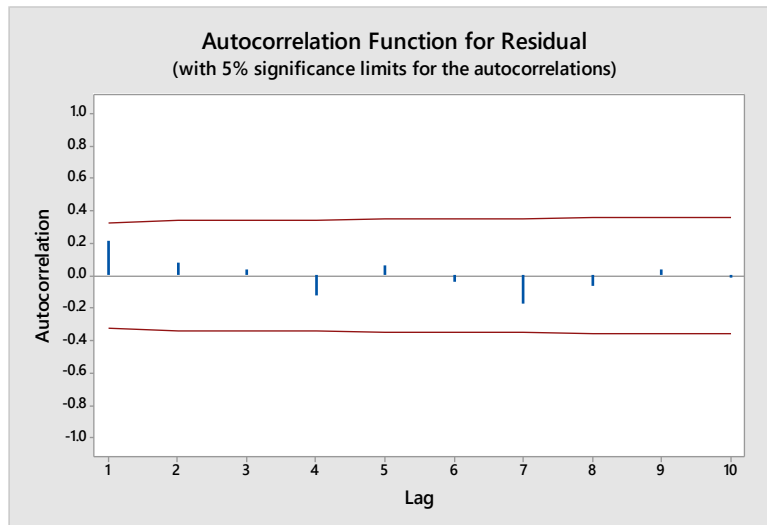
Apabila interval konfidensi memuat nilai nol, maka parameter tidak berpengaruh secara signifikan terhadap model. Berdasarkan Tabel 4.7 didapatkan dari 13 parameter terdapat 8 parameter yang berpengaruh signifikan terhadap model. Namun secara keseluruhan, keempat variabel prediktor berpengaruh secara signifikan terhadap variabel respon.

#### 4.2.2.7 Pemeriksaan Asumsi Residual

Dalam analisis regresi nonparametrik spline, residual hasil analisis regresi harus memenuhi asumsi IIDN. Pemeriksaan asumsi residual yang dilakukan untuk mengetahui apakah residual independen, identik, dan berdistribusi normal.

##### (i) Pemeriksaan Asumsi Independensi Residual

Salah satu cara yang dapat dilakukan untuk mendeteksi asumsi independen residual adalah dengan menggunakan plot ACF dari residual.



**Gambar 4.5** Plot ACF Residual

Berdasarkan Gambar 4.5 dapat dilihat bahwa nilai autokorelasi residual berada pada batas signifikan atau dengan kata lain tidak ada lag yang keluar batas. Sehingga dapat ditarik kesimpulan bahwa tidak ada korelasi antar residual. Asumsi independen terpenuhi.

**(ii) Pemeriksaan Asumsi Identik Residual**

Pemeriksaan asumsi identik digunakan untuk melihat homogenitas dari varians residual. Uji Glejser adalah satu metode yang dapat digunakan untuk mendeteksi adanya heterogenitas varians residual. Hipotesis yang digunakan adalah  $H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 = \sigma^2$  (residual identik)

$H_1$ : minimal ada satu  $\sigma_i^2 \neq \sigma^2, i = 1, 2, \dots, n$  (residual tidak identik)

**Tabel 4.8** ANOVA dari Uji Glejser

Sumber	df	SS	MS	Fhit
Regresi	12	11,32839	0,9440321	0,8804611
Error	25	26,80505	1,072202	
Total	37	38,13344		

Berdasarkan tabel 4.8 didapatkan nilai  $F_{hitung} = 0,88 < F_{tabel} = 2,16$ . Daerah penolakan  $H_0$  yaitu tolak  $H_0$  jika  $F_{hitung} > F_{tabel}$ . Sehingga gagal tolak  $H_0$ , maka dapat ditarik kesimpulan bahwa varians residual homogen. Asumsi identik terpenuhi.



### (iii) Pemeriksaan Asumsi Normalitas Residual

Pengujian asumsi normalitas residual dilakukan untuk memeriksa apakah residual mengikuti distribusi normal atau tidak. Hipotesis yang digunakan untuk pengujian normalitas residual adalah sebagai berikut.

$H_0$ : residual berdistribusi normal

$H_1$ : residual tidak berdistribusi normal

Berdasarkan uji normalitas dengan *Kolmogorov* didapatkan nilai *p-value* sebesar  $0,27 > \alpha = 0,05$  maka gagal tolak  $H_0$ . Sehingga dapat ditarik kesimpulan bahwa residual berdistribusi normal. Asumsi normalitas residual terpenuhi.

### 4.2.3 Interpretasi Model Regresi Nonparametrik Spline Truncated Multivariabel Dan Interval Konfidensi Pada Data IPG di Jawa Timur

Setelah melakukan pemeriksaan asumsi residual, selanjutnya akan dilakukan interpretasi model regresi nonparametrik. Berdasarkan analisis yang telah dilakukan, model terbaik yang terbentuk adalah sebagai berikut.

$$\hat{y}_i = -18,05 + 0,31x_{1i} - 0,81(x_{1i} - 69,51)_+^1 + 4,75(x_{1i} - 79,48)_+^1 - 4,68(x_{1i} - 80,47)_+^1 - 0,54x_{2i} - 0,17(x_{2i} - 20,63)_+^1 + 5,42(x_{2i} - 21,65)_+^1 + 0,81x_{3i} - 3,18(x_{3i} - 89,46)_+^1 + 2,62(x_{3i} - 91,32)_+^1 + 0,54x_{4i} - 1,58(x_{4i} - 58,56)_+^1$$

Interpretasi model terhadap variabel-variabel yang berpengaruh secara signifikan dan interval konfidensi yang didapatkan adalah sebagai berikut:

1. Apabila variabel  $x_2, x_3$ , dan  $x_4$  dianggap konstan, maka pengaruh angka partisipasi sekolah (APS) Sekolah Lanjutan Tingkat Atas (SLTA) penduduk perempuan ( $x_1$ ) terhadap IPG adalah:

$$\hat{y}_i = \begin{cases} 0,31x_{1i} & , x_{1i} < 69,51 \\ -0,5x_{1i} - 56,30 & , 69,51 \leq x_{1i} < 78,48 \\ 5,06x_{1i} - 377,53 & , 78,48 \leq x_{1i} < 80,47 \\ -4,37x_{1i} + 376,59 & , x_{1i} \geq 80,47 \end{cases}$$

Dari model tersebut dapat diinterpretasikan yaitu jika daerah yang memiliki APS SLTA penduduk perempuan kurang dari 69,51%, apabila APS SLTA naik sebesar 1% maka IPG akan cenderung naik 0,31. Daerah yang memiliki APS SLTA penduduk perempuan antara 69,51% hingga 78,48%, apabila APS

SLTA naik sebesar 1% maka IPG akan cenderung turun 0,5. Daerah yang memiliki perilaku tersebut adalah kabupatn Ponorogo, Banyuwangi, Bondowoso, Mojokerto, Jombang, Ngawi, Tuban, Lamongan, Kota Madiun, dan Kota Surabaya. Untuk daerah yang memiliki APS SLTA penduduk perempuan antara 78,48% hingga 80,47%, apabila APS SLTA naik sebesar 1% maka IPG akan cenderung naik 5,06. Sedangkan daerah yang memiliki APS SLTA penduduk perempuan lebih dari 80,47%, apabila APS SLTA naik sebesar 1% maka IPG akan cenderung turun 4,37. Daerah yang memiliki perilaku tersebut adalah Kabupaten Sidoarjo, Madiun, Magetan, Kota Kediri, Blitar, Malang, Mjokerto, Batu.

2. Apabila variabel  $x_1, x_3$ , dan  $x_4$  dianggap konstan, maka pengaruh angka kesakitan penduduk perempuan ( $x_2$ ) terhadap IPG adalah:

$$\hat{y}_i = \begin{cases} -0,54x_{2i} & , x_{2i} < 20,63 \\ -0,71x_{2i} + 3,51 & , 20,63 \leq x_{2i} < 21,65 \\ 4,88x_{2i} - 117,18 & , x_{2i} \geq 21,65 \end{cases}$$

Dari model tersebut dapat diinterpretasikan yaitu jika daerah yang memiliki angka kesakitan penduduk perempuan kurang dari 20,63%, apabila angka kesakitan naik sebesar 1% maka IPG akan cenderung turun 0,54.

3. Apabila variabel  $x_1, x_2$ , dan  $x_4$  dianggap konstan, maka pengaruh penolong terakhir kelahiran oleh medis ( $x_3$ ) terhadap IPG adalah:

$$\hat{y}_i = \begin{cases} 0,81x_{3i} & , x_{3i} < 89,46 \\ -2,37x_{3i} + 284,48 & , 89,46 \leq x_{3i} < 91,32 \\ 3,43x_{3i} - 239,26 & , x_{3i} \geq 91,32 \end{cases}$$

Dari model tersebut dapat diinterpretasikan yaitu jika daerah yang memiliki penolong terakhir kelahiran oleh medis kurang dari 89,46%, apabila penolong terakhir kelahiran oleh medis naik sebesar 1% maka IPG akan cenderung naik 0,81.

4. Apabila variabel  $x_1, x_2$ , dan  $x_3$  dianggap konstan, maka pengaruh Tingkat Partisipasi Angkatan Kerja (TPAK) penduduk perempuan ( $x_4$ ) terhadap IPG adalah:

$$\hat{y}_i = \begin{cases} 0,54x_{4i} & , x_{4i} < 58,56 \\ -1,04x_{4i} + 92,52 & , x_{4i} \geq 58,56 \end{cases}$$

Dari model tersebut dapat diinterpretasikan yaitu jika daerah yang memiliki TPAK penduduk perempuan kurang dari 58,56%, apabila TPAK naik sebesar 1% maka IPG akan cenderung naik 0,54. Sedangkan daerah yang memiliki TPAK penduduk perempuan lebih dari 58,56%, apabila TPAK naik sebesar 1% maka IPG akan cenderung turun 1,04. Daerah yang memiliki perilaku tersebut adalah Kabupaten Pacitan, Ponorogo, Trenggalek, Tulungagung, Sampang, Pamekasan, dan Sumenep.

*(halaman ini sengaja dikosongkan)*

**BAB 5**  
**KESIMPULAN DAN SARAN**

**5.1. Kesimpulan**

Berdasarkan Analisis data yang telah dilakukan dapat ditarik kesimpulan sebagai berikut.

1. Interval konfidensi pada penelitian ini dibangun dengan menggunakan konsep *pivotal quantity* dan interval konfidensi terpendek. Interval konfidensi untuk parameter regresi nonparametrik spline *truncated* multivariabel terbagi menjadi dua yaitu saat saat  $\sigma^2$  diketahui dan  $\sigma^2$  tidak diketahui. *Pivotal quantity* saat  $\sigma^2$  diketahui adalah

$$Z_{ju}(x_1, x_2, \dots, x_n) = \frac{\hat{\beta}_{ju} - \beta_{ju}}{\sqrt{\sigma^2 a_{jj}}}$$

Interval konfidensi saat  $\sigma^2$  diketahui adalah

$$P\left(\hat{\beta}_{ju} - z_{\alpha/2} \sqrt{\sigma^2 a_{jj}} \leq \beta_{ju} \leq \hat{\beta}_{ju} + z_{\alpha/2} \sqrt{\sigma^2 a_{jj}}\right) = 1 - \alpha$$

dengan  $a_{jj}$  merupakan elemen diagonal ke  $jj$  dari  $(\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))^{-1}$

*Pivotal quantity* saat  $\sigma^2$  diketahui adalah

$$T_{ju}(x_1, x_2, \dots, x_n) = \frac{\hat{\beta}_{ju} - \beta_{ju}}{\sqrt{\frac{y' [I - \mathbf{X}(\mathbf{K}) (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})'] y}{n - p(1+r)} (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))_{jj}^{-1}}}$$

Interval konfidensi saat saat  $\sigma^2$  tidak diketahui adalah

$$P\left(\hat{\beta}_{ju} - t_{\alpha/2, (n-p(r+1))} \sqrt{\frac{y' [I - \mathbf{X}(\mathbf{K}) (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})'] y}{n - p(1+r)} (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))_{jj}^{-1}} \leq \beta_{ju} \leq \hat{\beta}_{ju} + t_{\alpha/2, (n-p(r+1))} \sqrt{\frac{y' [I - \mathbf{X}(\mathbf{K}) (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))^{-1} \mathbf{X}(\mathbf{K})'] y}{n - p(1+r)} (\mathbf{X}(\mathbf{K})' \mathbf{X}(\mathbf{K}))_{jj}^{-1}}\right) = 1 - \alpha$$

2. Model regresi nonparametrik spline linier multivariabel terbaik yang terbentuk untuk kasus IPG beserta faktor-faktor yang diduga mempengaruhinya di provinsi Jawa Timur adalah dengan kombinasi titik knot 3, 2, 2, 1 adalah

$$\hat{y}_i = -18,05 + 0,31x_{1i} - 0,81(x_{1i} - 69,51)_+^1 + 4,75(x_{1i} - 79,48)_+^1 - \\ 4,68(x_{1i} - 80,47)_+^1 - 0,54x_{2i} - 0,17(x_{2i} - 20,63)_+^1 + \\ 5,42(x_{2i} - 21,65)_+^1 + 0,81x_{3i} - 3,18(x_{3i} - 89,46)_+^1 + \\ 2,62(x_{3i} - 91,32)_+^1 + 0,54x_{4i} - 1,58(x_{4i} - 58,56)_+^1$$

yang memiliki nilai  $R^2$  sebesar 87,48%. Penentuan variabel prediktor yang berpengaruh secara signifikan terhadap IPG dengan menggunakan interval konfidensi didapatkan APS Sekolah Lanjutan Tingkat Atas (SLTA) penduduk perempuan ( $x_1$ ), angka kesakitan penduduk perempuan ( $x_2$ ), penolong terakhir kelahiran oleh medis ( $x_3$ ), dan tingkat partisipasi angkatan kerja (TPAK) penduduk perempuan ( $x_4$ ) berpengaruh secara signifikan terhadap IPG di provinsi Jawa Timur.

### 5.1. Saran

1. Pada penelitian ini terbatas pada data *cross section*, pada penelitian berikutnya dapat dikembangkan interval konfidensi menggunakan data panel.
2. Pada penelitian ini dibangun interval konfidensi dengan menggunakan konsep *pivotal quantity* dan interval konfidensi terpendek. Untuk penelitian berikutnya dapat dikembangkan interval konfidensi untuk parameter regresi nonparametrik spline *truncated* multivariabel dengan pendekatan *Bayesian*.

## DAFTAR PUSTAKA

- Aryanto, D. (2014). *Pendugaan Area Kecil Terhadap Defisit Kesempatan Kerja Produktif Pada Level Kecamatan di Provinsi Maluku Dengan Pendekatan Empirical Bayes*. Surabaya:Institut Teknologi Sepuluh Nopember.
- Badan Pusat Statistik. (2014). *Indeks Pembangunan Gender 2014*. Jakarta:BPS.
- \_\_\_\_\_. (2014). *Hasil Survey Sosial Ekonomi Nasional Tahun 2014 Provinsi Jawa Timur*. Surabaya:BPS.
- \_\_\_\_\_. (2015). *Pembangunan Manusia Berbasis Gender Provinsi Jawa Timur 2015*. Surabaya:BPS.
- \_\_\_\_\_ (2015). *Profil Gender Provinsi Jawa Timur 2015*. Surabaya:BPS.
- \_\_\_\_\_. (2015). *Laporan Eksekutif Kesehatan Provinsi Jawa Timur 2015*, Surabaya:BPS.
- \_\_\_\_\_, (2016), *Indeks Pembangunan Manusia 2015*, Jakarta:BPS.
- Budiantara, I. N. (2000) *Metode UBR, GML, CV dan GCV dalam Regresi Nonparametrik Spline*. Majalah Ilmiah Himpunan Matematika Indonesia (MIHMI), 6, 285-290.
- Budiantara, I. N. (2004). Model Spline Multivariabel dalam Regresi Nonparametrik. *Makalah Seminar Nasional Matematika*. Surabaya:Institut Teknologi Sepuluh Nopember.
- Budiantara, I. N. (2006). Model Spline dengan Knots Optimal. *Jurnal Ilmu Dasar, FMIPA Universitas Jember, Vol.7, Hal. 77-85*.
- Budiantara, I.N (2009). *Spline dalam Regresi Nonparametrik dan Semiparametrik: Sebuah Pemodelan Statistika Masa Kini dan Masa Mendatang, Pidato Pengukuhan Untuk Jabatan Guru Besar pada Jurusan Statistika, Institut Teknologi Sepuluh Nopember (ITS) Surabaya*. Surabaya: ITS Press.
- Budiantara, I. N., Ratna, M., Zain, I., dan Wibowo, W. (2006). Modeling the Percentage of Poor People in Indonesia Using Spline Nonparametric Regression Approach. *International Journal of Basic & Applied Science IJBAS-IJENS, Vol.12, No.06*.

- Dewi, R. K. (2012). Faktor- Faktor yang Mempengaruhi Angka Gizi Buruk di Jawa Timur dengan Pendekatan Regresi Nonparametrik Spline. *Jurnal Sains dan Seni ITS Vol. 1, No. 1, (Sept. 2012) ISSN : 2301-928X*.
- Eubank, R. L. (1989). *Spline Smoothing and Nonparametric Regression*. New York: Marcel Dekker.
- Gujarati, D. N. (2004). *Basic Econometrics 4<sup>th</sup> edition*. New York: The Mc Graw Hill Companies.
- Gujarati, D. N., dan Porter, D. C. (2015). *Dasar-Dasar Ekonometrika Volume 1, 2nd Edition*. Jakarta: Penerbit Salemba Empat.
- Hafizh, U.Q. (2013). *Pemodelan Disparatis Gender di Jawa Timur dengan Pendekatan Model Regresi Probit Ordinal*. Surabaya: Institut Teknologi Sepuluh Nopember .
- Hakim, L. J. (2014) *Analisis Komponen Indeks Pembangunan Gender Dengan Geographically Weighted Multivariate Regression Model di Provinsi Kalimantan Timur dan Kalimantan Selatan Tahun 2011*. Surabaya: Institut Teknologi Sepuluh Nopember.
- Hardle, W. (1990) *Applied Nonparametrik Regression*. New York: Cambridge University Press.
- Intansari, I. A. S. (2016). *Inferensi Statistik Untuk Kurva Regresi Nonparametrik Spline Kuadratik dan Aplikasinya Pada Data ASFR (Age Specific Fertility Rate) di Bali*. Surabaya: Institut Teknologi Sepuluh Nopember.
- Mao, W. & Zhao, L. H. (2003). Free-knot polynomial splines with confidence intervals. *J.R. Statist.Soc.B., Volume 65, Part 4, pp. 901-919*
- Merdekawati, I. P. (2013) Pemodelan Regresi Spline Truncated Multivariabel pada Faktor- Faktor yang Mempengaruhi Kemiskinan di Kabupaten/ Kota Provinsi Jawa Tengah, *Jurnal Sains dan Seni POMITS Vol. 2, No. 1, (2013) 2337-3520 (2301-92X Print)*.
- Mongomery, D. C., Peck., E. A., and Vining, G. G. (2012). *Introduction to Linier Regression Analysis Fifth Eition*, New Jersey: John Wiley & Sons.
- Nychka, D. (1988). Bayesian confidence intervals for smoothing splines. *J Am. Statist. Ass., 83, 1134-1143*.



- Pratiwi, L. P. T., (2015). *Pengujian Hipotesis Model Spline Truncated Dalam Regresi Nonparametrik Multivariabel (Aplikasi: Data Angka Kematian Ibu di Provinsi Jawa Tengah)*, Surabaya: Institut Teknologi Sepuluh Nopember.
- Setiawan & Kusriani, D. E. (2010). *Ekonometrika*. Yogyakarta: Penerbit ANDI.
- Shao, J. (2008). *Mathematical Statistics (2<sup>nd</sup> ed.)*. New York : Springer.
- Wahba, G. (1983). Bayesian "confidence intervals" for the cross-validated smoothing spline. *J R. Statist. Soc. B*, 45, 133-150.
- Walpole, R. E. (1995). *Pengantar Statistika Edisi ke-3*. Jakarta: PT Gramedia Pustaka Utama.
- Wasono. (2014). *Model Regresi Nonparametrik Multivariabel Heteroskedastik Spline (Studi Kasus Kematian Bayi di Provinsi Jawa Timur)*. Surabaya: Institut Teknologi Sepuluh Nopember.

*(halaman ini sengaja dikosongkan)*

## LAMPIRAN

**Lampiran 1. Data yang Digunakan**

<b>Kota/Kabupaten</b>	<b>IPG</b>	<b>APS SLTA</b>	<b>Angka Kesakitan</b>	<b>Penolong Terakhir Kelahiran Oleh Medis</b>	<b>TPAK P</b>
pacitan	83,76	80,32	12,27	98,86	72,11
ponorogo	93,85	77,37	10,54	96,31	60,35
trenggalek	92,58	68,29	15,05	97,58	61,66
tulungagung	95,11	66,31	15,53	98,65	61,24
blitar	92,81	66,46	16,61	97,02	54,49
kediri	91,98	80,90	12,48	98,29	50,01
malang	87,89	64,53	15,85	93,69	49,32
lumajang	89,08	62,85	11,98	98,43	48,06
jember	83,74	46,89	13,68	91,20	43,58
banyuwangi	85,06	78,04	21,68	94,04	53,77
bondowoso	88,79	73,33	22,46	78,94	54,99
situbondo	86,64	62,55	18,24	91,74	49,51
probolinggo	83,40	47,95	16,56	83,31	54,28
pasuruan	89,95	64,21	19,21	95,22	56,26
sidoarjo	94,20	82,02	9,78	100,00	51,59
mojokerto	90,46	73,32	15,39	99,26	50,43
jombang	89,35	72,39	20,44	99,33	47,37
nganjuk	93,48	64,69	13,88	98,87	50,02
madiun	91,53	87,97	11,22	99,43	54,23
magetan	93,50	88,62	15,15	100,00	57,27
ngawi	92,03	78,31	12,87	100,00	51,85
bojonegoro	89,24	58,83	10,92	95,34	47,43
tuban	87,78	76,24	13,73	98,42	45,92
lamongan	87,21	75,05	18,63	99,60	49,71
gresik	89,01	79,84	12,11	98,53	47,17
bangkalan	85,52	49,79	10,20	74,55	57,82
sampang	82,62	40,60	26,40	69,63	68,90
pamekasan	84,68	55,14	18,35	89,74	67,84
sumenep	76,63	55,89	15,09	77,31	66,35

**Lampiran 1. Data yang Digunakan (Lanjutan 1)**

<b>Kota/Kabupaten</b>	<b>IPG</b>	<b>APS SLTA</b>	<b>Angka Kesakitan</b>	<b>Penolong Terakhir Kelahiran Oleh Medis</b>	<b>TPAK P</b>
kota kediri	95,15	89,45	12,70	100,00	55,98
kota blitar	98,23	83,14	12,86	97,63	52,58
kota malang	94,99	80,92	15,49	100,00	51,48
kota probolinggo	96,74	66,58	14,47	97,76	52,97
kota pasuruan	96,30	81,10	10,85	99,42	53,65
kota Mojokerto	93,27	83,97	12,32	100,00	52,94
kota Madiun	92,81	75,05	13,69	99,48	52,27
kota surabaya	93,65	69,71	15,91	96,31	53,26
kota batu	89,22	84,70	15,08	100,00	53,89

## Lampiran 2. Program GCV Spline *Truncated* Linier Satu Knot

```
GCV1=function(para)
{
data=read.table("f://THESIS//VALIDASI1//ipgp.txt",header=TRUE)
data=as.matrix(data)
p=length(data[,1])
q=length(data[1,])
m=ncol(data)-para-1
dataA=data[(para+2):q]
F=matrix(0,nrow=p,ncol=p)
diag(F)=1
nk= length(seq(min(data[,2]),max(data[,2]),length.out=50))
knot1=matrix(ncol=m,nrow=nk)
for (i in (1:m))
{
for (j in (1:nk))
{
a=seq(min(dataA[,i]),max(dataA[,i]),length.out=50)
knot1[j,i]=a[j]
}
}
a1=length(knot1[,1])
knot1=knot1[2:(a1-1),]
aa=rep(1,p)
data1=matrix(ncol=m,nrow=p)
data2=data[,2:q]
a2=nrow(knot1)
GCV=rep(NA,a2)
Rsqr=rep(NA,a2)
for (i in 1:a2)
{
for (j in 1:m)
{
for (k in 1:p)
{
if (data[k,(j+para+1)]<knot1[i,j]) data1[k,j]=0 else
data1[k,j]=data[k,(j+para+1)]-knot1[i,j]
}
}
}
mx=cbind(aa,data2,data1)
mx=as.matrix(mx)
C=pinv(t(mx)%*%mx)
B=C%*%(t(mx)%*%data[,1])
}
```

## Lampiran 2. Program GCV Spline *Truncated* Linier Satu Knot (Lanjutan)

```
yhat=mx%*%B
SSE=0
SSR=0
for (r in (1:p))
{
sum=(data[r,1]-yhat[r,])^2
sum1=(yhat[r,]-mean(data[,1]))^2
SSE=SSE+sum
SSR=SSR+sum1
}
Rsqr[i]=(SSR/(SSE+SSR))*100
MSE=SSE/p
A=mx%*%C%*%t(mx)
A1=(F-A)
A2=(sum(diag(A1))/p)^2
GCV[i]=MSE/A2
}
GCV=as.matrix(GCV)
Rsqr=as.matrix(Rsqr)
cat("=====", "\n")
cat("Nilai Knot dengan Spline linear 1 knot", "\n")
cat("=====", "\n")
print (knot1)
cat("=====", "\n")
cat("Rsqr dengan Spline linear 1 knot", "\n")
cat("=====", "\n")
print (Rsqr)
cat("=====", "\n")
cat("HASIL GCV dengan Spline linear 1 knot", "\n")
cat("=====", "\n")
print (GCV)
s1=min(GCV)
print(max(Rsqr))
cat("=====", "\n")
cat("HASIL GCV terkecil dengan Spline linear 1 knot", "\n")
cat("=====", "\n")
cat(" GCV =",s1, "\n")
write.csv(GCV,file="f://THESIS//VALIDASI1//validasi_GCV.csv")
write.csv(Rsqr,file="f://THESIS//VALIDASI1//validasi_Rsqr.csv")
write.csv(knot1,file="f://THESIS//VALIDASI1//validasi_knot.csv")
}
GCV1(0)
```

### Lampiran 3. Program GCV Spline *Truncated* Linier Dua Knot

```
GCV2=function(para)
{
data=read.table("f://THESIS//VALIDASI1//ipgp.txt", header=FALSE)
data=as.matrix(data)
p=length(data[,1])
q=length(data[1,])
m=ncol(data)-1
F=matrix(0,nrow=p,ncol=p)
diag(F)=1
nk= length(seq(min(data[,2]),max(data[,2]),length.out=50))
knot=matrix(ncol=m,nrow=nk)
for (i in (1:m))
{
for (j in (1:nk))
{
a=seq(min(data[,i+1]),max(data[,i+1]),length.out=50)
knot[j,i]=a[j]
}
}
z=(nk*(nk-1)/2)
knot2=cbind(rep(NA,(z+1)))
for (i in (1:m))
{
knot1=rbind(rep(NA,2))
for ( j in 1:(nk-1))
{
for (k in (j+1):nk)
{
xx=cbind(knot[j,i],knot[k,i])
knot1=rbind(knot1,xx)
}
}
knot2=cbind(knot2,knot1)
}
knot2=knot2[2:(z+1),2:(2*m+1)]
aa=rep(1,p)
data2=matrix(ncol=(2*m),nrow=p)
data1=data[,2:q]
a1=length(knot2[,1])
GCV=rep(NA,a1)
Rsqr=rep(NA,a1)
for (i in 1:a1)
```

### Lampiran 3. Program GCV Spline *Truncated* Linier Dua Knot (Lanjutan)

```
{
for (j in 1:(2*m))
{
if (mod(j,2)==1) b=floor(j/2)+1 else b=j/2
for (k in 1:p)
{
if (data1[k,b]<knot2[i,j]) data2[k,j]=0 else data2[k,j]=data1[k,b]-knot2[i,j]
}
}
}
mx=cbind(aa,data1,data2)
mx=as.matrix(mx)
C=pinv(t(mx)%*%mx)
B=C%*%(t(mx)%*%data[,1])
yhat=mx%*%B
SSE=0
SSR=0
for (r in (1:p))
{
sum=(data[r,1]-yhat[r,])^2
sum1=(yhat[r,]-mean(data[,1]))^2
SSE=SSE+sum
SSR=SSR+sum1
}
Rsqr[i]=(SSR/(SSE+SSR))*100
MSE=SSE/p
A=mx%*%C%*%t(mx)
A1=(F-A)
A2=(sum(diag(A1))/p)^2
GCV[i]=MSE/A2
}
GCV=as.matrix(GCV)
Rsqr=as.matrix(Rsqr)
cat("=====\n")
cat("Nilai Knot dengan Spline linear 2 knot","\n")
cat("=====\n")
print (knot2)
cat("=====\n")
cat("Rsqr dengan Spline linear 2 knot","\n")
```



### Lampiran 3. Program GCV Spline *Truncated* Linier Dua Knot (Lanjutan)

```
cat("=====\n")
print (Rsqr)
cat("=====\n")
cat("HASIL GCV dengan Spline linear 2 knot", "\n")
cat("=====\n")
print (GCV)
s1=min(GCV)
cat("=====\n")
cat("HASIL GCV terkecil dengan Spline linear 2 knot", "\n")
cat("=====\n")
cat(" GCV =", s1, "\n")
write.csv(GCV, file="f://THESIS//VALIDASI1//validasi_GCV2.csv")
write.csv(Rsqr, file="f://THESIS//VALIDASI1//validasi_Rsqr2.csv")
write.csv(knot2, file="f://THESIS//VALIDASI1//validasi_knot2.csv")
}
GCV2(0)
```

#### Lampiran 4. Program GCV Spline *Truncated* Linier Tiga Knot

```
GCV3=function(para)
{
data=read.table("f://THESIS//VALIDASI1//ipgp.txt",header=FALSE)
data=as.matrix(data)
p=length(data[,1])
q=length(data[1,])
m=ncol(data)-para-1
F=matrix(0,nrow=p,ncol=p)
dataA=data[(para+2):q]
diag(F)=1
nk= length(seq(min(data[,2]),max(data[,2]),length.out=50))
knot=matrix(ncol=m,nrow=nk)
for (i in (1:m))
{
for (j in (1:nk))
{
a=seq(min(dataA[,i]),max(dataA[,i]),length.out=50)
knot[j,i]=a[j]
}
}
knot=knot[2:(nk-1),]
a2=nrow(knot)
z=(a2*(a2-1)*(a2-2)/6)
knot1=cbind(rep(NA,(z+1)))
for (i in (1:m))
{
knot2=rbind(rep(NA,3))
for (j in 1:(a2-2))
{
for (k in (j+1):(a2-1))
{
for (g in (k+1):a2)
{
xx=cbind(knot[j,i],knot[k,i],knot[g,i])
knot2=rbind(knot2,xx)
}
}
}
}
knot1=cbind(knot1,knot2)
}
knot1=knot1[2:(z+1),2:(3*m+1)]
aa=rep(1,p)
```

#### Lampiran 4. Program GCV Spline *Truncated* Linier Tiga Knot (Lanjutan)

```
data1=matrix(ncol=(3*m),nrow=p)
data2=data[, (para+2):q]
a1=length(knot1[,1])
GCV=rep(NA,a1)
Rsqr=rep(NA,a1)
for (i in 1:a1)
{
for (j in 1:ncol(knot1))
{
b=ceiling(j/3)
for (k in 1:p)
{
if (data2[k,b]<knot1[i,j]) data1[k,j]=0 else data1[k,j]=data2[k,b]-knot1[i,j]
}
}
mx=cbind(aa,data[,2:q],data1)
mx=as.matrix(mx)
C=pinv(t(mx)%*%mx)
B=C%*%(t(mx)%*%data[,1])
yhat=mx%*%B
SSE=0
SSR=0
for (r in (1:p))
{
sum=(data[r,1]-yhat[r,])^2
sum1=(yhat[r,]-mean(data[,1]))^2
SSE=SSE+sum
SSR=SSR+sum1
}
Rsqr[i]=(SSR/(SSE+SSR))*100
MSE=SSE/p
A=mx%*%C%*%t(mx)
A1=(F-A)
A2=(sum(diag(A1))/p)^2
GCV[i]=MSE/A2
}
GCV=as.matrix(GCV)
Rsqr=as.matrix(Rsqr)
cat("=====", "\n")
cat("Nilai Knot dengan Spline linear 3 knot", "\n")
cat("=====", "\n")
print (knot1)
```

#### Lampiran 4. Program GCV Spline *Truncated* Linier Tiga Knot (Lanjutan)

```
cat("=====", "\n")
cat("Rsq dengan Spline linear 3 knot", "\n")
cat("=====", "\n")
print (Rsq)
r=max(Rsq)
print (r)
cat("=====", "\n")
cat("HASIL GCV dengan Spline linear 3 knot", "\n")
cat("=====", "\n")
print (GCV)
s1=min(GCV)
cat("=====", "\n")
cat("HASIL GCV terkecil dengan Spline linear 3 knot", "\n")
cat("=====", "\n")
cat(" GCV =",s1, "\n")
write.csv(GCV,file="f://THESIS//VALIDASI1//validasi_GCV3.csv")
write.csv(Rsq,file="f://THESIS//VALIDASI1//validasi_Rsq3.csv")
write.csv(knot1,file="f://THESIS//VALIDASI1//validasi_knot3.csv")
}
GCV3(0)
```

## Lampiran 5. Program GCV Kombinasi Spline *Truncated* Linier

```
GCVkom=function(para)
{
data=read.table("f://THESIS//VALIDASI1//ipgp.txt",header=FALSE)
data=as.matrix(data)
p1=length(data[,1])
q1=length(data[1,])
v=para+2
F=matrix(0,nrow=p1,ncol=p1)
diag(F)=1
x1=read.table("f://THESIS//VALIDASI1//1.txt")
x2=read.table("f://THESIS//VALIDASI1//2.txt")
x3=read.table("f://THESIS//VALIDASI1//3.txt")
x4=read.table("f://THESIS//VALIDASI1//4.txt")
n2=nrow(x1)
a=matrix(nrow=4,ncol=3^4)
m=0
for (ii in 1:3)
for (j in 1:3)
for (k in 1:3)
for (l in 1:3)
{
m=m+1
a[,m]=c(ii,j,k,l)
}
a=t(a)
GCV=matrix(nrow=nrow(x1),ncol=3^4)
R=matrix(nrow=nrow(x1),ncol=3^4)
for (i in 1:3^4)
{
for (h in 1:nrow(x1))
{
if (a[i,1]==1)
{
gab=as.matrix(x1[,1])
gen=as.matrix(data[,v])
aa=matrix(nrow=nrow(x1)*nrow(data),ncol=1)
for (j in 1:1)
for (w in 1:nrow(data))
{
if (gen[w,j]<gab[h,j]) aa[w,j]=0 else aa[w,j]=gen[w,j]-gab[h,j]
}
}
}
}
```

### Lampiran 5. Program GCV Kombinasi Spline *Truncated* Linier (Lanjutan)

```
else
if (a[i,1]==2)
{
gab=as.matrix(x1[,2:3])
gen=as.matrix(cbind(data[,v],data[,v]))
aa=matrix(nrow=nrow(x1)*nrow(data),ncol=2)
for (j in 1:2)
for (w in 1: nrow(data))
{
if (gen[w,j]<gab[h,j]) aa[w,j]=0 else aa[w,j]=gen[w,j]-gab[h,j]
}
}
}
else
{
gab=as.matrix(x1[,4:6])
gen=as.matrix(cbind(data[,v],data[,v],data[,v]))
aa=matrix(nrow=nrow(x1)*nrow(data),ncol=3)
for (j in 1:3)
for (w in 1:nrow(data))
{
if (gen[w,j]<gab[h,j]) aa[w,j]=0 else aa[w,j]=gen[w,j]-gab[h,j]
}
}
}
if (a[i,2]==1)
{
gab=as.matrix(x2[,1] )
gen=as.matrix(data[,v+1])
bb=matrix(nrow=nrow(x1)*nrow(data),ncol=1)
for (j in 1:1)
for (w in 1:nrow(data))
{
if (gen[w,j]<gab[h,j]) bb[w,j]=0 else bb[w,j]=gen[w,j]-gab[h,j]
}
}
}
else
if (a[i,2]==2)
{
gab=as.matrix(x2[,2:3] )
gen=as.matrix(cbind(data[,v+1],data[,v+1]))
bb=matrix(nrow=nrow(x1)*nrow(data),ncol=2)
for (j in 1:2)
for (w in 1:nrow(data))
```

### Lampiran 5. Program GCV Kombinasi Spline *Truncated* Linier (Lanjutan)

```
{
if (gen[w,j]<gab[h,j]) bb[w,j]=0 else bb[w,j]=gen[w,j]-gab[h,j]
}
}
else
{
gab=as.matrix(x2[,4:6])
gen=as.matrix(cbind(data[,v+1],data[,v+1],data[,v+1]))
bb=matrix(nrow=nrow(x1)*nrow(data),ncol=3)
for (j in 1:3)
for (w in 1:nrow(data))
{
if (gen[w,j]<gab[h,j]) bb[w,j]=0 else bb[w,j]=gen[w,j]-gab[h,j]
}
}
if (a[i,3]==1)
{
gab=as.matrix(x3[,1] )
gen=as.matrix(data[,v+2])
cc=matrix(nrow=nrow(x1)*nrow(data),ncol=1)
for (j in 1:1)
for (w in 1:nrow(data))
{
if (gen[w,j]<gab[h,j]) cc[w,j]=0 else cc[w,j]=gen[w,j]-gab[h,j]
}
}
else
if (a[i,3]==2)
{
gab=as.matrix(x3[,2:3] )
gen=as.matrix(cbind(data[,v+2],data[,v+2]))
cc=matrix(nrow=nrow(x1)*nrow(data),ncol=2)
for (j in 1:2)
for (w in 1:nrow(data))
{
if (gen[w,j]<gab[h,j]) cc[w,j]=0 else cc[w,j]=gen[w,j]-gab[h,j]
}
}
}
else
{
gab=as.matrix(x3[,4:6])
gen=as.matrix(cbind(data[,v+2],data[,v+2],data[,v+2]))
}
```

### Lampiran 5. Program GCV Kombinasi Spline *Truncated* Linier (Lanjutan)

```
cc=matrix(nrow=nrow(x1)*nrow(data),ncol=3)
for (j in 1:3)
for (w in 1:nrow(data))
{
if (gen[w,j]<gab[h,j]) cc[w,j]=0 else cc[w,j]=gen[w,j]-gab[h,j]
}
}
if (a[i,4]==1)
{
gab=as.matrix(x4[,1] )
gen=as.matrix(data[,v+3])
dd=matrix(nrow=nrow(x1)*nrow(data),ncol=1)
for (j in 1:1)
for (w in 1:nrow(data))
{
if (gen[w,j]<gab[h,j]) dd[w,j]=0 else dd[w,j]=gen[w,j]-gab[h,j]
}
}
else
if (a[i,4]==2)
{
gab=as.matrix(x4[,2:3] )
gen=as.matrix(cbind(data[,v+3],data[,v+3]))
dd=matrix(nrow=nrow(x1)*nrow(data),ncol=2)
for (j in 1:2)
for (w in 1:nrow(data))
{
if (gen[w,j]<gab[h,j]) dd[w,j]=0 else dd[w,j]=gen[w,j]-gab[h,j]
}
}
else
{
gab=as.matrix(x4[,4:6])
gen=as.matrix(cbind(data[,v+3],data[,v+3],data[,v+3]))
dd=matrix(nrow=nrow(x1)*nrow(data),ncol=3)
for (j in 1:3)
for (w in 1:nrow(data))
{
if (gen[w,j]<gab[h,j]) dd[w,j]=0 else dd[w,j]=gen[w,j]-gab[h,j]
}
}
}
ma=as.matrix(cbind(aa,bb,cc,dd))
```



### Lampiran 5. Program GCV Kombinasi Spline *Truncated* Linier (Lanjutan)

```
mx=cbind(rep(1,nrow(data)), data[,2:q1],na.omit(ma))
mx=as.matrix(mx)
C=pinv(t(mx)%*%mx)
B=C%*%(t(mx)%*%data[,1])
yhat=mx%*%B
SSE=0
SSR=0
for (r in 1:nrow(data))
{
sum=(data[r,1]-yhat[r,])^2
sum1=(yhat[r,]-mean(data[,1]))^2
SSE=SSE+sum
SSR=SSR+sum1
}
Rsqr=(SSR/(SSE+SSR))*100
R[i]=Rsqr
MSE=SSE/p1
A=mx%*%C%*%t(mx)
A1=(F-A)
A2=(sum(diag(A1))/p1)^2
GCV[h,i]=MSE/A2
}
if (a[i,1]==1) sp=x1[,1] else
if (a[i,1]==2) sp=x1[,2:3] else
sp=x1[,4:6]
if (a[i,2]==1) spl=x2[,1] else
if (a[i,2]==2) spl=x2[,2:3] else
spl=x2[,4:6]
if (a[i,3]==1) splin=x3[,1] else
if (a[i,3]==2) splin=x3[,2:3] else
splin=x3[,4:6]
if (a[i,4]==1) spline=x4[,1] else
if (a[i,4]==2) spline=x4[,2:3] else
spline=x4[,4:6]
kkk=cbind(sp,spl,splin,spline)
cat("=====", "\n")
print(i)
print(kkk)
print(Rsq)
}
write.csv(GCV,file="f://THESIS//VALIDASI1//output GCV kombinasi.csv")
}
```

## Lampiran 6. Program Estimasi Parameter Spline *Truncated Linier*

```
uji=function(alpha,para)
{
data=read.table("f://THESIS//VALIDASI1//ipgp.txt",header=FALSE)
knot=read.table("f://THESIS//VALIDASI1//knot.txt",header=FALSE)
data=as.matrix(data)
knot=as.matrix(knot)
ybar=mean(data[,1])
m=para+2
p=nrow(data)
q=ncol(data)
dataA=cbind(data[,m],data[,m],data[,m],data[,m+1],data[,m+1],data[,m+2],data[,
m+2],data[,m+3])
dataA=as.matrix(dataA)
satu=rep(1,p)
n1=ncol(knot)
data.knot=matrix(ncol=n1,nrow=p)
for (i in 1:n1)
{
for(j in 1:p)
{
if (dataA[j,i]<knot[1,i]) data.knot[j,i]=0 else data.knot[j,i]=dataA[j,i]-knot[1,i]
}
}
mx=cbind(satu,
data[,2],data.knot[,1:3],data[,3],data.knot[,4:5],data[,4],data.knot[,6:7],data[,5],dat
a.knot[,8])
mx=as.matrix(mx)
B=(pinv(t(mx)%*%mx))%*%t(mx)%*%data[,1]
cat("=====", "\n")
cat("Estimasi Parameter", "\n")
cat("=====", "\n")
print (B)
n1=nrow(B)
yhat=mx%*%B
res=data[,1]-yhat
SSE=sum((data[,1]-yhat)^2)
SSR=sum((yhat-ybar)^2)
SST=SSR+SSE
MSE=SSE/(p-n1)
MSR=SSR/(n1-1)
Rsqr=(SSR/(SSR+SSE))*100
```

## Lampiran 6. Program Estimasi Parameter Spline *Truncated* Linier (Lanjutan)

```

#uji F (uji serentak)
Fhit=MSR/MSE
pvalue=pf(Fhit,(n1-1),(p-n1),lower.tail=FALSE)
if (pvalue<=alpha)
{
cat("-----", "\n")
cat("Kesimpulan hasil uji serentak", "\n")
cat("-----", "\n")
cat("Tolak Ho yakni minimal terdapat 1 prediktor yang signifikan", "\n")
cat("", "\n")
}
else
{
cat("-----", "\n")
cat("Kesimpulan hasil uji serentak", "\n")
cat("-----", "\n")
cat("Gagal Tolak Ho yakni semua prediktor tidak berpengaruh signifikan", "\n")
cat("", "\n")
}

#uji t (uji individu)
thit=rep(NA,n1)
pval=rep(NA,n1)
SE=sqrt(diag(MSE*(pinv(t(mx)%*%mx))))
cat("-----", "\n")
cat("Kesimpulan hasil uji individu", "\n")
cat("-----", "\n")
thit=rep(NA,n1)
pval=rep(NA,n1)
for (i in 1:n1)
{
thit[i]=B[i,1]/SE[i]
pval[i]=2*(pt(abs(thit[i]),(p-n1),lower.tail=FALSE))
if (pval[i]<=alpha) cat("Tolak Ho yakni prediktor signifikan dengan
pvalue", pval[i], "\n") else cat("Gagal tolak Ho yakni prediktor tidak signifikan
dengan pvalue", pval[i], "\n")
}
thit=as.matrix(thit)
cat("=====", "\n")
cat("nilai t hitung", "\n")
cat("=====", "\n")
print (thit)

```

### Lampiran 6. Program Estimasi Parameter Spline *Truncated* Linier (Lanjutan)

```
cat("Analysis of Variance", "\n")
cat("=====", "\n")
cat("Sumber    df    SS    MS    Fhit", "\n")
cat("Regresi    ", (n1-1), " ", SSR, " ", MSR, " ", Fhit, "\n")
cat("Error      ", p-n1, " ", SSE, " ", MSE, "\n")
cat("Total      ", p-1, " ", SST, "\n")
cat("=====", "\n")
cat("s=", sqrt(MSE), " Rsq=", Rsq, "\n")
cat("pvalue(F)=", pvalue, "\n")
write.csv(res, file="f://THESIS//VALIDASI1//output uji residual.csv")
write.csv(pval, file="f://THESIS//VALIDASI1//output uji pvalue.csv")
write.csv(mx, file="f://THESIS//VALIDASI1//output uji mx.csv")
write.csv(yhat, file="f://THESIS//VALIDASI1//output uji yhat.csv")
write.csv(B, file="f://THESIS//VALIDASI1//output B.csv")
write.csv(thit, file="f://THESIS//VALIDASI1//output thit.csv")
}
uji(0.05, 0)
```

## Lampiran 7. Program Interval Konfidensi Spline *Truncated Linier*

```
interval=function(alpha,para)
{
data=read.table("f://THESIS//VALIDASI1//ipgp.txt",header=FALSE)
knot=read.table("f://THESIS//VALIDASI1//knot.txt",header=FALSE)
data=as.matrix(data)
knot=as.matrix(knot)
ybar=mean(data[,1])
m=para+2
p=nrow(data)
q=ncol(data)
dataA=cbind(data[,m],data[,m],data[,m],data[,m+1],data[,m+1],data[,m+2],data[,
m+2],data[,m+3])
dataA=as.matrix(dataA)
satu=rep(1,p)
n1=ncol(knot)
data.knot=matrix(ncol=n1,nrow=p)
for (i in 1:n1)
{
for(j in 1:p)
{
if (dataA[j,i]<knot[1,i]) data.knot[j,i]=0 else data.knot[j,i]=dataA[j,i]-knot[1,i]
}
}
mx=cbind(satu,
data[,2],data.knot[,1:3],data[,3],data.knot[,4:5],data[,4],data.knot[,6:7],data[,5],dat
a.knot[,8])
mx=as.matrix(mx)
B=(pinv(t(mx)%*%mx))%*%t(mx)%*%data[,1]
cat("=====", "\n")
cat("Estimasi Parameter", "\n")
cat("=====", "\n")
print (B)
n1=nrow(B)
yhat=mx%*%B
res=data[,1]-yhat
SSE=sum((data[,1]-yhat)^2)
SSR=sum((yhat-ybar)^2)
SST=SSR+SSE
MSE=SSE/(p-n1)
MSR=SSR/(n1-1)
Rsq=(SSR/(SSR+SSE))*100
```

**Lampiran 7. Program Interval Konfidensi Spline *Truncated* Linier (Lanjutan)**

```

#uji F (uji serentak)
Fhit=MSR/MSE
pvalue=pf(Fhit,(n1-1),(p-n1),lower.tail=FALSE)
if (pvalue<=alpha)
{
cat("-----", "\n")
cat("Kesimpulan hasil uji serentak", "\n")
cat("-----", "\n")
cat("Tolak Ho yakni minimal terdapat 1 prediktor yang signifikan", "\n")
cat("", "\n")
}
else
{
cat("-----", "\n")
cat("Kesimpulan hasil uji serentak", "\n")
cat("-----", "\n")
cat("Gagal Tolak Ho yakni semua prediktor tidak berpengaruh signifikan", "\n")
cat("", "\n")
}

#uji t (uji individu)
thit=rep(NA,n1)
pval=rep(NA,n1)
SE=sqrt(diag(MSE*(pinv(t(mx)%*%mx))))
cat("-----", "\n")
cat("Kesimpulan hasil uji individu", "\n")
cat("-----", "\n")
thit=rep(NA,n1)
pval=rep(NA,n1)
for (i in 1:n1)
{
thit[i]=B[i,1]/SE[i]
pval[i]=2*(pt(abs(thit[i]),(p-n1),lower.tail=FALSE))
if (pval[i]<=alpha) cat("Tolak Ho yakni prediktor signifikan dengan
pvalue", pval[i], "\n") else cat("Gagal tolak Ho yakni prediktor tidak signifikan
dengan pvalue", pval[i], "\n")
}
thit=as.matrix(thit)
cat("=====", "\n")
cat("nilai t hitung", "\n")
cat("=====", "\n")
print (thit)

```

**Lampiran 7. Program Interval Konfidensi Spline *Truncated* Linier (Lanjutan)**

```

cat("Analysis of Variance","\n")
cat("=====", "\n")
cat("Sumber      df      SS      MS      Fhit","\n")
cat("Regresi      ,(n1-1)," ",SSR," ",MSR,"",Fhit,"\n")
cat("Error        ",p-n1," ",SSE,"",MSE,"\n")
cat("Total        ",p-1," ",SST,"\n")
cat("=====", "\n")
cat("s=",sqrt(MSE)," Rsq=",Rsq,"\n")
cat("pvalue(F)=" ,pvalue,"\n")

#interval konfidensi
ba=rep(NA,n1)
bb=rep(NA,n1)
SE=sqrt(diag(MSE*(pinv(t(mx)%*%mx))))
for(i in 1:n1)
{
ba[i]=B[i,1]+(2.05*SE[i])
bb[i]=B[i,1]-(2.05*SE[i])
}
ba=as.matrix(ba)
bb=as.matrix(bb)
ik=cbind(bb,ba)
cat("----interval konfidensi----","\n")
cat("--batas bawah  batas atas--","\n")
print(ik)
write.csv(bb,file="f://THESIS//VALIDASI1//output bb.csv")
write.csv(ba,file="f://THESIS//VALIDASI1//output ba.csv")
}
interval(0.05,0)

```

## Lampiran 8. Program Uji Glejser

```
glejser=function(alpha,para)
{
data=read.table("f://THESIS//VALIDASI1//ipgp.txt")
knot=read.table("f://THESIS//VALIDASI1//knot.txt",header=FALSE)
res=read.table("f://THESIS//VALIDASI1//resi.txt")
data=as.matrix(data)
knot=as.matrix(knot)
res=abs(res)
res=as.matrix(res)
rbar=mean(res)
m=para+2
p=nrow(data)
q=ncol(data)
dataA=cbind(data[,m],data[,m],data[,m],data[,m+1],data[,m+1],data[,m+2],data[,
m+2],data[,m+3])
dataA=as.matrix(dataA)
satu=rep(1,p)
n1=ncol(knot)
data.knot=matrix(ncol=n1,nrow=p)
for (i in 1:n1)
{
for(j in 1:p)
{
if (dataA[j,i]<knot[1,i]) data.knot[j,i]=0 else data.knot[j,i]=dataA[j,i]-knot[1,i]
}
}
mx=cbind(satu,
data[,2],data.knot[,1:3],data[,3],data.knot[,4:5],data[,4],data.knot[,6:7],data[,5],dat
a.knot[,8])
mx=as.matrix(mx)
B=(pinv(t(mx)%*%mx))%*%t(mx)%*%res
n1=nrow(B)
yhat=mx%*%B
SSE=sum((res-yhat)^2)
SSR=sum((yhat-rbar)^2)
SST=SSR+SSE
MSE=SSE/(p-n1)
MSR=SSR/(n1-1)
Rsq=(SSR/SST)*100

#uji F (uji serentak)
Fhit=MSR/MSE
```



### Lampiran 8. Program Uji Glejser (Lanjutan)

```
pvalue=pf(Fhit,(n1-1),(p-n1),lower.tail=FALSE)
if (pvalue<=alpha)
{
cat("-----", "\n")
cat("Kesimpulan hasil uji serentak", "\n")
cat("-----", "\n")
cat("Tolak Ho yakni minimal terdapat 1 prediktor yang signifikan atau terjadi
heteroskedastisitas", "\n")
cat("", "\n")
}
else
{
cat("-----", "\n")
cat("Kesimpulan hasil uji serentak", "\n")
cat("-----", "\n")
cat("Gagal Tolak Ho yakni semua prediktor tidak berpengaruh signifikan atau
tidak terjadi heteroskedastisitas", "\n")
cat("", "\n")
}
cat("Analysis of Variance", "\n")
cat("=====", "\n")
cat("Sumber      df      SS      MS      Fhit", "\n")
cat("Regresi      ,(n1-1), " ",SSR, " ",MSR, """,Fhit, "\n")
cat("Error        ",p-n1, " ",SSE, """,MSE, "\n")
cat("Total        ",p-1, " ",SST, "\n")
cat("=====", "\n")
cat("s=",sqrt(MSE), " Rsq=",Rsq, "\n")
cat("pvalue(F)=",pvalue, "\n")
}
glejser(0.05,0)
```

### Lampiran 9. Program Uji Normalitas

```
resid=read.table("f://THESIS//VALIDASI1//resi.txt")
ks.test(resid, "pnorm")
```

*(halaman ini sengaja dikosongkan)*

## BIOGRAFI PENULIS



Penulis lahir di Kota Jember, Provinsi Jawa Timur pada tanggal 4 Februari 1993 dengan nama lengkap Rifani Nur Sindy Setiawan, sebagai anak dari pasangan Alex Setiawan dan Eni Suryanti. Penulis menempuh pendidikan formal di TK Dharmawanita Puger Wetan, SD Negeri Puger Kulon 2, SMP Negeri 1 Puger dan SMA Negeri 2 Jember. Penulis kemudian melanjutkan jenjang S1 di Prodi Statistika Universitas Airlangga. Penulis melanjutkan studi ke jenjang S2 di Program Pascasarjana Statistika FMIPA Institut Teknologi Sepuluh Nopember Surabaya dengan jalur mandiri.

Saran, kritik, dan pertanyaan seputar tesis ini dapat disampaikan ke alamat email [setiawanrifani93@gmail.com](mailto:setiawanrifani93@gmail.com).

*(halaman ini sengaja dikosongkan)*