



TUGAS AKHIR – SS141501

**PEMODELAN JUMLAH KASUS TUBERKULOSIS
DI PROVINSI JAWA BARAT MENGGUNAKAN
REGRESI BINOMIAL NEGATIF DAN REGRESI
NONLINIER**

**DWI PUSPITA FIRDAUS
NRP 1313 100 023**

**Dosen Pembimbing
Dra. Wiwiek Setya Winahju, M.S.
Dr. Irhamah, S.Si, M.Si.**

**PROGRAM STUDI SARJANA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA 2017**



TUGAS AKHIR – SS141501

**PEMODELAN JUMLAH KASUS TUBERKULOSIS
DI PROVINSI JAWA BARAT MENGGUNAKAN
REGRESI BINOMIAL NEGATIF DAN REGRESI
NONLINIER**

**DWI PUSPITA FIRDAUS
NRP 1313 100 023**

**Dosen Pembimbing
Dra. Wiwiek Setya Winahju, M.S.
Dr. Irhamah, S.Si, M.Si.**

**PROGRAM STUDI SARJANA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA 2017**



FINAL PROJECT – SS141501

**MODELLING FOR THE NUMBER OF
TUBERCULOSIS IN THE PROVINCE OF WEST
JAVA USING NEGATIVE BINOMIAL REGRESSION
AND NONLINEAR REGRESSION**

**DWI PUSPITA FIRDAUS
NRP 1313 100 023**

**Supervisor
Dra. Wiwiek Setya Winahju, M.S.
Dr. Irhamah, S.Si, M.Si.**

**UBDERGRADUATE PROGRAMME
DEPARTMENT OF STATISTICS
FACULTY OF MATHEMATICS AND NATURAL SCIENCE
SEPULUH NOPEMBER INSTITUTE OF TECHNOLOGY
SURABAYA 2017**

LEMBAR PENGESAHAN

**PEMODELAN JUMLAH KASUS TUBERKULOSIS
DI PROVINSI JAWA BARAT MENGGUNAKAN
REGRESI BINOMIAL NEGATIF DAN REGRESI
NONLINIER**

TUGAS AKHIR

Diajukan Untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Sains
pada
Program Studi Sarjana Departemen Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Institut Teknologi Sepuluh Nopember

Oleh :
DWI PUSPITA FIRDAUS
NRP. 1313 100 023

Disetujui oleh Pembimbing:
Dra. Wiwiek Setya Winahju, M.S.
NIP. 19560424 198303 2 001

(*Wiwiek Setya Winahju*)

Dr. Irhamah, S.Si., M.Si.
NIP. 19780406 200112 2 002

(*Irhamah*)



Mengetahui,
Kepala Departemen

[Signature]
Dr. Suhartono
19710929 199512 1 001

SURABAYA, JULI 2017

PEMODELAN JUMLAH KASUS TUBERKULOSIS DI PROVINSI JAWA BARAT REGRESI BINOMIAL NEGATIF DAN REGRESI NONLINIER

Nama Mahasiswa : Dwi Puspita Firdaus
NRP : 1313 100 023
Departemen : Statistika
Dosen Pembimbing 1 : Dra. Wiwiek Setya Wnahju, M. S.
Dosen Pembimbing 2 : Dr. Irhamah, S.Si., M.Si.

Abstrak

*Tuberkulosis adalah penyakit radang parenkim paru yang menu-
lar karena infeksi kuman TB, yaitu Mycobacterium tuberculosis. Preva-
lensi tuberkulosis di Indonesia menempati urutan ketiga setelah negara
India dan Cina. Di Indonesia, jumlah kasus tuberkulosis terbanyak ta-
hun 2015 terdapat pada Provinsi Jawa Barat, yaitu sebanyak 31.231 ka-
sus. Jumlah kasus tuberkulosis merupakan salah satu contoh data count.
Salah satu metode yang dapat digunakan untuk menganalisis data count
adalah regresi Poisson, tetapi regresi Poisson sering terjadi kasus over-
dispersi. Salah satu metode yang dapat digunakan untuk mengatasi
overdispersi adalah regresi Binomial Negatif. Selain itu pada penelitian
sering ditemukan kasus data yang memiliki pola nonlinier pada model
regresi, sehingga perlu juga dilakukan pemodelan menggunakan regresi
nonlinier. Pemodelan menggunakan regresi Binomial Negatif didapati-
kan hasil bahwa model terbaik dengan nilai AIC terkecil sebesar 447,04
terdapat pada kombinasi enam variabel, serta didapatkan tiga variabel
yang berpengaruh signifikan terhadap jumlah kasus tuberkulosis, yaitu
jumlah keluarga miskin, rata-rata konsumsi protein perkapita perhari,
dan kepadatan penduduk. Kemudian pemodelan menggunakan regresi
nonlinier didapatkan hasil bahwa model yang tepat adalah model
dengan variabel signifikan yang memiliki nilai MSE dan S terkecil.
Hasil perbandingan ketiga model regresi didapatkan kesimpulan, model
regresi yang tepat digunakan untuk memodelkan jumlah kasus tuber-
kulosis di Jawa Barat adalah regresi Nonlinier.*

**Kata kunci : Jawa Barat, Overdispersi, Regresi Binomial Negatif,
Regresi Nonlinier, Tuberkulosis.**

(Halaman ini sengaja dikosongkan)

MODELLING FOR THE NUMBER OF TUBERCULOSIS IN WEST JAVA USING NEGATIVE BINOMIAL REGRESSION AND NONLINEAR REGRESSION

Name : Dwi Puspita Firdaus
Student Number : 1313 100 023
Department : Statistics
Supervisor 1 : Dra. Wiwiek Setya Wnahju, M. S.
Supervisor 2 : Dr. Irhamah, S.Si., M.Si.

Abstract

Tuberculosis is a lung parenchymal inflammatory disease that spreads due to infection with TB germ, Mycobacterium tuberculosis. The prevalence of tuberculosis in Indonesia ranks third after India and China. In Indonesia, the highest number of tuberculosis in 2015 was found in the Province of West Java, which is 31.231 cases. The number of tuberculosis is an example of count data. A method that can be used to analyze the count data is Poisson regression, but it often occurs overdispersion. A method that can be used to overcome the overdispersion is Binomial Negative regression. In addition, the research is often found cases of data that has a nonlinear pattern. Therefore its is also necessary to do the modeling using nonlinear regression. Modeling using Binomial Negative regression was obtained that the best model with the lowest AIC of 447,04 were in combination of six variables and there were three variables that have a significant effect to the number of tuberculosis, namely the number of poor families, the average protein consumption percapita perday, and population density. The modeling using nonlinear regression found that the appropriate model is using 2 significant variables which has the lowest MSE. The comparison between three regression models is that the right nonlinear regression is the appropriate regression used to model the number of tuberculosis in West Java.

Keywords : Binomial Negative Regression, Nonlinear Regression, Overdispersion, Tuberculosis, West Java.

(Halaman ini sengaja dikosongkan)

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan atas kehadiran Allah SWT, serta shalawat dan salam selalu tercurah pada Nabi Muhammad SAW atas suri tauladannya sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul **“Pemodelan Jumlah Kasus Tuberkulosis di Provinsi Jawa Barat Menggunakan Regresi Binomial Negatif dan Regresi Nonlinier”**. terselesaikannya Tugas Akhir ini tidak lepas dari bantuan berbagai pihak, penulis ingin mengucapkan terima kasih kepada:

1. Bapak Dr. Suhartono selaku Kepala Departemen Statistika FMIPA ITS.
2. Ibu Dra. Wiwiek Setya Winahju, M.S. dan Dr. Irhamah, S.Si, M.Si selaku dosen pembimbing yang telah meluangkan waktunya untuk memberikan bimbingan, arahan, dan masukan selama penyusunan Tugas Akhir ini.
3. Bapak Dr. Sutikno, S.Si, M.Si selaku kaprodi S1 Departemen Statistika dan dosen penguji yang telah memberikan saran dan kritik demi kesempurnaan Tugas Akhir ini.
4. Bapak Prof. Drs. Nur Iriawan, M.Ikom, Ph.D selaku dosen wali dan dosen penguji yang telah memberikan saran dan kritik demi kesempurnaan Tugas Akhir ini.
5. Para dosen pengajar Departemen Statistika ITS atas segala ilmu yang diberikan, serta seluruh staf dan karyawan atas kerja keras dan bantuannya selama ini.
6. Kedua orangtua yang telah memberikan doa dan semangat serta dukungan.

Penulis berharap semoga Tugas Akhir ini bermanfaat untuk menambah wawasan keilmuan bagi semua pihak, serta kritik dan saran diharapkan dari semua pihak untuk tahap pengembangan selanjutnya.

Surabaya, Juli 2017

Penulis

(Halaman ini sengaja dikosongkan)

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
TITLE PAGE	iii
LEMBAR PENGESAHAN	v
ABSTRAK	vii
ABSTRACT	ix
KATA PENGANTAR	xi
DAFTAR ISI	xiii
DAFTAR GAMBAR	xv
DAFTAR TABEL	xvii
DAFTAR LAMPIRAN	xix
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
1.3 Tujuan.....	4
1.4 Manfaat.....	4
1.5 Batasan Masalah.....	5
BAB II TINJAUAN PUSTAKA	7
2.1 Statistika Deskriptif.....	7
2.2 Multikolinieritas.....	8
2.3 Regresi Poisson.....	8
2.4 Estimasi Parameter Regresi Poisson.....	9
2.5 Pengujian Parameter Regresi Poisson.....	10
2.6 Pemeriksaan Equidispersi.....	11
2.7 Regresi Binomial Negatif.....	12
2.8 Estimasi Parameter Regresi Binomial Negatif.....	15
2.9 Pengujian Parameter Regresi Binomial Negatif.....	15
2.10 Uji Deteksi Nonlinier.....	16
2.11 Regresi Nonlinier.....	17
2.12 Estimasi Parameter Regresi Nonlinier.....	18
2.13 Pengujian Parameter Regresi Nonlinier.....	19
2.14 AIC.....	20
2.15 Tuberkulosis.....	20

BAB III METODOLOGI PENELITIAN	23
3.1 Sumber Data.....	23
3.2 Kerangka Konseptual	23
3.3 Variabel Penelitian	24
3.4 Langkah Analisis.....	26
3.5 Diagram Alir	28
BAB IV ANALISIS DAN PEMBAHASAN	31
4.1 Karakteristik Jumlah Kasus Tuberkulosis di Jawa Barat ..	31
4.2 Pemodelan Banyaknya Kasus Tuberkulosis di Jawa Barat Menggunakan Regresi Poisson	33
4.3 Pemodelan Banyaknya Kasus Tuberkulosis di Jawa Barat Menggunakan Regresi Binomial Negatif	37
4.4 Pemodelan Banyaknya Kasus Tuberkulosis di Jawa Barat Menggunakan Regresi Nonlinier.....	40
4.4.1 Regresi Nonlinier Menggunakan 8 Variabel Prediktor	42
4.4.2 Regresi Nonlinier Menggunakan Variabel Signifikan	44
4.4.3 Perbandingan Model Regresi Nonlinier	47
4.5 Pemilihan Model Terbaik.....	47
BAB V KESIMPULAN DAN SARAN	49
5.1 Kesimpulan	49
5.2 Saran.....	50
DAFTAR PUSTAKA	51
LAMPIRAN	55

DAFTAR GAMBAR

	Halaman
Gambar 3.1 Kerangka Konseptual Penelitian.....	23
Gambar 3.2 Diagram Alir Penelitian	28
Gambar 4.1 Jumlah Kasus Tuberkulosis di Kabupaten/Kota di Jawa Barat.....	31
Gambar 4.2 <i>Scatter Plot</i> Residual dengan Fits Model Regresi Binomial Negatif.....	40
Gambar 4.3 <i>Scatter Plot</i> antara Variabel Respon dan Variabel Prediktor	41
Gambar 4.4 <i>Scatter Plot</i> Residual dengan Fits Model Regresi Nonlinier Menggunakan 8 Variabel Prediktor.....	44
Gambar 4.5 <i>Scatter Plot</i> Residual dengan Fits Model Regresi Nonlinier Menggunakan Variabel Signifikan.....	46

(Halaman ini sengaja dikosongkan)

DAFTAR TABEL

	Halaman
Tabel 2.1 Beberapa Contoh Model Regresi Nonlinier	17
Tabel 3.1 Variabel Penelitian	24
Tabel 3.2 Struktur Data.....	26
Tabel 4.1 Statistika Deskriptif Variabel Penelitian	32
Tabel 4.2 Koefisien Korelasi antar Variabel Prediktor.....	33
Tabel 4.3 Nilai VIF dari Variabel Prediktor	34
Tabel 4.4 Estimasi Parameter Model Regresi Poisson	35
Tabel 4.5 <i>Deviance</i> Model Regresi Poisson	37
Tabel 4.6 Kemungkinan Model Regresi Binomial Negatif dari Kombinasi Variabel Prediktor	38
Tabel 4.7 Estimasi Parameter Regresi Binomial Negatif	38
Tabel 4.8 Hasil Pengujian Terasvirta.....	41
Tabel 4.9 Nilai Inilisiasi Parameter dengan Seluruh Variabel....	42
Tabel 4.10 Estimasi Parameter Regresi Nonlinier dengan 8 Variabel	42
Tabel 4.11 Nilai Inilisiasi Parameter dengan Dua Variabel Signifikan	45
Tabel 4.12 Estimasi Parameter Regresi Nonlinier dengan Variabel Signifikan.....	45
Tabel 4.13 Perbandingan Model Regresi Nonlinier	47
Tabel 4.14 Pemilihan Model Terbaik	48

(Halaman ini sengaja dikosongkan)

DAFTAR LAMPIRAN

	Halaman
Lampiran 1. Jumlah Kasus Tuberkulosis di Provinsi Jawa Barat Tahun 2015.....	55
Lampiran 2. Variabel Prediktor Penelitian	56
Lampiran 3. Output Koefisien Korelasi.....	58
Lampiran 4. Output Multikolinieritas (VIF).....	59
Lampiran 5. Syntax SAS untuk Regresi Poisson.....	60
Lampiran 6. Output SAS untuk Regresi Poisson.....	61
Lampiran 7. Macro SAS untuk Overdispersi	63
Lampiran 8. Output SAS untuk Overdispersi.....	64
Lampiran 9. Syntax dan Output R Regresi Binomial Negatif dengan 7 Variabel Prediktor.....	65
Lampiran 10. Syntax dan Output R Regresi Binomial Negatif dengan 6 Variabel Prediktor.....	66
Lampiran 11. Syntax dan Output R Regresi Binomial Negatif dengan 5 Variabel Prediktor.....	67
Lampiran 12. Syntax dan Output R Regresi Binomial Negatif dengan 4 Variabel Prediktor.....	68
Lampiran 13. Syntax dan Output R Regresi Binomial Negatif dengan 3 Variabel Prediktor.....	69
Lampiran 14. Syntax dan Output R Regresi Binomial Negatif dengan 2 Variabel Prediktor.....	70
Lampiran 15. Syntax dan Output R Regresi Binomial Negatif dengan 1 Variabel Prediktor.....	71
Lampiran 16. Output Regresi Linier Y dengan 8 Variabel	72
Lampiran 17. Macro SAS Regresi Nonlinier (8 Variabel).....	73
Lampiran 18. Output SAS Regresi Nonlinier (8 Variabel)	74
Lampiran 19. Output Regresi Linier Y dengan 2 Variabel Signifikan	75

Lampiran 20. Macro SAS Regresi Nonlinier (2 Variabel Signifikan)	76
Lampiran 21. Output SAS Regresi Nonlinier (2 Variabel Signifikan)	77
Lampiran 22. Surat Pernyataan Legalitas Data.....	78

BAB I

PENDAHULUAN

1.1 Latar Belakang

Tuberkulosis merupakan salah satu penyakit menular mematikan yang sampai saat ini masih menjadi perhatian dunia. Berdasarkan laporan *World Health Organization* (WHO) tahun 2016, terdapat 10,4 juta orang menderita penyakit tuberkulosis, dimana 1,8 juta diantaranya meninggal karena tuberkulosis. Prevalensi tuberkulosis di Indonesia menempati urutan ketiga setelah India dan Cina dengan jumlah sebanyak 700 ribu kasus (WHO, 2013). Kemudian dalam *Global Tuberculosis Report*, WHO (2016) menyatakan bahwa pada tahun 2015 ditemukan sebanyak 330.729 kasus tuberkulosis di Indonesia, namun Indonesia masih tetap berada pada peringkat ketiga negara dengan kasus tuberkulosis terbanyak. Tuberkulosis merupakan penyakit menular langsung yang menyerang berbagai organ serta jaringan tubuh yang disebabkan oleh bakteri *Mycrobacterium Tuberculosis* (Soedarto,2009). Bakteri ini berbentuk batang dan bersifat tahan asam, sehingga dikenal juga sebagai Batang Tahan Asam (BTA). Kuman ini dapat menyerang semua bagian tubuh manusia, tetapi yang paling sering terkena adalah organ paru. Seseorang yang terinfeksi kuman TB belum tentu mengidap penyakit tuberkulosis karena kuman TB tersebut tidak aktif. Kuman tersebut dapat menjadi aktif apabila seseorang memiliki kekebalan tubuh yang lemah.

Kasus tuberkulosis pada tahun 2015 di Indonesia jumlahnya mencapai angka 330.729 kasus. Angka tersebut meningkat jika dibandingkan dengan tahun 2014, yaitu sebesar 324.539 kasus. Jumlah kasus tertinggi terdapat pada provinsi dengan jumlah penduduk yang tinggi, yaitu Jawa Barat, Jawa Timur, dan Jawa Tengah. Kasus tuberkulosis di tiga provinsi tersebut merupakan 38% dari seluruh kasus di Indonesia. Provinsi Jawa Barat merupakan provinsi dengan jumlah kasus tuberkulosis tertinggi di Indonesia sejak tahun 2012. Pada tahun 2015, jumlah kasus tuberkulosis di Jawa Barat mencapai angka 31.231 kasus (Depkes, 2015). Jumlah kasus tuberkulosis di Jawa Barat ini perlu menjadi perhatian karena me-

rupakan peringkat pertama jumlah kasus tuberkulosis terbanyak dalam lima tahun terakhir.

Jumlah kasus tuberkulosis merupakan salah satu contoh data *count* yang mengikuti distribusi Poisson, sehingga untuk mengetahui faktor-faktor yang mempengaruhi secara signifikan dapat digunakan regresi Poisson. Salah satu asumsi pada regresi Poisson adalah equidisersi, dimana nilai varian sama dengan rata-ratanya (Famoye, Wulu, dan Singh, 2004). Jika asumsi tersebut tidak terpenuhi atau dengan kata lain terdapat kondisi *under/over* dispersi dalam data yang dimodelkan, maka bisa mengakibatkan dugaan regresi Poisson menjadi kurang akurat. Pada umumnya, sering ditemui data diskrit dengan varian lebih besar dibandingkan dengan rata-ratanya atau disebut dengan overdispersi, sehingga pemodelan dengan regresi Poisson tidak dapat digunakan. Metode yang dapat digunakan untuk mengatasi masalah overdispersi pada data *count* salah satunya adalah regresi Binomial Negatif. Pemodelan dengan menggunakan regresi Binomial Negatif dapat digunakan karena metode ini tidak mengharuskan terpenuhinya asumsi equidisersi atau nilai rata-rata sama dengan nilai varian seperti pada regresi Poisson.

Pada penelitian juga sering ditemukan kasus data yang memiliki pola nonlinier pada model regresi, sehingga perlu digunakan model regresi nonlinier dengan asumsi kelinieran ditolak yang disebabkan oleh adanya data yang menyimpang dari pola garis linier. Beberapa contoh model regresi non linier adalah ekponensial, *invers*, *compound*, *power*, kuadratik, dan kubik. Untuk mendapatkan model regresi nonlinier terbaik dapat digunakan kriteria nilai S dan MSE. Model regresi nonlinier terbaik merupakan model yang memiliki nilai S dan MSE paling kecil.

Penelitian dengan menggunakan metode regresi Binomial Negatif pernah dilakukan oleh Saraswati (2016) yang memodelkan faktor-faktor yang memengaruhi kematian ibu di Kota Surabaya dan didapatkan hasil bahwa regresi Binomial Negatif menghasilkan model dengan AIC yang lebih kecil dibandingkan menggunakan regresi Poisson. Selain itu, Fatmasari (2012) juga melakukan penelitian menggunakan regresi Binomial Negatif yang memodelkan kasus DBD di Kota Malang. Hasil yang didapatkan adalah nilai

AIC terkecil didapatkan dari kombinasi lima variabel prediktor yang mempengaruhi jumlah kasus DBD di Kota Malang.

Penelitian dengan menggunakan metode regresi nonlinier pernah dilakukan oleh Jayadewa (2012) yang memodelkan biaya tak langsung proyek konstruksi di PT. Wijaya Karya. Hasil yang didapatkan adalah perbandingan antara regresi linier, regresi nonlinier menggunakan Algoritma *Leverberg-Marquardt*, dan regresi nonlinier menggunakan Algoritma Genetika, model yang paling sesuai untuk memodelkan biaya tak langsung proyek konstruksi adalah menggunakan regresi nonlinier dengan Algoritma Genetika karena menghasilkan MSE yang paling kecil. Selain itu, pernah dilakukan oleh Harifuddin (2007) yang memodelkan kebutuhan daya listrik Sulawesi Selatan. Hasil yang diperoleh adalah model nonlinier paling tepat untuk digunakan memodelkan kebutuhan daya listrik Sulawesi Selatan adalah model parabola.

Penelitian yang membahas mengenai jumlah kasus tuberkulosis pernah dilakukan oleh Indahwati (2016) di Kota Surabaya tahun 2014 menggunakan metode GWNBR yang menghasilkan kesimpulan bahwa kelompok yang terbentuk berdasarkan variabel signifikan sebanyak dua kelompok, dimana kelompok pertama merupakan kelompok yang dipengaruhi oleh variabel kepadatan penduduk, dan persentase penderita HIV, sedangkan kelompok kedua merupakan kelompok yang dipengaruhi oleh variabel rumah tangga berperilaku hidup bersih dan sehat. Selain itu, Meilania (2016) memodelkan faktor-faktor yang mempengaruhi jumlah penderita tuberkulosis di Kabupaten Probolinggo menggunakan *Generalized Poisson Regression* (GPR). Hasil yang didapatkan dari penelitian tersebut adalah faktor-faktor yang mempengaruhi jumlah penderita tuberkulosis adalah persentase rumah tangga ber-PHBS, persentase penduduk yang memiliki akses sanitasi layak, persentase rumah sehat, persentase penduduk miskin, dan kepadatan penduduk.

Faktor-faktor yang mempengaruhi jumlah kasus tuberkulosis sangat beragam, sehingga untuk mengetahui faktor-faktor risiko yang berpengaruh signifikan terhadap jumlah tuberkulosis tersebut, peneliti ingin mendapatkan hubungan antara jumlah kasus tuberkulosis yang terjadi di Jawa Barat dengan variabel-variabel prediktor yang diduga mempengaruhi dengan cara mendapatkan mo-

del hubungan yang terbaik menggunakan analisis regresi Binomial Negatif dan regresi nonlinier berdasarkan kriteria nilai AIC yang terkecil.

1.2 Rumusan Masalah

Tuberkulosis merupakan penyakit ancaman global yang serius dari kelompok penyakit menular. Indonesia menempati peringkat ketiga negara dengan prevalensi tuberkulosis terbanyak di dunia setelah India dan Cina pada tahun 2013. Tuberkulosis merupakan penyebab kematian terbanyak kedua di Indonesia. Jumlah kasus tuberkulosis tertinggi di Indonesia pada tahun 2015 terdapat di provinsi dengan jumlah penduduk yang tinggi, yaitu Provinsi Jawa Barat. Faktor-faktor yang mempengaruhi kasus tuberkulosis sangat beragam, seperti faktor ekonomi, lingkungan, dan kesehatan, sehingga perlu dilakukan analisis yang juga mempertimbangkan faktor-faktor tersebut. Berdasarkan permasalahan tersebut, akan dilakukan pemodelan jumlah kasus tuberkulosis di Provinsi Jawa Barat menggunakan metode regresi Binomial Negatif dan regresi nonlinier, serta membandingkan kedua metode regresi tersebut berdasarkan kriteria AIC terkecil.

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang dipaparkan, tujuan yang ingin dicapai dari penelitian ini adalah untuk mendeskripsikan karakteristik Provinsi Jawa Barat jika ditinjau dari penderita tuberkulosis dan mengetahui faktor faktor yang berpengaruh signifikan terhadap jumlah kasus tuberkulosis di Jawa Barat menggunakan regresi Binomial Negatif dan regresi non linier, serta mendapatkan model terbaik untuk memodelkan jumlah kasus tuberkulosis di Jawa Barat tahun 2015 berdasarkan nilai AIC terkecil .

1.4 Manfaat Penelitian

Hasil penelitian ini diharapkan mampu memberikan tambahan informasi terkait faktor-faktor yang mempengaruhi jumlah kasus tuberkulosis sehingga dapat dijadikan bahan pertimbangan oleh

pemerintah, khususnya Provinsi Jawa Barat dalam pengambilan kebijakan guna meminimalisir jumlah kasus tuberkulosis di Jawa Barat. Selain itu, penelitian ini diharapkan dapat menjadi bahan referensi untuk penelitian-penelitian selanjutnya.

1.5 Batasan Masalah

Batasan masalah yang digunakan dalam penelitian ini adalah jumlah kasus tuberkulosis di Provinsi Jawa Barat tahun 2015 dengan unit penelitian 27 kabupaten/kota.

(Halaman ini sengaja dikosongkan)

BAB II TINJAUAN PUSTAKA

2.1 Statistika Deskriptif

Statistika deskriptif merupakan metode-metode yang berkaitan dengan pengumpulan dan penyajian suatu gugus data sehingga memberikan informasi yang berguna (Walpole, 1995). Data dapat didefinisikan menjadi dua, yaitu dideskripsikan ke dalam grafik atau tabel dan dideskripsikan secara numerik. Mendeskripsikan data dalam bentuk grafik dapat menggunakan histogram. Histogram merupakan sebuah representasi grafik yang menampilkan impresi visual dari distribusi data atau sebuah estimasi distribusi probabilitas dari variabel kontinyu. Histogram terdiri dari frekuensi tabular, ditunjukkan sebagai balok yang berdekatan, didirikan sepanjang interval yang berlainan, dengan luas yang sama dengan frekuensi dari observasi di dalam interval. Tinggi dari balok juga sama dengan densitas frekuensi dari interval, yaitu frekuensi yang dibagi oleh lebar dari interval. Keseluruhan luas atau area dari histogram sama dengan jumlah data yang ada.

Mendeskripsikan data secara numerik yaitu dengan menggunakan angka, dalam hal ini dapat meliputi ukuran pemusatan data atau ukuran penyebaran data. Salah satu dari ukuran penyebaran data adalah varian, sedangkan salah satu dari ukuran pemusatan data adalah *mean*. *Mean* atau rata-rata dapat didefinisikan sebagai jumlah seluruh data dibagi dengan banyaknya data (Walpole, 1995).

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \quad (2.1)$$

Varian adalah suatu angka yang menunjukkan ukuran variabilitas yang dapat dihitung dengan cara mengkuadratkan standar deviasi (Walpole, 1995).

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1} \quad (2.2)$$

2.2 Multikolinieritas

Multikolinieritas berarti terjadi korelasi yang kuat atau hampir sempurna antar variabel bebas. Tepatnya multikolinieritas berkenaan dengan adanya lebih dari satu hubungan linier pasti dan istilah kolinieritas berkenaan dengan terdapatnya satu hubungan linier. Pemeriksaan adanya masalah multikolinieritas dapat dilakukan dengan beberapa metode, diantaranya (Draper dan Smith, 1992) :

a. Koefisien Korelasi

Multikolinieritas dapat diduga dari tingginya nilai korelasi antara variabel bebas, yaitu dapat diduga dari tingginya nilai korelasi antar variabel bebas dengan melihat nilai dari koefisien korelasi sederhana yang cukup tinggi ($0,9 \leq R \leq 1,0$).

b. Dengan melihat nilai VIF (*Variance Inflation Factor*)

VIF multikolinieritas dalam variabel bebas dapat diperoleh berdasarkan fakta bahwa kenaikan varian tergantung dari σ^2 dan VIF itu sendiri. VIF dinyatakan dengan rumus sebagai berikut.

$$VIF = \frac{1}{1 - R_j^2} \quad (2.3)$$

dimana :

VIF = 1 : mengindikasikan tidak ada korelasi yang signifikan antar variabel prediktor.

VIF > 1 : mengindikasikan adanya korelasi antar variabel prediktor.

VIF > 5-10 : mengindikasikan ada salah satu variabel prediktor yang merupakan fungsi dari variabel prediktor yang lain.

(Draper dan Smith, 1992).

2.3 Regresi Poisson

Regresi Poisson merupakan salah satu jenis analisis regresi yang digunakan untuk menganalisis variabel respon bertipe diskrit dan integer tidak negatif atau data *count*. Metode ini biasanya diterapkan pada penelitian kesehatan masyarakat, biologi, dan teknik, dimana observasi dari variabel responnya (y) berupa cacahan objek yang merupakan fungsi dari sejumlah karakteristik tertentu (x).

Probabilitas dari y “banyaknya suatu kejadian” yang berdistribusi Poisson, yaitu (Agresti, 2002) :

$$P(y; \mu) = \frac{e^{-\mu} \mu^y}{y!}, y = 0, 1, 2, \dots \quad (2.4)$$

dengan μ adalah rata-rata jumlah kejadian yang berdistribusi Poisson.

Distribusi Poisson digunakan untuk memodelkan kejadian yang relatif jarang terjadi selama periode waktu yang dipilih. Pada model regresi Poisson, fungsi yang digunakan adalah log yaitu $\ln(\mu_i) = \eta_i$, sehingga fungsi hubungan untuk model regresi Poisson jika X_1, X_2, \dots, X_k adalah variabel bebas mempunyai persamaan.

$$\ln(\mu_i) = \beta_0 + \sum_{j=1}^k \beta_{1j} X_{ij} \quad (2.5)$$

$$\mu_i = \exp\left(\beta_0 + \sum_{j=1}^k \beta_{1j} X_{ij}\right) \quad (2.6)$$

Distribusi data jumlahan yang paling sederhana adalah Poisson. Distribusi Poisson juga termasuk dalam keluarga eksponensial. Hal ini ditunjukkan oleh Agresti (2002), fungsi peluang densitasnya sebagai berikut.

$$P(y; \mu) = \frac{e^{-\mu} \mu^y}{y!} = \frac{\exp(-\mu) \exp(y \ln(\mu))}{y!} \quad (2.7)$$

Berdasarkan persamaan (2.7), maka dapat dikatakan bahwa $\theta = \mu$, $a(\mu) = \exp(-\mu)$, $b(y) = \frac{1}{y!}$, dan $Q(\mu) = \ln(\mu)$. Oleh karena itu terbukti bahwa Poisson termasuk dalam keluarga eksponensial.

2.4 Estimasi Parameter Regresi Poisson

Penaksiran parameter model regresi Poisson dilakukan dengan menggunakan metode *Maximum Likelihood Estimation* yaitu dengan cara memaksimumkan fungsi *likelihood* (Cameron dan Trivedi, 1998). Fungsi *likelihood* dari regresi Poisson yaitu.

$$\ln L(\boldsymbol{\beta}) = \ln \left(\prod_{i=1}^n \frac{\exp(-\mu_i) \mu_i^{y_i}}{y_i!} \right)$$

$$\begin{aligned}
&= \sum_{i=1}^n \ln \left(\frac{\exp(-\mu_i) \mu_i^{y_i}}{y_i!} \right) \\
&= \sum_{i=1}^n (\ln(e^{-\mu_i}) + \ln(\mu_i^{y_i}) - \ln(y_i!)) \\
&= \sum_{i=1}^n (-\mu_i + y_i \ln(\mu_i) - \ln(y_i!)) \\
&= \sum_{i=1}^n \left(-e^{\mathbf{x}_i^T \boldsymbol{\beta}} + y_i \ln \left(e^{\mathbf{x}_i^T \boldsymbol{\beta}} \right) - \ln(y_i!) \right) \\
&= -\sum_{i=1}^n e^{\mathbf{x}_i^T \boldsymbol{\beta}} + \sum_{i=1}^n y_i \mathbf{x}_i^T \boldsymbol{\beta} - \sum_{i=1}^n \ln(y_i!) \quad (2.8)
\end{aligned}$$

Taksiran maksimum *likelihood* untuk parameter β_k dinyatakan dengan $\hat{\beta}_k$ yang merupakan penyelesaian turunan pertama fungsi logaritma natural dari *likelihood*. Selanjutnya, $\ln L(\boldsymbol{\beta})$ diturunkan terhadap $\boldsymbol{\beta}^T$ menjadi turunan kedua.

$$\frac{\partial \ln L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T} = -\sum_{i=1}^n \mathbf{x}_i \exp(\mathbf{x}_i^T \boldsymbol{\beta}) + \sum_{i=1}^n y_i \mathbf{x}_i \quad (2.9)$$

Kemudian disamadengankan nol untuk mendapatkan penyelesaiannya, tetapi penyelesaian dengan cara tersebut sering kali tidak mendapatkan hasil yang eksplisit sehingga alternatif yang bisa digunakan untuk mendapatkan penyelesaian dari MLE adalah dengan metode iterasi numerik Newton-Raphson.

2.5 Pengujian Parameter Regresi Poisson

Pengujian kelayakan model regresi Poisson, terlebih dahulu ditentukan dua buah fungsi *likelihood* yang meliputi $L(\hat{\Omega})$ yaitu nilai *likelihood* untuk model lengkap dengan melibatkan variabel prediktor dan $L(\hat{\omega})$ yaitu nilai *likelihood* untuk model sederhana tanpa melibatkan variabel prediktor. Salah satu metode yang digunakan untuk menentukan statistik uji dalam pengujian parameter model regresi Poisson adalah dengan menggunakan metode *Maximum Likelihood Ratio Test* (MLRT) dengan hipotesis.

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$

H_1 : paling sedikit ada satu $\beta_j \neq 0$ dengan $j = 1, 2, \dots, k$

Dengan statistik uji sebagai berikut.

$$G^2 = 2[\ln L(\hat{\Omega}) - \ln L(\hat{\omega})] \sim \chi_k^2 \quad (2.10)$$

Keputusan yang akan diambil adalah tolak H_0 jika $G^2 > \chi^2_{\alpha, k}$ dengan k adalah banyaknya parameter model di bawah populasi dikurangi dengan banyaknya parameter dibawah H_0 . Parameter model regresi Poisson yang telah dihasilkan dari estimasi parameter belum tentu mempunyai pengaruh yang signifikan terhadap model. Untuk itu perlu dilakukan pengujian terhadap parameter model regresi Poisson secara individu dengan menggunakan hipotesis sebagai berikut.

$H_0: \beta_j = 0$ (pengaruh variabel ke- j tidak signifikan)

$H_1: \beta_j \neq 0$ (pengaruh variabel ke- j signifikan)

Statistik uji yang digunakan adalah:

$$Z = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \quad (2.11)$$

$(SE(\hat{\beta}_j))$ dapat dari elemen diagonal ke $(j+1)$ dari $[-H^{-1}(\boldsymbol{\beta})]$, dimana:

$$H^{-1}(\boldsymbol{\beta}) = \frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \quad (2.12)$$

Keputusan yang akan diambil adalah tolak H_0 jika $|Z_{hit}| > Z_{\alpha/2}$, dimana α adalah tingkat signifikansi yang digunakan.

2.6 Pemeriksaan Equidispersi

Metode regresi Poisson mewajibkan equidispersi, yaitu kondisi dimana nilai rata-rata dan varians dari variabel respon bernilai sama. *Overdispersion* adalah kondisi dimana data variabel respon menunjukkan nilai varians lebih besar dari nilai rata-ratanya (Darnah, 2011). *Underdispersion* adalah kondisi dimana data variabel respon menunjukkan nilai varians lebih kecil dari nilai rata-ratanya. *Overdispersion* ataupun *underdispersion* akan menghasilkan nilai devians model menjadi sangat besar sehingga model yang dihasilkan kurang tepat. Apabila model Poisson diaplikasikan untuk data yang terjadi *overdispersi*, maka dapat mengakibatkan *standard error* menjadi *underestimate*, sehingga beberapa variabel penjelas menjadi tidak signifikan. Terdapat dua cara yang dapat digunakan untuk mendeteksi *overdispersion*, yaitu:

1. *Deviance*

$$\phi_1 = \frac{D^2}{db}, D^2 = 2 \sum_{i=1}^n \left\{ y_i \ln \left(\frac{y_i}{\hat{\mu}_i} \right) - (y_i - \hat{\mu}_i) \right\} \quad (2.13)$$

db = n-k dengan k merupakan banyaknya parameter termasuk konstanta, n merupakan banyaknya pengamatan, dan D^2 adalah nilai *deviance*.

2. *Pearson Chi-Square*

$$\phi_2 = \frac{\chi^2}{db}, \chi^2 = \sum_{i=1}^n \frac{(y_i - \mu_i)^2}{\text{var}(y_i)} \quad (2.14)$$

dimana db = n-k dengan k merupakan banyaknya parameter termasuk konstanta, n merupakan banyaknya pengamatan, dan χ^2 adalah *Pearson Chi-Square*.

Jika ϕ_1 dan ϕ_2 bernilai lebih dari satu, maka dapat disimpulkan terjadi *overdispersion* pada data.

2.7 Regresi Binomial Negatif

Binomial Negatif merupakan hasil dari distribusi mixture Poisson-Gamma. Distribusi mixture Poisson terjadi jika nilai *mean* dari distribusi Poisson dinyatakan dalam suatu distribusi lain. Jika *mean* dinyatakan dalam distribusi Gamma maka model dari distribusi tersebut disebut regresi Binomial Negatif. Model regresi Binomial Negatif yang dibangun memiliki sebaran Binomial Negatif dengan parameter μ dan k , sehingga diperoleh nilai tengah dan ragamnya adalah:

$$E(Y) = \mu \text{ dan } V(Y) = \mu + \alpha\mu^2 \quad (2.15)$$

Fungsi probabilitas dari suatu variabel acak Y yang berdistribusi Binomial Negatif adalah sebagai berikut (Cameron dan Trivedi, 1998).

$$f(y) = \frac{\Gamma(y + \alpha^{-1})}{\Gamma(y + 1) \Gamma(\alpha^{-1})} \left(\frac{\mu}{\alpha^{-1} + \mu} \right)^y \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu} \right)^{\alpha^{-1}} \quad (2.16)$$

Misalkan ingin diketahui hubungan antara suatu variabel respon Y dengan variabel penjelas X_1, X_2, \dots, X_k . Variabel respon Y berupa data *count* dan menyatakan banyaknya kejadian yang diamati pada suatu populasi tertentu. Model regresi linier menggunakan hubungan antara variabel respon Y dengan variabel penjelas X_1, X_2, \dots, X_k sebagai berikut:

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki} + \varepsilon_i \quad (2.17)$$

Misal nilai ekspektasi untuk Y_i adalah $E(Y_i | X_{1i} = x_{1i}, \dots, x_{ki}) = \mu_i$ maka diperoleh:

$$\mu_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki} \quad (2.18)$$

atau bila dinyatakan dalam vektor menjadi

$$\mu_i = \mathbf{x}_i^T \boldsymbol{\beta} \quad (2.19)$$

Misalkan variabel acak Y memiliki distribusi probabilitas yang bergantung pada parameter θ yang dianggap sebagai parameter lokasi dan terdapat parameter lain yaitu ϕ yang disebut parameter pengganggu. Distribusi dari Y merupakan anggota dari keluarga eksponensial jika fungsi probabilitasnya memiliki bentuk sebagai berikut.

$$f(y; \theta, \phi) = \frac{y\theta - b(\theta)}{a(\phi)} + c(y; \phi) \quad (2.20)$$

dimana θ merupakan parameter kanonik, $a(\phi)$ merupakan parameter skala, dan $b(\theta)$ adalah fungsi unit *cumulant*. Misalkan Y adalah suatu variabel acak yang berdistribusi Binomial Negatif dengan fungsi probabilitas pada persamaan (2.16), sehingga akan diperoleh:

$$\begin{aligned} f(y; \mu, \alpha) &= \exp\left(\ln\left(\frac{\Gamma(y + \alpha^{-1})}{\Gamma(y + 1) \Gamma(\alpha^{-1})} \left(\frac{\mu}{\alpha^{-1} + \mu}\right)^y \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu}\right)^{\alpha^{-1}}\right)\right) \\ &= \exp\left(\ln\left(\frac{\Gamma(y + \alpha^{-1})}{\Gamma(y + 1) \Gamma(\alpha^{-1})}\right) + y \ln\left(\frac{\mu}{\alpha^{-1} + \mu}\right) \right. \\ &\quad \left. + \alpha^{-1} \ln\left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu}\right)\right) \end{aligned}$$

$$\begin{aligned}
&= \exp\left(y \ln\left(\frac{\mu}{\alpha^{-1} + \mu}\right) + \ln\left(\frac{\Gamma(y + \alpha^{-1})}{\Gamma(y + 1) \Gamma(\alpha^{-1})}\right)\right) \\
&\quad + \alpha^{-1} \ln\left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu}\right)
\end{aligned} \tag{2.21}$$

Dengan menyesuaikan persamaan (2.20) dan (2.21) maka didapatkan

- $\theta = \ln\left(\frac{\mu}{\alpha^{-1} + \mu}\right)$
- $b(\theta) = -\alpha^{-1} \ln\left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu}\right)$
- $c(y; \phi) = \ln\left(\frac{\Gamma(y + \alpha^{-1})}{\Gamma(y + 1) \Gamma(\alpha^{-1})}\right)$
- $\alpha(\phi) = 1.$

Komponen pada bentuk keluarga eksponensial yang merupakan *link function* adalah θ . Berdasarkan hasil tersebut, *link function* kanonik untuk Binomial Negatif menghasilkan bentuk yang cukup rumit dan interpretasi dari parameter model akan menjadi lebih sulit, sehingga jarang digunakan. Model Binomial Negatif pada umumnya menggunakan fungsi penghubung logaritma atau *log link* (Hilbe, 2011):

$$\ln(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} \tag{2.22}$$

Model Binomial Negatif dapat menggunakan *log link* karena $\ln(\mu_i)$ dan $\mathbf{x}_i^T \boldsymbol{\beta}$ akan terdefinisi di dalam interval $(-\infty, \infty)$ dan interpretasi parameter regresi akan menjadi lebih mudah. Setelah diperoleh fungsi penghubung, maka selanjutnya dapat dinyatakan model regresi Binomial Negatif untuk memodelkan data *count* yaitu:

$$\ln\{E(Y_i | \mathbf{x}_i)\} = \ln(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} \text{ untuk } i = 1, 2, \dots, n \tag{2.23}$$

sehingga dapat diperoleh model regresi Binomial Negatif sebagai berikut.

$$\mu_i = \exp(\mathbf{x}_i^T \boldsymbol{\beta}) = \exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}) \tag{2.24}$$

2.8 Estimasi Parameter Regresi Binomial Negatif

Metode *Maximum Likelihood Estimation* (MLE) digunakan untuk estimasi parameter dalam regresi Binomial Negatif. Setelah itu mensubstitusikan $\mu_i = \exp(\mathbf{x}_i^T \boldsymbol{\beta})$ ke dalam fungsi probabilitas bersyarat pada persamaan (2.12) akan diperoleh

$$f(y; \boldsymbol{\beta}, \alpha) = \frac{\Gamma(y + \alpha^{-1})}{\Gamma(y + 1) \Gamma(\alpha^{-1})} \left(\frac{\exp(\mathbf{x}_i^T \boldsymbol{\beta})}{\alpha^{-1} + \exp(\mathbf{x}_i^T \boldsymbol{\beta})} \right)^y \left(\frac{\alpha^{-1}}{\alpha^{-1} + \exp(\mathbf{x}_i^T \boldsymbol{\beta})} \right)^{\alpha^{-1}} \quad (2.25)$$

Fungsi *Likelihood* dari regresi Binomial Negatif adalah sebagai berikut.

$$\begin{aligned} L(\boldsymbol{\beta}, \alpha) &= \prod_{i=1}^n f(y; \boldsymbol{\beta}, \alpha) \\ &= \prod_{i=1}^n \frac{\Gamma(y + \alpha^{-1})}{\Gamma(y + 1) \Gamma(\alpha^{-1})} \left(\frac{\exp(\mathbf{x}_i^T \boldsymbol{\beta})}{\alpha^{-1} + \exp(\mathbf{x}_i^T \boldsymbol{\beta})} \right)^y \left(\frac{\alpha^{-1}}{\alpha^{-1} + \exp(\mathbf{x}_i^T \boldsymbol{\beta})} \right)^{\alpha^{-1}} \end{aligned} \quad (2.26)$$

Estimasi regresi Binomial Negatif menggunakan metode iterasi *Newton Rhapsion* untuk memaksimumkan fungsi *Likelihood*.

2.9 Pengujian Parameter Regresi Binomial Negatif

Pengujian parameter model regresi Binomial Negatif dilakukan untuk menguji kelayakan model yang diperoleh dari penaksiran parameter. Hal ini bertujuan untuk mengetahui peubah bebas yang terdapat dalam model berpengaruh nyata atau tidak. Berikut ini merupakan hipotesis untuk pengujian parameter model secara keseluruhan menggunakan uji devians (Hosmer dan Lemeshow, 1995).

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1 : \text{paling sedikit terdapat satu } \beta_j \neq 0 \quad ; j=1, 2, \dots, k$$

dengan menggunakan hipotesis tersebut, statistik uji yang digunakan adalah sebagai berikut.

$$G^2 = -2 \ln \left[\frac{L(\hat{\omega})}{L(\hat{\Omega})} \right] = 2 \ln L(\hat{\Omega}) - 2 \ln L(\hat{\omega}) \quad (2.27)$$

dengan kriteria penolakannya adalah tolak H_0 jika $G^2 > \chi_{\alpha,p}^2$.

Pengujian signifikansi parameter secara parsial dilakukan dengan menggunakan uji Wald. Uji Wald dilakukan untuk mengetahui parameter mana saja yang memberikan pengaruh yang signifikan terhadap model dengan hipotesis sebagai berikut.

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0 \quad ; j=1, 2, \dots, k$$

dengan menggunakan hipotesis tersebut, statistik uji Wald yang digunakan adalah sebagai berikut.

$$Z_{hit} = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \quad (2.28)$$

Kriteria penolakannya adalah tolak H_0 jika $|Z_{hit}| > Z_{\alpha/2}$. Tolak H_0 berarti bahwa parameter ke- j signifikan terhadap model regresi Binomial Negatif.

2.10 Uji Deteksi Nonlinier

Uji deteksi nonlinier digunakan untuk menemukan kasus data yang memiliki pola non linier pada model regresi. Uji deteksi nonlinier dapat dilakukan dengan menggunakan uji Terasvirta. Uji Terasvirta termasuk dalam kelompok uji *Lagrange Multiplier* dengan pendekatan ekspansi Taylor yang menggunakan statistik uji χ^2 dengan derajat bebas ν . Prosedur uji Terasvirta adalah sebagai berikut (Terasvirta, Lin, dan Granger, 1993):

- Meregresikan y_i pada $1, x_1, x_2, \dots, x_p$ dan menghitung nilai-nilai residual u_i .
- Meregresikan \hat{u}_i pada $1, x_1, x_2, \dots, x_p$ dan m prediktor tambahan dan hitung koefisien determinasi dari regresi R^2 .
- Hitung $\chi^2 = nR^2$, dimana n adalah jumlah pengamatan yang digunakan.

Hipotesis yang digunakan dalam uji Terasvirta adalah sebagai berikut.

H_0 : model linier

H_1 : model non linier

dengan menggunakan hipotesis tersebut, statistik uji yang digunakan adalah sebagai berikut.

$$\chi^2 = nR^2 \quad (2.29)$$

Kriteria penolakannya adalah tolak H_0 jika p -value dari statistik uji χ^2 kurang dari α .

2.11 Regresi Nonlinier

Secara umum, regresi adalah suatu metode untuk meramalkan nilai harapan yang bersyarat. Regresi dikatakan linier apabila hubungan antara variabel independen dan variabel dependennya adalah linier. Apabila hubungan antara variabel independent dan variabel dependen tidak linier, maka regresi dikatakan regresi nonlinier. *Error* pada regresi non linier diasumsikan untuk mempunyai nilai harapan sebesar nol, ragam yang konstan dan tidak dikorelasikan, sama seperti asumsi *error* pada model regresi linier (Neter, Kutner, Nachtsheim, dan Wasserman, 1996). Bentuk dari regresi nonlinier adalah:

$$Y_i = f(X_i, \beta) + \varepsilon_i \quad (2.30)$$

dimana f adalah fungsi nonlinier dari satu atau lebih p parameter β_1, \dots, β_p dan ε adalah random error. Terdapat beberapa bentuk model regresi nonlinier pada Tabel 2.1.

Tabel 2.1 Beberapa Contoh Model Regresi Nonlinier

Nama Model	Model
<i>Exponential</i>	$y = \alpha e^{\beta_1 X}$
<i>Logarithmic</i>	$y = \alpha + \beta_1 \ln(X)$
<i>Inverse</i>	$y = \alpha + \frac{\beta_1}{X}$
<i>Compound</i>	$y = \alpha \beta_1^X$

Tabel 2.2 Beberapa Contoh Model Regresi Nonlinier (Lanjutan)

Nama Model	Model
<i>Power</i>	$y = \alpha X^{\beta_1}$
<i>Quadratic</i>	$y = \alpha + \beta_1 X + \beta_2 X^2$
<i>Qubic</i>	$y = \alpha + \beta_1 X + \beta_2 X^2 + \beta_3 X^3$

2.12 Estimasi Parameter Regresi Nonlinier

Parameter β dalam fungsi nonlinier dapat diestimasi dengan metode *nonlinear least square*. Penaksiran β dengan metode ini bertujuan untuk mendapatkan nilai β yang meminimumkan *residual sum of squares* $S(\beta)$. Metode *Levenberg Marquardt* menghasilkan solusi numerik untuk meminimalkan sebuah fungsi nonlinier terhadap parameter di fungsi tersebut. Aplikasi utama dari metode *Levenberg Marquardt* adalah pada masalah kuadrat terkecil yang bertujuan untuk mengoptimasi parameter β , sehingga *Residual Sum Square* (RSS) menjadi minimal nilainya. *Residual Sum Square* (RSS) dapat dituliskan sebagai berikut.

$$S(\beta) = \sum_{i=1}^n [y_i - f(X_i, \beta)]^2 \quad (2.31)$$

Metode *Levenberg-Marquardt* menggunakan prosedur iterasi. Untuk memulai proses minimisasi, langkah pertama adalah melakukan perkiraan nilai dari parameter vektor, β . Pada setiap tahap iterasi, parameter vektor, β , akan diganti dengan nilai estimasi baru, yaitu $\beta + \delta$. Untuk mencari nilai δ , fungsi $f(X_i, \beta)$ didekati dengan menlinierkan.

$$(X_i, \beta + \delta) \approx f(X_i, \beta) + J_i \delta \quad (2.32)$$

dimana J_i merupakan gradient (vektor baris) dari f terhadap parameter β .

$$J_i = \frac{\partial f(X_i, \beta)}{\partial \beta} \quad (2.33)$$

Aproksimasi dari $f(X_i, \beta + \delta)$ akan menghasilkan.

$$S(\beta + \delta) \approx \sum_{i=1}^n [y_i - f(X_i, \beta) - J_i \delta]^2 \quad (2.34)$$

atau jika ditulis dalam notasi vektor menjadi

$$S(\beta + \delta) \approx \|y - f(\beta) - J\delta\|^2 \quad (2.35)$$

2.13 Pengujian Parameter Regresi Nonlinier

Pengujian parameter dalam model regresi dilakukan untuk mengetahui parameter menunjukkan hubungan yang nyata antara variabel dependen dengan variabel independen. Dalam pengujian parameter terdapat dua tahap, yaitu uji serentak dan uji parsial. Hipotesis yang digunakan dalam pengujian secara serentak adalah:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1 : \text{paling sedikit terdapat satu } \beta_j \neq 0 \quad ; j=1, 2, \dots, k$$

dengan menggunakan hipotesis tersebut, statistik uji yang digunakan adalah sebagai berikut.

$$F_{hit} = \frac{MSR}{MSE} \quad (2.36)$$

Kriteria penolakannya adalah tolak H_0 jika $F_{hit} > F_{\alpha, v1, v2}$, dimana $v1 = p$ dan $v2 = n - k - 1$. Tolak H_0 berarti bahwa parameter ke- p signifikan terhadap model regresi.

Setelah melakukan pengujian secara serentak, akan dilanjutkan ke pengujian secara parsial. Pengujian secara parsial bertujuan untuk mengetahui variabel mana saja yang berpengaruh signifikan terhadap variabel respon. Hipotesis yang digunakan dalam pengujian secara parsial adalah:

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0 \quad ; j=1, 2, \dots, k$$

dengan menggunakan hipotesis tersebut, statistik uji yang digunakan adalah sebagai berikut.

$$t_{hit} = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \quad (2.37)$$

Kriteria penolakannya adalah tolak H_0 apabila $|t_{hit}| > t_{df; 1-\alpha/2}$, dimana $df = n - k$. Tolak H_0 berarti bahwa β_j tidak sama dengan nol atau variabel independen tersebut memberikan pengaruh yang signifikan pada variabel dependennya.

2.14 AIC

Terdapat beberapa metode dalam menentukan model terbaik, salah satunya adalah *Akaike Information Criterion* (AIC). Menurut Bozdogan (2000), AIC didefinisikan sebagai berikut.

$$AIC = -2 \ln L(\hat{\beta}) + 2k \quad (2.38)$$

dimana $L(\hat{\beta})$ adalah nilai likelihood dan k adalah jumlah parameter bebas. Parameter bebas adalah parameter dari variabel bebas yang tidak saling berinteraksi satu sama lain. Model terbaik adalah model yang mempunyai nilai AIC terkecil.

2.15 Tuberkulosis

Tuberkulosis adalah penyakit radang parenkim paru yang menular karena infeksi kuman TB, yaitu *Mycobacterium tuberculosis*. Sebagian besar kuman TB menyerang paru, tetapi dapat juga mengenai organ tubuh lainnya. Keluhan yang dirasakan penderita tuberkulosis dapat bermacam-macam, seperti demam, batuk, sesak napas, nyeri dada, dan malaise. Sumber penularan adalah pasien TB BTA (Basil Tahan Asam). Pada waktu batuk atau bersin, pasien TB menyebarkan kuman ke udara dalam bentuk percikan dahak.

Departemen Kesehatan RI (2009) mengklasifikasikan tuberkulosis menjadi dua klasifikasi, yaitu tuberkulosis paru dan tuberkulosis ekstra paru. Tuberkulosis paru adalah penyakit tuberkulosis yang menyerang jaringan paru, sedangkan tuberkulosis ekstra paru merupakan penyakit tuberkulosis yang menyerang organ tubuh lain selain paru, diantaranya organ selaput otak, selaput jantung (*pericardium*), kelenjar getah bening, tulang, limfa, persendian, kulit, usus, ginjal, saluran kencing, dan lain-lain.

Beberapa faktor yang dapat mempengaruhi terjadinya penyakit tuberkulosis adalah kondisi sosial ekonomi masyarakat, yaitu status gizi dan sanitasi lingkungan. Semakin rendah status gizi dan sanitasi lingkungan menyebabkan semakin rendahnya daya tahan tubuh, sehingga lebih mudah tertular tuberkulosis pada saat sakit (Entjang, 2000). Faktor lingkungan memegang peranan penting dalam penularan penyakit tuberkulosis, terutama lingku-

ngan rumah yang tidak memenuhi syarat. Lingkungan rumah merupakan salah satu faktor yang memberikan pengaruh besar terhadap status kesehatan penghuninya (Notoatmodjo, 2003).

(Halaman ini sengaja dikosongkan)

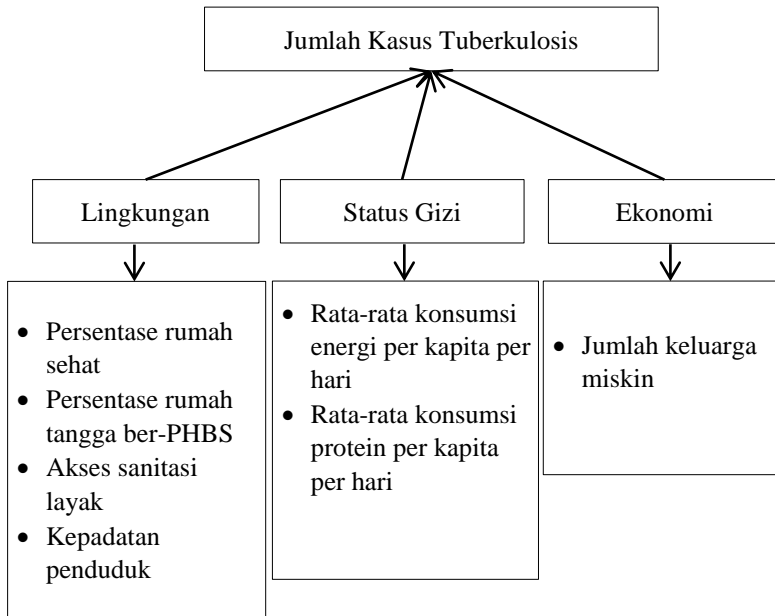
BAB III METODOLOGI PENELITIAN

3.1 Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder. Data yang digunakan diperoleh dari publikasi Profil Kesehatan Provinsi Jawa Barat tahun 2015 oleh Dinas Kesehatan Provinsi Jawa Barat dan juga didapatkan dari publikasi Provinsi Jawa Barat Dalam Angka oleh Badan Pusat Statistik Provinsi Jawa Barat. Data yang digunakan adalah data untuk masing-masing 27 kabupaten/kota di Provinsi Jawa Barat.

3.2 Kerangka Konseptual

Diagram jalur model kerangka konseptual untuk menggambarkan hubungan antara variabel respon dengan variabel prediktor dalam penelitian ini sebagai berikut.



Gambar 3.1 Kerangka Konseptual Penelitian.

Angka kejadian tuberkulosis dipengaruhi oleh beberapa faktor seperti pada Gambar 3.1. Beberapa penelitian sebelumnya yang menunjukkan hubungan sebab akibat antara variabel respon dan prediktor adalah sebagai berikut.

- a. Meilania (2016) menyebutkan bahwa variabel yang berpengaruh terhadap jumlah tuberkulosis adalah persentase rumah tangga ber-PHBS, persentase penduduk yang memiliki akses sanitasi layak, persentase rumah sehat, dan kepadatan penduduk.
- b. Kondisi gizi seseorang akan mempengaruhi perjalanan suatu infeksi. Menurut Badan Pusat Statistik Jawa Barat (2015), tingkat kecukupan gizi mencakup konsumsi kalori dan protein. Angka kecukupan gizi yang dianjurkan adalah suatu kecukupan rata-rata zat gizi setiap hari bagi semua orang menurut golongan umur, jenis kelamin, ukuran tubuh, aktivitas tubuh untuk mencapai derajat kesehatan yang optimal.
- c. WHO (2013) menyebutkan bahwa penderita Tuberkulosis di dunia menyerang kelompok sosial ekonomi lemah atau miskin. Meilania (2016) menyebutkan bahwa variabel yang berpengaruh terhadap jumlah tuberkulosis adalah jumlah keluarga miskin.

3.3 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini terdiri dari satu variabel respon (Y) dan 8 variabel prediktor (X) yang memengaruhinya seperti ditunjukkan pada Tabel 3.1.

Tabel 3.1 Variabel Penelitian

Variabel	Keterangan	Skala
Y	Jumlah kasus tuberkulosis	Rasio
X ₁	Jumlah keluarga miskin	Rasio
X ₂	Rata-rata konsumsi energi per kapita per hari	Rasio
X ₃	Rata-rata konsumsi protein per kapita per hari	Rasio
X ₄	Persentase rumah tangga berperilaku hidup bersih dan sehat	Rasio

Tabel 3.1 Variabel Penelitian (Lanjutan)

Variabel	Keterangan	Skala
X ₅	Persentase penduduk dengan akses sanitasi layak	Rasio
X ₆	Persentase rumah sehat	Rasio
X ₇	Persentase tempat pengelolaan makanan yang memenuhi sanitasi	Rasio
X ₈	Kepadatan penduduk	Rasio

Penjelasan masing-masing variabel adalah sebagai berikut.

1. Jumlah kasus tuberkulosis di setiap kabupaten / kota di Provinsi Jawa Barat tahun 2015 (Y). Jawa Barat terdiri dari 18 kabupaten dan 9 kota. Penelitian ini meneliti kasus tuberkulosis yang terjadi di 18 kabupaten dan 9 kota tersebut.
2. Jumlah keluarga miskin (X₁) merupakan jumlah keluarga yang memiliki rata-rata pengeluaran perkapita perbulan dibawah garis kemiskinan (dibawah Rp 600.000 perbulan).
3. Rata-rata konsumsi energi per kapita per hari (X₂) merupakan rata-rata dari jumlah konsumsi kalori yang dihitung berdasarkan hasil kali antara kuantitas setiap makanan yang dikonsumsi dengan besarnya kandungan kalori dalam setiap makanan yang dikonsumsi tersebut.
4. Rata-rata konsumsi protein per kapita per hari (X₃) merupakan rata-rata dari jumlah konsumsi protein yang dihitung berdasarkan hasil kali antara kuantitas setiap makanan yang dikonsumsi dengan besarnya kandungan protein di setiap makanan yang dikonsumsi tersebut.
5. Persentase rumah tangga berperilaku hidup bersih dan sehat (X₄) merupakan hasil bagi dari jumlah rumah tangga yang berperilaku hidup bersih dan sehat terhadap jumlah rumah tangga yang dipantau di tiap kabupaten/kota di Provinsi Jawa Barat dikalikan dengan 100%.
6. Persentase penduduk dengan akses sanitasi layak (X₅) merupakan hasil bagi antara jumlah penduduk dengan akses sanitasi layak terhadap jumlah penduduk yang diperiksa akses sanitasinya dikalikan 100%.

7. Persentase rumah sehat (X_6) merupakan hasil bagi dari jumlah rumah yang tergolong rumah sehat terhadap jumlah rumah yang dipantau di tiap kabupaten/kota di Provinsi Jawa Barat dikalikan dengan 100%.
8. Persentase tempat pengelolaan makanan yang memenuhi sanitasi (X_7) merupakan hasil bagi dari jumlah TPM yang memenuhi syarat sanitasi terhadap jumlah TPM yang dipantau di tiap kabupaten/kota di Provinsi Jawa Barat dikalikan dengan 100%.
9. Kepadatan penduduk (X_8) merupakan hasil bagi dari jumlah penduduk terhadap luas wilayah di tiap kabupaten/kota di Provinsi Jawa Barat dalam satuan (orang/km²).

Struktur data dalam penelitian ini adalah sebagai berikut.

Tabel 3.2 Struktur Data

Kabupaten/ Kota	Y	X ₁	X ₂	...	X ₈
Kabupaten Bogor	y_1	$x_{1;1}$	$x_{2;1}$...	$x_{8;1}$
Kabupaten Sukabumi	y_2	$x_{1;2}$	$x_{2;2}$...	$x_{8;2}$
Kabupaten Cianjur	y_3	$x_{1;3}$	$x_{2;3}$...	$x_{8;3}$
⋮	⋮	⋮	⋮	⋮	⋮
Kota Banjar	y_{27}	$x_{1;27}$	$x_{2;27}$...	$x_{8;27}$

3.4 Langkah Analisis

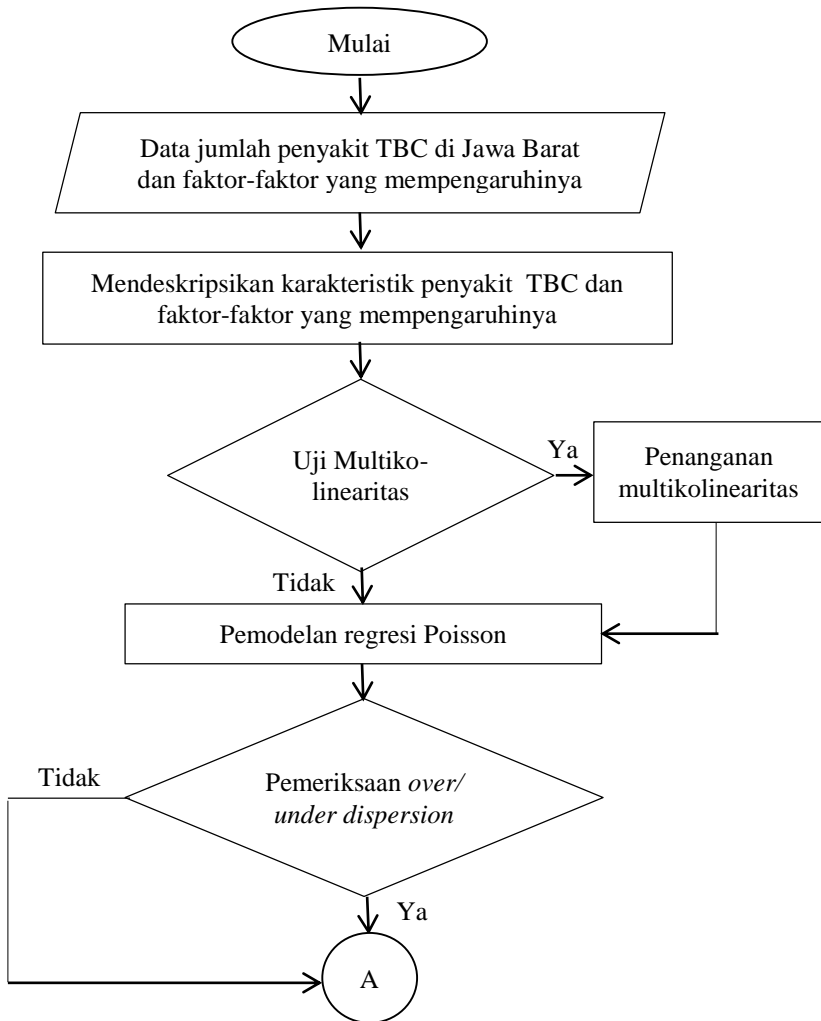
Langkah-langkah analisis yang akan dilakukan dalam penelitian ini adalah sebagai berikut.

1. Mendeskripsikan karakteristik jumlah kasus tuberkulosis dan faktor-faktor yang mempengaruhinya.
2. Pengujian multikolinieritas antar variabel prediktor menggunakan koefisien korelasi Pearson dan nilai VIF.
3. Mendapatkan model regresi Poisson pada jumlah kasus tuberkulosis di Jawa Barat.
 - a. Pengujian signifikansi model regresi Poisson secara serentak menggunakan uji devians.
 - b. Pengujian signifikansi model regresi Poisson secara parsial menggunakan uji Wald.

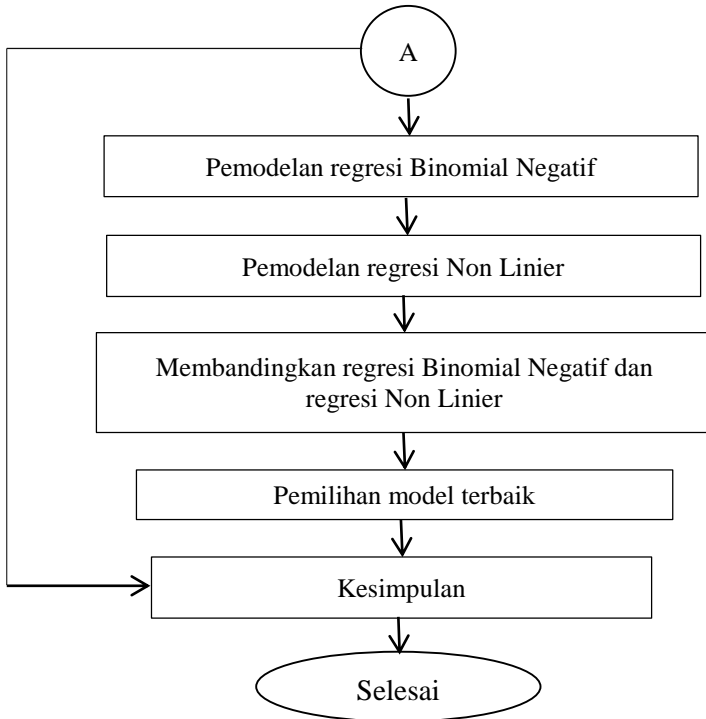
4. Mendeteksi adanya overdispersi pada data dengan melihat nilai *Pearson chi-square* dan *deviance* yang dibagi dengan derajat bebasnya.
5. Mendapatkan kombinasi variabel prediktor regresi Binomial Negatif dengan kombinasi satu variabel prediktor hingga 8 variabel prediktor.
6. Mendapatkan model terbaik regresi Binomial Negatif pada jumlah kasus tuberkulosis di Jawa Barat.
 - a. Pengujian signifikansi model regresi Binomial Negatif secara serentak menggunakan uji devians.
 - b. Pengujian signifikansi model regresi Binomial Negatif secara parsial menggunakan uji Wald.
7. Interpretasi model regresi Binomial Negatif.
8. Melakukan uji deteksi nonlinier menggunakan *scatter plot* dan uji Terasvirta.
9. Mendapatkan model regresi nonlinier pada jumlah kasus tuberkulosis di Jawa Barat.
 - a. Melakukan pemodelan menggunakan semua variabel prediktor.
 - b. Melakukan pemodelan menggunakan variabel yang signifikan.
 - c. Mendapatkan model terbaik regresi nonlinier dengan membandingkan kedua pemodelan menggunakan kriteria nilai S dan MSE terkecil.
10. Membandingkan model regresi Binomial Negatif dan regresi nonlinier berdasarkan kriteria nilai AIC terkecil.
11. Menarik kesimpulan.

3.5 Diagram Alir

Berdasarkan langkah-langkah analisis data tersebut dapat disajikan dalam bentuk diagram alir pada Gambar 3. berikut ini.



Gambar 3.2 Diagram Alir Penelitian.



Gambar 3.2 Diagram Alir Penelitian (Lanjutan).

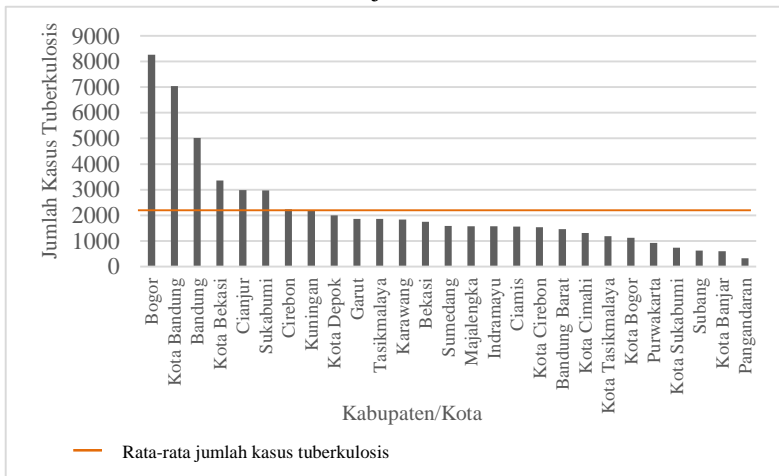
(Halaman ini sengaja dikosongkan)

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini dibahas beberapa hal untuk menjawab rumusan masalah dalam penelitian ini yaitu mengenai karakteristik jumlah kasus tuberkulosis di Jawa Barat pada tahun 2015 dan faktor-faktor yang diduga mempengaruhinya menggunakan statistika deskriptif dan pemodelan jumlah kasus tuberkulosis menggunakan metode regresi Binomial Negatif. Bagian awal disajikan tentang deskripsi jumlah kasus tuberkulosis dan karakteristik faktor-faktor yang diduga mempengaruhinya.

4.1 Karakteristik Banyaknya Kasus Tuberkulosis di Jawa Barat

Jawa Barat merupakan salah satu provinsi di Indonesia dan merupakan provinsi penyumbang terbesar jumlah kasus tuberkulosis di Indonesia. Tuberkulosis di Jawa Barat terjadi di 18 kabupaten dan 7 kota di Jawa Barat pada tahun 2015 dengan jumlah total sebanyak 59.446 kasus. Karakteristik mengenai jumlah kasus tuberkulosis di Jawa Barat disajikan dalam Gambar 4.1.



Gambar 4.1 Jumlah Kasus Tuberkulosis di Kabupaten/Kota di Jawa Barat

Gambar 4.1 menunjukkan bahwa Kabupaten Bogor merupakan kabupaten di Jawa Barat yang memiliki jumlah kasus tuberkulosis tertinggi pada tahun 2015 dengan jumlah sebanyak 8271 kasus. Kemudian diikuti dengan Kota Bandung, Kabupaten Bandung, dan Kota Bekasi. Selain itu, Kabupaten Pangandaran merupakan kabupaten dengan jumlah kasus tuberkulosis terendah, yaitu sebanyak 318 kasus. Kemudian garis merah menunjukkan rata-rata jumlah kasus tuberkulosis di Jawa Barat, dimana rata-rata jumlah kasus tuberkulosis di Jawa Barat sebanyak 2202 kasus. Berdasarkan rata-rata tersebut, masih terdapat beberapa kabupaten dan kota di Jawa Barat yang berada di atas rata-rata, yaitu Kabupaten Bogor, Kota Bandung, Kabupaten Bandung, Kota Bekasi, Kabupaten Cianjur, Kabupaten Sukabumi, dan Kabupaten Cirebon.

Setelah diketahui karakteristik untuk variabel jumlah kasus tuberkulosis di Jawa Barat, berikut merupakan karakteristik untuk 8 variabel prediktor yang diduga mempengaruhi jumlah kasus tuberkulosis di Jawa Barat yang disajikan dalam Tabel 4.1.

Tabel 4.1 Statistika Deskriptif Variabel Penelitian

Variabel	Minimum	Maksimum	Rata-Rata	Varian
Jumlah keluarga miskin (X_1)	14630,00	446040,00	181277,00	13.920.589.981
Rata-rata konsumsi energi per kapita per hari (X_2)	1853,70	2339,00	2052,50	16.108
Rata-rata konsumsi protein per kapita per hari (X_3)	48,26	66,94	57,81	18
Persentase rumah tangga berperilaku hidup bersih dan sehat (X_4)	37,60	77,50	53,77	103
Persentase penduduk dengan akses sanitasi layak (X_5)	17,92	93,75	65,91	315
Persentase rumah sehat (X_6)	39,31	100,00	70,20	215
Persentase TPM yang memenuhi sanitasi (X_7)	14,10	87,90	46,61	436
Kepadatan penduduk (X_8)	387,00	14937,00	3825,00	21.844.793

Berdasarkan Tabel 4.1, dapat diketahui bahwa nilai varian tertinggi dari beberapa variabel prediktor yang diduga mempengaruhi jumlah tuberkulosis di Jawa Barat terdapat pada variabel X_1 (jumlah keluarga miskin), dimana jumlah keluarga miskin yang paling sedikit terdapat pada Kota Banjar, sedangkan jumlah keluarga miskin paling banyak terdapat pada Kabupaten Bogor. Varian yang be-

sar pada variabel X_1 menunjukkan bahwa setiap kabupaten/kota di Jawa Barat memiliki jumlah keluarga miskin yang berbeda-beda.

Berdasarkan Peraturan Menteri Kesehatan RI Nomor 75 Tahun 2013, rata-rata kecukupan energi dan protein bagi penduduk Indonesia masing-masing sebesar 2150 kkal dan 57 gram protein. Rata-rata konsumsi energi per kapita per hari (X_2) di kabupaten/kota di Jawa Barat memiliki nilai rata-rata sebesar 2025,50 kkal. Hal ini menunjukkan bahwa rata-rata konsumsi energi di kabupaten/kota di Jawa Barat masih belum memenuhi standard kecukupan konsumsi energi. Berbeda dengan konsumsi energi yang belum memenuhi, rata-rata konsumsi protein di kabupaten/kota di Jawa Barat (X_3) telah memenuhi standard kecukupan konsumsi energi karena memiliki nilai rata-rata sebesar 57,81 gram.

4.2 Pemodelan Banyaknya Kasus Tuberkulosis di Jawa Barat Menggunakan Regresi Poisson

Sebelum melakukan analisis menggunakan regresi Poisson, terlebih dahulu dilakukan pemeriksaan multikolinieritas untuk mengetahui ada atau tidaknya korelasi antar variabel prediktor. Pemeriksaan multikolinieritas dilakukan dengan melihat nilai korelasi Pearson dan nilai VIF. Berikut merupakan koefisien korelasi yang dihasilkan antar variabel prediktor.

Tabel 4.2 Koefisien Korelasi antar Variabel Prediktor

	X_1	X_2	X_3	X_4	X_5	X_6	X_7
X_2	-0,101						
	(0,617)						
X_3	-0,324	0,848					
	(0,099)	(0,000)*					
X_4	-0,414	0,104	0,410				
	(0,032)*	(0,606)	(0,034)*				
X_5	-0,034	0,099	0,242	0,554			
	(0,867)	(0,623)	(0,223)	(0,003)*			
X_6	0,167	-0,025	0,125	0,521	0,533		
	(0,406)	(0,903)	(0,543)	(0,005)*	(0,004)*		
X_7	-0,011	-0,067	-0,158	-0,058	0,075	0,043	
	(0,958)	(0,741)	(0,430)	(0,775)	(0,709)	(0,830)	
X_8	-0,488	-0,407	-0,020	0,426	0,119	0,059	-0,222
	(0,010)*	(0,035)*	(0,923)	(0,027)*	(0,556)	(0,771)	(0,267)

Jika koefisien korelasi Pearson antar variabel prediktor lebih dari 0,90 maka diduga terdapat kasus multikolinieritas. Berdasarkan Tabel 4.2, dapat diketahui bahwa semua variabel prediktor memiliki koefisien korelasi Pearson yang kurang dari 0,90 yang berarti bahwa tidak terdapat kasus multikolinieritas. Kriteria lain yang dapat digunakan untuk mendeteksi adanya multikolinieritas adalah nilai VIF yang disajikan dalam Tabel 4.3.

Tabel 4.3 Nilai VIF dari Variabel Prediktor

Variabel	Nilai VIF
X ₁	2,056
X ₂	8,744
X ₃	8,152
X ₄	2,922
X ₅	1,698
X ₆	2,017
X ₇	1,189
X ₈	2,972

Jika nilai VIF lebih dari 10 maka dapat disimpulkan terdapat kasus multikolinieritas. Tabel 4.3 menunjukkan bahwa nilai VIF untuk masing-masing variabel prediktor kurang dari 10. Hal itu menunjukkan bahwa tidak terdapat kasus multikolinieritas. Pemeriksaan multikolinieritas dengan menggunakan koefisien korelasi Pearson dan nilai VIF diperoleh hasil yang sama, sehingga dapat dilanjutkan ke tahap selanjutnya, yaitu pemodelan menggunakan regresi Poisson.

Setelah dilakukan pengecekan terhadap kasus multikolinieritas, selanjutnya akan dicari hubungan antara jumlah kasus tuberkulosis di Jawa Barat dengan kedelapan variabel prediktor menggunakan analisis regresi Poisson. Dalam pemodelan menggunakan regresi Poisson, langkah pertama akan dilakukan penaksiran atau estimasi parameter model regresi. Penaksiran parameter model regresi Poisson terangkum dalam Tabel 4.4. Nilai AIC yang dihasilkan dalam pemodelan menggunakan regresi Poisson adalah sebesar 10043.

Tabel 4.4 Estimasi Parameter Model Regresi Poisson

Parameter	Estimate	Std.Error	Z _{hitung}	P Value
β_0	3,0020	0,000031	95349,30	<0,0001*
β_1	$5,492 \times 10^{-6}$	0	Infty	<0,0001*
β_2	$4,819 \times 10^{-3}$	0,000071	68,23	<0,0001*
β_3	$-1,414 \times 10^{-1}$	0,002722	-51,94	<0,0001*
β_4	$2,825 \times 10^{-2}$	0,000775	36,46	<0,0001*
β_5	$-0,122 \times 10^{-3}$	0,000366	-3,34	0,0025*
β_6	$-0,191 \times 10^{-3}$	0,000423	-4,51	0,0001*
β_7	$-0,003 \times 10^{-3}$	0,000239	-0,11	0,9132
β_8	$0,131 \times 10^{-4}$	0,000001	114,69	<0,0001*

*) signifikan dengan taraf nyata 5%

Nilai estimasi parameter yang diperoleh kemudian diuji secara serentak dan parsial untuk mengetahui signifikansi pengaruh dari variabel bebas yang digunakan tersebut. Berdasarkan hasil pengujian serentak dengan taraf signifikansi 5%, didapatkan $\chi^2_{(8;0,05)}$ sebesar 15,507. Nilai tersebut lebih kecil dari nilai *deviance*, yaitu 10025 sehingga dapat diambil keputusan tolak H_0 yang berarti bahwa paling sedikit ada satu variabel prediktor yang berpengaruh signifikan terhadap variabel respon.

Selanjutnya untuk mengetahui variabel prediktor yang berpengaruh signifikan terhadap jumlah kasus tuberkulosis di Jawa Barat, dilakukan pengujian signifikansi parameter secara parsial. Berdasarkan hasil pengujian secara parsial pada Tabel 4.4, dapat diketahui bahwa *p-value* untuk semua parameter kurang dari 0,05 kecuali untuk parameter β_7 . Oleh karena itu dapat diambil keputusan tolak H_0 yang berarti bahwa parameter $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6$, dan β_8 berpengaruh signifikan terhadap model, sehingga didapatkan model regresi Poisson sebagai berikut.

$$\hat{\mu} = \exp(3,0020 + 0,000005492X_1 + 0,004819X_2 - 0,1414X_3 + 0,02825X_4 - 0,00122X_5 - 0,00191X_6 - 0,00003X_7 + 0,000131X_8)$$

Dari model regresi Poisson di atas, didapatkan variabel yang berpengaruh terhadap jumlah kasus tuberkulosis di Jawa Barat adalah jumlah keluarga miskin (X_1), rata-rata konsumsi energi per kapita per hari (X_2), rata-rata konsumsi protein per kapita per hari (X_3), persentase rumah tangga ber-PHBS (X_4), persentase penduduk dengan akses sanitasi layak (X_5), persentase rumah sehat (X_6), dan kepadatan penduduk (X_8). Peningkatan maupun penurunan jumlah kasus tuberkulosis setiap kabupaten/kota di Jawa Barat tergantung dari nilai koefisien masing-masing variabel.

Berdasarkan hasil estimasi yang terbentuk dapat dijelaskan bahwa untuk setiap pertambahan 1 keluarga miskin (X_1) akan meningkatkan rata-rata jumlah kasus tuberkulosis sebesar 1,000005 ($e^{5,492 \times 10^{-6}}$) kasus dengan asumsi variabel lain konstan. Kemudian setiap pertambahan rata-rata konsumsi energi (X_2) sebesar 1 kkal maka akan meningkatkan rata-rata jumlah kasus tuberkulosis sebanyak 1,0048 ($e^{4,819 \times 10^{-3}}$) kasus dengan asumsi variabel lain konstan. Kemudian setiap pertambahan rata-rata konsumsi protein (X_3) sebesar 1 gram maka akan menurunkan rata-rata jumlah kasus tuberkulosis sebanyak 0,8681 ($e^{-0,1414}$) kasus dengan asumsi variabel lain konstan. Setiap pertambahan persentase rumah tangga ber-PHBS (X_4) sebesar 1 persen maka akan meningkatkan rata-rata jumlah kasus tuberkulosis sebanyak 1,0286 ($e^{0,02825}$) kasus dengan asumsi variabel lain konstan. Setiap pertambahan persentase penduduk dengan akses sanitasi layak (X_5) sebesar 1 persen maka akan menurunkan rata-rata jumlah kasus tuberkulosis sebanyak 0,9987 ($e^{-0,00122}$) kasus dengan asumsi variabel lain konstan. Setiap pertambahan persentase rumah sehat (X_6) sebesar 1 persen maka akan menurunkan rata-rata jumlah kasus tuberkulosis sebanyak 0,9980 ($e^{-0,00191}$) kasus dengan asumsi variabel lain konstan. Setiap pertambahan persentase TPM yang memenuhi sanitasi (X_7) sebesar 1 persen maka akan menurunkan rata-rata jumlah kasus tuberkulosis sebanyak 0,999 ($e^{-0,000003}$) kasus dengan asumsi variabel lain konstan. Selanjutnya setiap pertambahan kepadatan penduduk (X_8) akan meningkatkan rata-rata jumlah kasus tuberkulosis sebanyak 1,0001 ($e^{0,000131}$) kasus dengan asumsi variabel lain konstan.

4.3 Pemodelan Banyaknya Kasus Tuberkulosis di Jawa Barat Menggunakan Analisis Regresi Binomial Negatif

Langkah awal dalam pemodelan menggunakan regresi Binomial Negatif adalah pemeriksaan overdispersi. Apabila terdapat kasus overdispersi, maka pemodelan menggunakan regresi Binomial Negatif dapat digunakan. Data dikatakan terjadi overdispersi apabila nilai varians variabel responnya lebih besar dari nilai rata-ratanya. Overdispersi dapat diketahui dari nilai *deviance* dan *Pearson chi-square* pada model regresi Poisson yang dibagi dengan derajat bebasnya seperti yang ditunjukkan pada Tabel 4.5.

Tabel 4.5 *Deviance* Model Regresi Poisson

Kriteria	Nilai	db	Nilai/db
<i>Deviance</i>	9774,378	18	543,021
<i>Pearson chi-square</i>	10155,448	18	564,192

Nilai *deviance*/ db pada Tabel 4.5 menunjukkan nilai lebih besar dari satu, sehingga dapat disimpulkan bahwa terjadi kasus overdispersi pada data jumlah kasus tuberkulosis di Jawa Barat. Selain itu, hasil menggunakan kriteria *Pearson chi-square*/db didapatkan hasil yang sama, yaitu terdapat kasus overdispersi karena nilai *Pearson chi-square*/db yang dihasilkan lebih dari satu. Kasus overdispersi dapat diatasi dengan melakukan pemodelan menggunakan regresi Binomial Negatif, dimana metode tersebut dapat meminimalkan parameter dispersi.

Setelah diketahui adanya kasus overdispersi pada model Regresi Poisson, maka akan ditangani menggunakan regresi Binomial Negatif. Ketujuh variabel prediktor yang signifikan pada pemodelan menggunakan regresi Poisson tersebut membentuk 128 kemungkinan model regresi Binomial Negatif. Oleh karena itu, akan dilakukan pemodelan menggunakan beberapa kemungkinan model regresi Binomial Negatif. Model dengan nilai AIC terkecil adalah model terbaik untuk regresi Binomial Negatif. Setiap kombinasi variabel, mulai dari satu hingga tujuh variabel dengan taraf signifikan 5% ditampilkan pada Tabel 4.6.

Tabel 4.6 Kemungkinan Model Regresi Binomial Negatif dari Kombinasi Variabel Prediktor

Variabel yang digunakan	AIC	Parameter Signifikan
X_8	466,97	β_0
X_1, X_8	443,73	$\beta_0, \beta_1, \beta_8$
X_1, X_3, X_8	445,38	$\beta_0, \beta_1, \beta_8$
X_1, X_3, X_4, X_8	451,10	$\beta_0, \beta_1, \beta_8$
X_1, X_2, X_3, X_4, X_8	445,06	$\beta_0, \beta_1, \beta_2, \beta_3, \beta_8$
$X_1, X_2, X_3, X_4, X_6, X_8$	447,04	$\beta_0, \beta_1, \beta_3, \beta_8$
$X_1, X_2, X_3, X_4, X_5, X_6, X_8$	449,01	$\beta_0, \beta_1, \beta_3, \beta_8$

Nilai estimasi parameter dari masing-masing kemungkinan model tersebut ditampilkan pada Lampiran 9. Pada Tabel 4.6 dapat diketahui model yang memiliki parameter signifikan paling banyak dengan nilai AIC terkecil di masing-masing kombinasi variabel prediktor. Pada kombinasi enam variabel prediktor adalah kombinasi yang memiliki parameter signifikan lebih banyak dengan nilai AIC terkecil daripada kombinasi lainnya, dimana AIC yang dihasilkan sebesar 447,04 dengan parameter yang signifikan sebanyak empat, yaitu $\beta_0, \beta_1, \beta_3, \beta_8$. Oleh karena itu, kombinasi enam variabel prediktor ini yang akan dianalisis lebih lanjut untuk mendapatkan model regresi Binomial Negatif. Hasil estimasi parameter regresi Binomial Negatif ditampilkan pada Tabel 4.7.

Tabel 4.7 Estimasi Parameter Model Regresi Binomial Negatif

Parameter	Estimate	Std.Error	Z _{hitung}	P Value
β_0	4,923	1,784	2,760	0,00578
β_1	$5,238 \times 10^{-6}$	$9,286 \times 10^{-7}$	5,640	$1,7 \times 10^{-8}$
β_2	$3,523 \times 10^{-3}$	$1,848 \times 10^{-3}$	1,907	0,05655
β_3	$-1,187 \times 10^{-1}$	$5,241 \times 10^{-2}$	-2,265	0,02349
β_4	$1,682 \times 10^{-2}$	$1,254 \times 10^{-2}$	1,341	0,17981
β_6	$-1,173 \times 10^{-3}$	$7,390 \times 10^{-3}$	-0,159	0,87384
β_8	$1,211 \times 10^{-4}$	$2,816 \times 10^{-5}$	4,299	$1,71 \times 10^{-5}$

*) signifikan dengan taraf nyata 5%

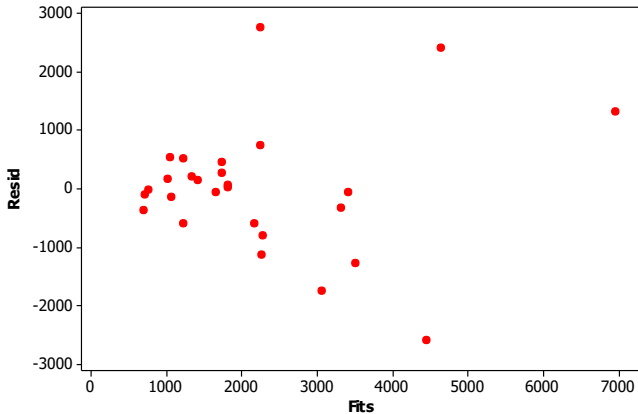
Setelah mengetahui estimasi parameter regresi Binomial Negatif, selanjutnya adalah melakukan pengujian signifikansi parameter secara serentak dan parsial. Berdasarkan hasil pengujian serentak dengan taraf signifikansi 5%, didapatkan $\chi^2_{(6;0,05)}$ sebesar 12,591. Nilai tersebut lebih kecil dari nilai *deviance*, yaitu 27,719 sehingga dapat diambil keputusan tolak H_0 yang berarti bahwa paling sedikit ada satu variabel prediktor yang berpengaruh signifikan terhadap variabel respon.

Kemudian dilanjutkan dengan melakukan pengujian signifikansi parameter secara parsial. Hasil dari uji parsial pada Tabel 4.7 didapatkan tiga variabel yang signifikan, yaitu jumlah keluarga miskin (X_1), rata-rata konsumsi protein (X_3), dan kepadatan penduduk (X_8). Berikut merupakan model regresi Binomial Negatif yang dihasilkan.

$$\hat{\mu} = \exp(4,923 + 0,000005238X_1 - 0,1414X_3 + 0,0001211X_8)$$

Berdasarkan model regresi Binomial Negatif yang terbentuk dapat disimpulkan bahwa untuk setiap penambahan 1 keluarga miskin (X_1) maka akan meningkatkan rata-rata jumlah kasus tuberkulosis sebesar 1,000005 ($e^{5,238 \times 10^{-6}}$) kasus dengan asumsi variabel lain konstan. Kemudian setiap penambahan rata-rata konsumsi protein (X_3) sebesar 1 gram maka akan menurunkan rata-rata jumlah kasus tuberkulosis sebanyak 0,8880 ($e^{-0,1187}$) kasus dengan asumsi variabel lain konstan. Selanjutnya setiap penambahan kepadatan penduduk (X_8) maka akan meningkatkan rata-rata jumlah kasus tuberkulosis sebanyak 1,000121 ($e^{1,211 \times 10^{-4}}$) kasus dengan asumsi variabel lain konstan.

Setelah diketahui model regresi Binomial Negatif, kemudian dilakukan pemeriksaan asumsi residual identik. Pemeriksaan residual identik digunakan untuk melihat homogenitas dari varian residual. Pemeriksaan asumsi identik ini dapat dilakukan secara visual. Suatu data dikatakan memenuhi asumsi identik apabila *plot* residualnya menyebar secara acak dan tidak membentuk pola tertentu. Berikut merupakan hasil *plot residual* dengan *fits* yang didapatkan.

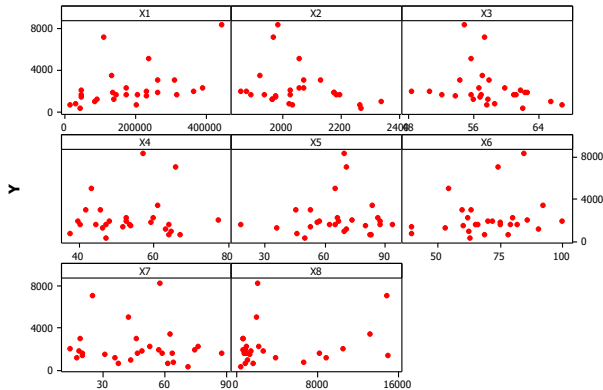


Gambar 4.2 *Scatter Plot* Residual dengan Fits Model Regresi Binomial Negatif.

Gambar 4.2 menunjukkan bahwa asumsi residual identik tidak terpenuhi. Hal ini dikarenakan titik-titik pada *scatter plot* tersebut tidak menyebar atau mengumpul di suatu titik tertentu. Oleh karena itu, asumsi residual identik pada model regresi Binomial Negatif tidak terpenuhi.

4.4 Pemodelan Banyaknya Kasus Tuberkulosis di Kabupaten/ Kota di Jawa Barat Menggunakan Analisis Regresi Non-Linier

Sebelum melakukan pemodelan menggunakan regresi nonlinier, akan diperiksa terlebih dahulu hubungan nonlinier masing-masing variabel prediktor dengan variabel respon. Identifikasi awal untuk menentukan adanya hubungan nonlinier adalah secara visual, yaitu dengan menggunakan *scatter plot* antara variabel respon dengan masing-masing variabel prediktor. Setelah dilakukan identifikasi awal menggunakan *scatter plot*, selanjutnya dapat dilakukan uji deteksi nonlinier dengan menggunakan uji Terasvirta. *Scatter plot* yang didapatkan antara variabel respon dengan masing-masing variabel prediktor terdapat pada Gambar 4.3.



Gambar 4.3 Scatter Plot antara Variabel Respon dan Variabel Prediktor.

Gambar 4.3 menunjukkan bahwa hubungan dari variabel respon dengan masing-masing variabel prediktor tidak menunjukkan adanya hubungan linier karena titik-titik pengamatan tidak mengikuti atau mendekati garis linier. Selanjutnya dilakukan uji Terasvirta untuk mendeteksi adanya hubungan nonlinier antara variabel respon dengan masing-masing variabel prediktor dalam Tabel 4.8.

Tabel 4.8 Hasil Pengujian Terasvirta

Variabel	P-value
X ₁	0,0095
X ₂	0,3789
X ₃	0,3049
X ₄	0,8845
X ₅	0,2314
X ₆	0,9129
X ₇	0,7771
X ₈	0,2148

Berdasarkan hasil uji Terasvirta, dapat disimpulkan bahwa variabel prediktor yang memiliki hubungan nonlinier dengan variabel respon adalah variabel jumlah keluarga miskin (X₁) karena p-value yang dihasilkan kurang dari 0,05. Pendeteksian hubungan

nonlinier menggunakan *scatterplot* dan uji Terasvirta didapatkan hasil yang berbeda, sehingga pada analisis selanjutnya akan tetap digunakan kedelapan variabel prediktor.

4.4.1 Regresi Nonlinier Menggunakan 8 Variabel Prediktor

Pada penelitian ini, model regresi nonlinier yang digunakan adalah model eksponensial. Algoritma *Levenberg Marquardt* menggunakan prosedur iterasi. Proses minimasi dimulai dengan melakukan perkiraan nilai dari parameter vektor. Nilai tersebut didapatkan dari nilai koefisien pada pemodelan regresi linier, dimana nilai tersebut terdapat pada Tabel 4.9.

Tabel 4.9 Nilai Iniliasi Parameter dengan Seluruh Variabel

Parameter	Koefisien
β_0	5,457
β_1	0,00000506
β_2	0,002641
β_3	-0,09823
β_4	0,01801
β_5	-0,000549
β_6	0,000561
β_7	-0,001082
β_8	0,00010550

Berdasarkan nilai pada Tabel 4.9, maka selanjutnya dilakukan regresi nonlinier menggunakan model eksponensial. Estimasi parameter model regresi nonlinier didapatkan hasil sebagai berikut.

Tabel 4.10 Estimasi Parameter Regresi Non Linier dengan 8 Variabel

Parameter	Estimate	Std.Error	t _{hit}
β_0	5,576	1,861	2,996
β_1	$6,708 \times 10^{-7}$	$1,605 \times 10^{-7}$	4,180
β_2	$0,368 \times 10^{-3}$	$0,324 \times 10^{-3}$	1,138
β_3	$-1,340 \times 10^{-2}$	$0,905 \times 10^{-2}$	-1,480

Tabel 4.10 Estimasi Parameter Regresi Non Linier dengan 8 Variabel (Lanjutan)

Parameter	Estimate	Std.Error	t_{hit}
β_4	$2,570 \times 10^{-3}$	$0,232 \times 10^{-2}$	1,108
β_5	$-0,110 \times 10^{-3}$	$0,102 \times 10^{-2}$	-0,108
β_6	$0,083 \times 10^{-3}$	$0,131 \times 10^{-2}$	0,063
β_7	$-0,017 \times 10^{-2}$	$0,718 \times 10^{-3}$	-0,241
β_8	$0,014 \times 10^{-3}$	$4,943 \times 10^{-6}$	2,850

Setelah mengetahui estimasi parameter regresi nonlinier, selanjutnya adalah melakukan pengujian signifikansi parameter secara serentak dan parsial. Berdasarkan hasil pengujian serentak dengan taraf signifikansi 5%, didapatkan $F_{0,05;8;18}$ sebesar 2,51. Nilai tersebut lebih kecil dari nilai F_{hit} , yaitu 634,9 sehingga dapat diambil keputusan tolak H_0 yang berarti bahwa paling sedikit ada satu variabel prediktor yang berpengaruh signifikan terhadap jumlah kasus tuberkulosis di Jawa Barat.

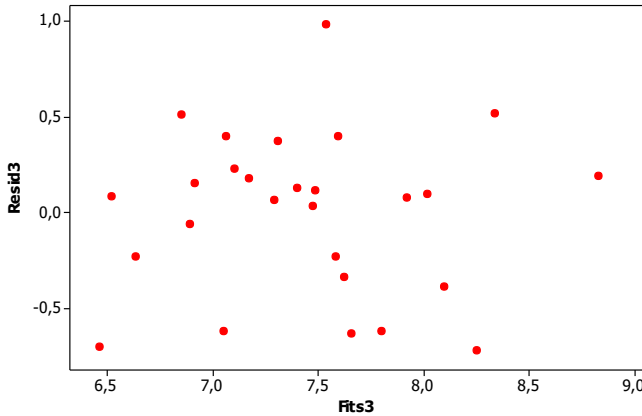
Kemudian dilanjutkan dengan melakukan pengujian signifikansi parameter secara parsial. Hasil dari uji parsial pada Tabel 4.10 didapatkan dua parameter yang signifikan, yaitu β_1 dan β_8 karena memiliki nilai $|t_{hit}|$ lebih dari $t_{19;0,05}$ (1,729). Oleh karena itu, variabel jumlah keluarga miskin (X_1) dan kepadatan penduduk (X_8) berpengaruh signifikan terhadap jumlah kasus tuberkulosis di Jawa Barat. Berikut merupakan model regresi nonlinier yang terbentuk.

$$\hat{y} = 5,576 \exp(0,0000006708X_1 + 0,000014X_8)$$

Berdasarkan model regresi nonlinier yang terbentuk dapat disimpulkan bahwa untuk setiap penambahan 1 keluarga miskin (X_1) maka akan meningkatkan rata-rata jumlah kasus tuberkulosis sebesar 1,000001 ($e^{6,708 \times 10^{-7}}$) kasus dengan asumsi variabel lain konstan. Kemudian setiap penambahan kepadatan penduduk (X_8) maka akan meningkatkan rata-rata jumlah kasus tuberkulosis sebanyak 1,000014 ($e^{0,014 \times 10^{-3}}$) kasus dengan asumsi variabel lain konstan.

Setelah diketahui model regresi nonlinier, kemudian dilakukan pemeriksaan asumsi residual identik. Pemeriksaan ini dapat di-

lakukan secara visual. Berikut merupakan hasil *plot residual* dengan *fits* yang didapatkan.



Gambar 4.4 *Scatter Plot* Residual dengan Fits Model Regresi Nonlinier Menggunakan 8 Variabel Prediktor.

Gambar 4.4 menunjukkan bahwa asumsi residual identik telah terpenuhi. Hal ini dikarenakan titik-titik pada *scatter plot* tersebut menyebar atau tidak mengumpul di suatu titik tertentu. Oleh karena itu, asumsi residual identik pada model regresi nonlinier menggunakan seluruh variabel prediktor telah terpenuhi.

4.4.2 Regresi Nonlinier Menggunakan Variabel Signifikan

Setelah diketahui estimasi parameter regresi nonlinier menggunakan 8 variabel prediktor, selanjutnya dilakukan pemodelan regresi nonlinier dengan menggunakan dua variabel prediktor yang signifikan. Model regresi nonlinier yang digunakan adalah model eksponensial. Estimasi parameter model regresi nonlinier menggunakan non linier *least squares*, dimana metode iterasinya menggunakan Algoritma *Levenberg Marquardt*. Proses minimasi dimulai dengan melakukan perkiraan nilai dari parameter vektor. Nilai tersebut didapatkan dari nilai koefisien pada pemodelan regresi

linier antara variabel respon dengan dua variabel yang signifikan, dimana nilai tersebut terdapat pada Tabel 4.11.

Tabel 4.11 Nilai Iniliasi Parameter dengan Variabel Signifikan

Parameter	Koefisien
β_0	6,1405
β_1	0,00000514
β_8	0,00009669

Berdasarkan nilai pada Tabel 4.11, maka selanjutnya dilakukan analisis regresi nonlinier dengan dua variabel signifikan menggunakan model eksponensial. Estimasi parameter model regresi nonlinier didapatkan hasil sebagai berikut.

Tabel 4.12 Estimasi Parameter Regresi Non Linier dengan Variabel Signifikan

Parameter	Estimate	Std.Error	t_{hit}
β_0	6,245	0,211	29,496
β_1	$6,779 \times 10^{-7}$	$1,212 \times 10^{-7}$	5,594
β_8	$0,013 \times 10^{-3}$	$3,053 \times 10^{-6}$	4,257

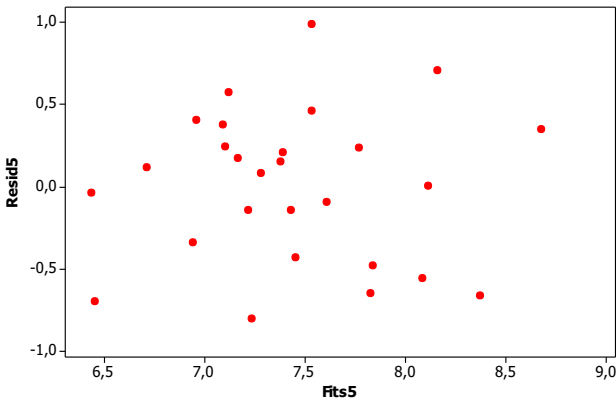
Hasil pengujian secara serentak dengan taraf signifikansi 5% didapatkan $F_{0,05;2;24}$ sebesar 3,40. Nilai tersebut lebih kecil dari nilai F_{hit} , yaitu 2170 sehingga dapat diambil keputusan tolak H_0 yang berarti bahwa paling sedikit ada satu variabel prediktor yang berpengaruh signifikan terhadap jumlah kasus tuberkulosis di Jawa Barat.

Selanjutnya dilakukan pengujian signifikansi parameter secara parsial. Hasil dari uji parsial pada didapatkan dua parameter yang signifikan, yaitu β_1 dan β_8 karena memiliki nilai $|t_{hit}|$ lebih dari $t_{25;0,05}$ (1,708). Oleh karena itu, variabel jumlah keluarga miskin (X_1) dan kepadatan penduduk (X_8) berpengaruh signifikan terhadap jumlah kasus tuberkulosis di Jawa Barat, sehingga didapatkan model regresi nonlinier menggunakan variabel signifikan sebagai berikut.

$$\hat{y} = 6,245 \exp(0,000006779X_1 + 0,000013X_8)$$

Berdasarkan model di atas, dapat disimpulkan bahwa untuk setiap penambahan 1 keluarga miskin (X_1) maka akan meningkatkan rata-rata jumlah kasus tuberkulosis sebesar 1 ($e^{6,708 \times 10^{-7}}$) kasus dengan asumsi variabel lain konstan. Kemudian setiap penambahan kepadatan penduduk (X_8) maka akan meningkatkan rata-rata jumlah kasus tuberkulosis sebanyak 1,000013 ($e^{0,014 \times 10^{-3}}$) kasus dengan asumsi variabel lain konstan.

Setelah diketahui model regresi nonlinier menggunakan dua variabel signifikan, kemudian dilakukan pemeriksaan asumsi residual identik. Pemeriksaan ini dapat dilakukan secara visual menggunakan *scatter plot* antara nilai *residual* dan *fits* yang dihasilkan dari pemodelan regresi non linier seperti pada Gambar 4.5.



Gambar 4.5 *Scatter Plot* Residual dengan Fits Model Regresi Nonlinier Menggunakan Variabel Signifikan.

Scatter plot antara nilai *residual* dan *fits* yang dihasilkan dari pemodelan regresi nonlinier menggunakan dua variabel signifikan pada Gambar 4.5 menunjukkan hasil bahwa asumsi residual identik telah terpenuhi. Hal ini dikarenakan titik-titik merah pada *scatter plot* tersebut menyebar atau tidak mengumpul di suatu titik tertentu. Oleh karena itu, asumsi residual identik pada model regresi nonlinier menggunakan dua variabel signifikan telah terpenuhi.

4.4.3 Perbandingan Model Regresi Nonlinier

Perbandingan model regresi nonlinier dilakukan untuk mengetahui model regresi nonlinier yang paling baik. Perbandingan model regresi nonlinier menggunakan 8 variabel prediktor dengan menggunakan dua variabel signifikan dilakukan berdasarkan kriteria nilai MSE dan S terkecil yang dihasilkan. Nilai MSE dan S yang dihasilkan dari kedua model regresi nonlinier terdapat pada Tabel 4.13.

Tabel 4.13 Perbandingan Model Regresi Nonlinier

Model	MSE	S
Regresi Nonlinier 8 Variabel	0,236	0,486
Regresi Nonlinier Variabel Signifikan	0,230	0,480

Model regresi nonlinier dengan nilai MSE dan S yang didapatkan untuk masing-masing model regresi ditunjukkan pada Tabel 4.13. Semakin kecil nilai MSE dan S maka model regresi nonlinier semakin baik. Berdasarkan hasil dari estimasi parameter, didapatkan model regresi nonlinier menggunakan variabel yang signifikan merupakan model regresi non linier yang paling baik karena menghasilkan nilai MSE dan S yang lebih kecil. Regresi nonlinier dengan menggunakan dua variabel signifikan menghasilkan nilai MSE sebesar 0,230, sedangkan nilai S sebesar 0,480. Oleh karena itu, dapat disimpulkan bahwa model regresi nonlinier menggunakan dua variabel signifikan merupakan model yang lebih baik daripada menggunakan seluruh variabel prediktor. Berdasarkan hasil tersebut, maka model regresi non linier menggunakan dua variabel signifikan akan dibandingkan dengan regresi Poisson dan regresi Binomial Negatif untuk mendapatkan model regresi yang paling baik.

4.5 Pemilihan Model Terbaik

Perbandingan model regresi Poisson, regresi Binomial Negatif, dan regresi nonlinier dilakukan untuk mengetahui model yang lebih baik digunakan dalam pemodelan jumlah kasus tuberkulosis

tiap kabupaten/kota di Jawa Barat. Kriteria pemilihan model terbaik yang digunakan adalah AIC terkecil.

Tabel 4.14 Pemilihan Model Terbaik

Model	Nilai AIC
Regresi Poisson	10043,00
Regresi Binomial Negatif (6 Variabel Prediktor)	447,04
Regresi Nonlinier (Variabel Signifikan)	170,18

Hasil perbandingan nilai AIC diperoleh nilai AIC terkecil terdapat pada model regresi Nonlinier dengan menggunakan variabel signifikan. Oleh karena itu model terbaik untuk jumlah kasus tuberkulosis tiap kabupaten/kota di Jawa Barat diperoleh dari model regresi Nonlinier.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil analisis dan pembahasan maka dapat diperoleh kesimpulan sebagai berikut.

1. Secara umum, gambaran jumlah kasus tuberkulosis di kabupaten/kota di Jawa Barat yang tertinggi terdapat pada Kabupaten Bogor dan yang terendah terdapat pada Kabupaten Pangandaran. Terdapat beberapa kabupaten/kota yang masih berada di atas rata-rata, yaitu Kabupaten Bogor, Kota Bandung, Kabupaten Bandung, Kota Bekasi, Kabupaten Cianjur, Kabupaten Sukabumi, dan Kabupaten Cirebon.
2. Pemodelan jumlah kasus tuberkulosis di Jawa Barat menggunakan regresi Poisson menghasilkan bahwa terdapat tujuh variabel prediktor yang berpengaruh signifikan terhadap jumlah kasus tuberkulosis di Jawa Barat, yaitu jumlah keluarga miskin, rata-rata konsumsi energi perkapita perhari, rata-rata konsumsi protein perkapita perhari, persentase rumah tangga ber-PHBS, persentase penduduk dengan akses sanitasi layak, persentase rumah sehat, dan kepadatan penduduk. Kemudian dari regresi Poisson dapat diketahui bahwa data jumlah kasus tuberkulosis di Jawa Barat terdapat kasus overdispersi.
3. Pemodelan regresi Binomial Negatif untuk kombinasi variabel prediktor menghasilkan model dengan jumlah parameter signifikan dan AIC terkecil terdapat pada kombinasi enam variabel prediktor. Kemudian dari enam variabel prediktor tersebut didapatkan tiga variabel yang berpengaruh signifikan terhadap jumlah kasus tuberkulosis di Jawa Barat, yaitu variabel jumlah keluarga miskin, rata-rata konsumsi protein perkapita perhari, dan kepadatan penduduk.
4. Pemodelan menggunakan regresi nonlinier yang menghasilkan nilai MSE dan S yang paling kecil adalah menggunakan dua variabel prediktor yang signifikan, yaitu variabel jumlah keluarga miskin dan kepadatan penduduk.

5. Perbandingan regresi Poisson, regresi Binomial Negatif, dan regresi Nonlinier didapatkan hasil bahwa model regresi dengan nilai AIC terkecil adalah menggunakan regresi Nonlinier, sehingga regresi nonlinier merupakan pemodelan yang paling tepat untuk memodelkan jumlah kasus tuberkulosis di Jawa Barat.

5.2 Saran

Berdasarkan penelitian yang telah dilakukan, saran yang direkomendasikan untuk penelitian selanjutnya adalah penambahan variabel prediktor yang memberikan pengaruh meningkatnya jumlah kasus tuberkulosis di Provinsi Jawa Barat. Kemudian untuk analisis menggunakan regresi nonlinier direkomendasikan untuk menggunakan Algoritma *Leverberg-Marquardt* dan Algoritma *Genetika*, sehingga dapat dibandingkan untuk mendapatkan model regresi nonlinier yang paling tepat.

DAFTAR PUSTAKA

- Agresti, A. (1990). *Categorical Data Analysis Second Edition*. New York: John Wiley & Sons.
- Cameron, A. C. dan Trivedi, P. K. (1998). *Regression Analysis of Count Data*. Cambridge: Cambridge University Press.
- Crofton, J., Horne, N., dan Miller, F. (2002). *Tuberkulosis Klinis*. Alih Bahasa: Harun M., Sutiono E., Citraningtyas T., Cho P., Noviani E. D., Abidin A. N., Jakarta: Widya Medika.
- Darnah. (2011). Mengatasi Overdispersi pada Model Regresi Poisson dengan Generalized Poisson Regression I. *Jurnal Eksponensial*, 2(2), 5-10.
- Dinas Kesehatan Provinsi Jawa Timur. (2013). *Profil Kesehatan Provinsi Jawa Timur Tahun 2012*. Surabaya : Dinas Kesehatan Provinsi Jawa Timur.
- Departemen Kesehatan RI. (2009). *Buku Saku Kader Program Penanggulangan TB*. Tangerang : Direktorat Jenderal Pengendalian Penyakit dan Penyehatan Lingkungan Departemen Kesehatan RI.
- Departemen Kesehatan RI. (2013). *Pedoman Survei Kesehatan Rumah Tangga*. Departemen Kesehatan RI.
- Draper, N. H. dan Smith, H. (1992). *Analisis Regresi Terapan*. Alih Bahasa : Bambang Sumantri, Jakarta: PT Gramedia Pustaka Utama.
- Entjang, I. (2000). *Ilmu Kesehatan Masyarakat*. Bandung: PT Citra Aditya Bakti.
- Famoye, F., Wulu, J. T., dan Singh, K. P. (2004). On The Generalized Poisson Regression Model with an Application to Accident Data. *Journal of Data Science* 2, 287-295.
- Harifuddin. (2007). Estimasi Kebutuhan Daya Listrik Sulawesi Selatan Sampai Tahun 2017. *Media Elektrik, Volume 2, Nomor 2*.
- Hilbe, J. M. (2011). *Negative Binomial Regression 2nd edition*. New York: Cambridge University Press.

- Hosmer, D. W. dan Lemeshow S. (1995). *Applied Logistic Regression*. New York : John Wiley & Sons.
- Indahwati, S. dan Salamah, M. (2016). Analisis Faktor-Faktor yang Mempengaruhi Jumlah Kasus Tuberkulosis di Surabaya Menggunakan *Geographically Weighted Negative Binomial Regression*. *Jurnal Sains dan Seni ITS Vol. 5 No. 2*.
- Jayadewa, O. F., Irhamah, Kusriani, D. E. (2012). *Pemodelan Biaya Tak Langsung Proyek Konstruksi di PT. Wijaya Karya (Studi Kasus: Proyek Konstruksi di Provinsi Kalimantan Timur)*. Surabaya: Tugas Akhir Jurusan Statistika FMIPA-ITS Surabaya.
- Meilania, I. (2016). Faktor-Faktor yang Mempengaruhi Jumlah Penderita Tuberkulosis di Kabupaten Probolinggo Tahun 2014 Menggunakan Pendekatan Generalized Poisson Regression (GPR). Surabaya: Tugas Akhir Jurusan Statistika FMIPA-ITS Surabaya.
- Neter, J., Kutner, M. H., Nachtsheim, C. J., dan Wasserman, W. (1996). *Applied Linear Regression Models*. New York: The Mc-Graw Hill Company.
- Notoatmodjo, S. (2003). *Ilmu Kesehatan Masyarakat, Prinsip-Prinsip Dasar*. Jakarta: Rineka Cipta.
- Purwanti, N. (2015). *Pemodelan Infeksi Tuberkulosis Paru Berdasarkan Tingkat Ketahanan Pangan Rumah Tangga di Wilayah Pesisir Pantai Surabaya Menggunakan Regresi Logistik Biner Stratifikasi*. Surabaya: Tugas Akhir Jurusan Statistika FMIPA-ITS Surabaya.
- Selwyn, P. A., Hartel, D., dan Lewis, V. A. (1989). A Prospective Study of the Risk of Tuberculosis among Intravenous Drug Users with Human Immunodeficiency Virus Infection. *N Engl J Med* 320: 545-550.
- Soedarto. (2009). *Penyakit Menular di Indonesia*. Jakarta: Sagung Seto.
- Terasvirta, T., Lin, C. F., dan Granger, C. W. J. (1993). Power of the Neural Networks Linearity Test. *Journal of Time Series Analysis*, 14 : 159-171.

World Health Organization (WHO). (2013). *Global Tuberculosis Report 2012*. Perancis: WHO Press.

World Health Organization (WHO). (2016). *Global Tuberculosis Report 2015*. Perancis: WHO Press.

(Halaman ini sengaja dikosongkan)

LAMPIRAN

Lampiran 1 Jumlah Kasus Tuberkulosis di Provinsi Jawa Barat Tahun 2015

No	Kabupaten / Kota	Y
1	Bogor	8271
2	Sukabumi	2970
3	Cianjur	2987
4	Bandung	5014
5	Garut	1864
6	Tasikmalaya	1864
7	Ciamis	1557
8	Kuningan	2179
9	Cirebon	2229
10	Majalengka	1575
11	Sumedang	1578
12	Indramayu	1566
13	Subang	620
14	Purwakarta	924
15	Karawang	1834
16	Bekasi	1748
17	Bandung Barat	1461
18	Pangandaran	318
19	Kota Bogor	1123
20	Kota Sukabumi	739
21	Kota Bandung	7044
22	Kota Cirebon	1536
23	Kota Bekasi	3355
23	Kota Depok	1996
25	Kota Cimahi	1312
26	Kota Tasikmalaya	1180
27	Kota Banjar	602

Lampiran 2. Variabel Prediktor Penelitian

No	Kabupaten / Kota	X ₁	X ₂	X ₃	X ₄
1	Bogor	446040	1984.97	54.85	57.1
2	Sukabumi	265480	2130.53	57.9	45.7
3	Cianjur	311110	2071.47	54.32	41.7
4	Bandung	238830	2056.56	55.67	43.3
5	Garut	365390	1876.52	48.26	39.6
6	Tasikmalaya	233240	1853.67	50.46	48.3
7	Ciamis	174530	1935.61	52.01	44.5
8	Kuningan	174790	2071.14	56.69	60.1
9	Cirebon	390540	2057.35	59.9	52.7
10	Majalengka	207150	2196.66	61.22	53.5
11	Sumedang	145340	2185.21	60.99	47.3
12	Indramayu	319530	2026.64	56.94	40.2
13	Subang	201780	2265.51	66.94	64.1
14	Purwakarta	84720	2338.96	65.61	64.8
15	Karawang	264830	2175.39	62.38	52.7
16	Bekasi	136670	2176.43	62.58	59.5
17	Bandung Barat	232690	1977.2	53.65	54
18	Pangandaran	41970	2269.04	62	47.3
19	Kota Bogor	91710	1965.1	55.9	63.4
20	Kota Sukabumi	30430	2021.94	58.64	37.6
21	Kota Bandung	110280	1967.99	57.34	65.9
22	Kota Cirebon	44960	1889.37	55.68	64.3
23	Kota Bekasi	134170	1923.03	57.11	61.3
23	Kota Depok	47130	2027.62	61.78	77.5
25	Kota Cimahi	46440	1974.64	56.72	51.8
26	Kota Tasikmalaya	140110	1965.64	57.68	46.3
27	Kota Banjar	14630	2032.03	57.55	67.3

Lampiran 2. (Lanjutan)

No	Kabupaten / Kota	X₅	X₆	X₇	X₈
1	Bogor	69.58	84.5	58.1	2014
2	Sukabumi	52.94	63.11	19.1	587
3	Cianjur	45.36	59.62	46.1	584
4	Bandung	64.99	54.43	42.6	1999
5	Garut	66.75	71.81	57.4	829
6	Tasikmalaya	57.48	70.15	74.7	680
7	Ciamis	17.92	66.18	63.6	826
8	Kuningan	86.45	62.1	76.7	950
9	Cirebon	65.97	80.18	52.8	2160
10	Majalengka	62.42	79.34	58.7	982
11	Sumedang	87.79	74.88	47	749
12	Indramayu	65.31	65.01	19.8	829
13	Subang	83.1	68.79	61.6	808
14	Purwakarta	69.36	62.45	43.7	1116
15	Karawang	87.74	100	18.6	1376
16	Bekasi	56.09	75.18	49.2	2650
17	Bandung Barat	80.55	60	31.2	1248
18	Pangandaran	49.83	63.03	71.7	387
19	Kota Bogor	70.81	90.53	36.2	8843
20	Kota Sukabumi	46.16	39.51	64.2	6593
21	Kota Bandung	70.91	74.02	25.2	14800
22	Kota Cirebon	93.75	81.8	87.9	8231
23	Kota Bekasi	83.9	92.15	63	13140
23	Kota Depok	73.49	85.93	14.1	10515
25	Kota Cimahi	52.96	39.31	20	14937
26	Kota Tasikmalaya	35.6	53.02	17.5	3831
27	Kota Banjar	82.44	78.35	37.8	1599

Lampiran 3. Output Koefisien Korelasi

Correlations: X1; X2; X3; X4; X5; X6; X7; X8							
	X1	X2	X3	X4	X5	X6	X7
X2	-0,101 0,617						
X3	-0,324 0,099	0,848 0,000					
X4	-0,414 0,032	0,104 0,606	0,410 0,034				
X5	-0,034 0,867	0,099 0,623	0,242 0,223	0,554 0,003			
X6	0,167 0,406	-0,025 0,903	0,125 0,534	0,521 0,005	0,533 0,004		
X7	-0,011 0,958	-0,067 0,741	-0,158 0,430	-0,058 0,775	0,075 0,709	0,043 0,830	
X8	-0,488 0,010	-0,407 0,035	-0,020 0,923	0,426 0,027	0,119 0,556	0,059 0,771	-0,222 0,267
Cell Contents: Pearson correlation P-Value							

Lampiran 4. Output Multikolinieritas (VIF)

Predictor	Coef	SE Coef	T	P	VIF
Constant	-8055	6928	-1,16	0,260	
x1	0,013101	0,003549	3,69	0,002	2,056
x2	9,927	6,805	1,46	0,162	8,744
x3	-287,0	193,7	-1,48	0,156	8,152
x4	67,16	49,04	1,37	0,188	2,922
x5	-8,00	21,43	-0,37	0,713	1,698
x6	-6,25	28,23	-0,22	0,827	2,017
x7	6,22	15,25	0,41	0,688	1,189
x8	0,3031	0,1077	2,81	0,012	2,972

S = 1489,25 R-Sq = 55,3% R-Sq(adj) = 35,4%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	8	49394951	6174369	2,78	0,034
Residual Error	18	39921616	2217868		
Total	26	89316568			

Lampiran 5. Syntax SAS untuk Regresi Poisson

```
data poisson;
input y x1 x2 x3 x4 x5 x6 x7 x8;
datalines;
8271 446040 1984.97 54.85 57.1 69.58 84.5 58.1 2014
2970 265480 2130.53 57.9 45.7 52.94 63.11 19.1 587
2987 311110 2071.47 54.32 41.7 45.36 59.62 46.1 584
.
.
.
602 14630 2032.03 57.55 67.3 82.44 78.35 37.8 1599
;
run;
title 'poisson model';
proc nlmixed data=poisson start hess tech=newwrap;
parms a0=0 a1=0 a2=0 a3=0 a4=0 a5=0 a6=0 a7=0 a8=0;
eta=a0+a1*x1+a2*x2+a3*x3+a4*x4+a5*x5+a6*x6+a7*x7+a
8*x8;
lambda=exp(eta);
model y ~ poisson(lambda);
predict _ll out=LL_1;
run;
```

Lampiran 6. Output SAS untuk Regresi Poisson

07:35 Thursday, May 20, 2017

The NLMIXED Procedure

Iteration History

Iter	Calls	NegLogLike	Diff	MaxGrad	Slope
1	22	128721.921	285227.5	1.093E10	-1.688E8
2*	33	31844.0322	96877.89	2.9237E9	-2195041
3*	44	22034.6274	9809.405	6.9355E9	-57591.6
4*	55	6899.58846	15135.04	1.3595E9	-26023
5*	66	5094.84987	1804.739	1.9787E8	-3220.77
6	77	5012.91199	81.93788	35944214	-156.708
7	88	5012.46344	0.448545	4793987	-0.84415
8	99	5012.45842	0.005018	840072	-0.00858
9	110	5012.45828	0.000146	140937.4	-0.00025
10	121	5012.45827	4.266E-6	24297.1	-7.29E-6

NOTE: GCONV convergence criterion satisfied.

Fit Statistics

-2 Log Likelihood	10025
AIC (smaller is better)	10043
AICC (smaller is better)	10054
BIC (smaller is better)	10055

Lampiran 6. (Lanjutan)

Parameter Estimates					
Parameter	Estimate	Standard Error	DF	t Value	Pr > t
a0	3.0020	0.000031	27	95349.3	<.0001
a1	5.492E-6	0	27	Infty	<.0001
a2	0.004819	0.000071	27	68.23	<.0001
a3	-0.1414	0.002722	27	-51.94	<.0001
a4	0.02825	0.000775	27	36.46	<.0001
a5	-0.00122	0.000366	27	-3.34	0.0025
a6	-0.00191	0.000423	27	-4.51	0.0001
a7	-0.00003	0.000239	27	-0.11	0.9132
a8	0.000131	1.138E-6	27	114.69	<.0001

Lampiran 7. Macro SAS untuk Overdispersi

```
data cari_devians;
input y x1 x2 x3 x4 x5 x6 x7 x8;
cards;
8271 446040 1984.97 54.85 57.1 69.58 84.5 58.1 2014
2970 265480 2130.53 57.9 45.7 52.94 63.11 19.1 587
2987 311110 2071.47 54.32 41.7 45.36 59.62 46.1 584
.
.
.
602 14630 2032.03 57.55 67.3 82.44 78.35 37.8 1599
;
run;
proc genmod data=cari_devians;
    model Y = x1 x2 x3 x4 x5 x6 x7 x8/dist=poisson
    link = log
type1
    type3 wald
scale=deviance;
run;
```

Lampiran 8. Output SAS untuk Overdispersi

The GENMOD Procedure			
Model Information			
Data Set		WORK.CARI_DEVIANS	
Distribution			Poisson
Link Function			Log
Dependent Variable			y
Observations Used			27
Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	18	9774.3783	543.0210
Scaled Deviance	18	18.0000	1.0000
Pearson Chi-Square	18	10155.4484	564.1916
Scaled Pearson X2	18	18.7018	1.0390
Log Likelihood		753.0279	

Lampiran 9. Syntax dan Output R Regresi Binomial Negatif dengan 7 Variabel Prediktor

```
> data=read.table("E://tujuh.txt",header=TRUE)
> library(MASS)
> negbin=glm.nb(Y~X1+X2+X3+X4+X5+X6+X8,data=data)
> summary(negbin)
```

Call:

```
glm.nb(formula = Y ~ X1 + X2 + X3 + X4 + X5 + X6 + X8,
data = data, init.theta = 6.141608359, link = log)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.89863	-0.90171	-0.05301	0.44819	2.26318

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	4.962e+00	1.787e+00	2.777	0.00549 **
X1	5.236e-06	9.328e-07	5.613	1.99e-08 ***
X2	3.488e-03	1.847e-03	1.888	0.05901 .
X3	-1.182e-01	5.238e-02	-2.256	0.02409 *
X4	1.626e-02	1.331e-02	1.221	0.22193
X5	9.043e-04	5.770e-03	0.157	0.87546
X6	-1.574e-03	7.646e-03	-0.206	0.83687
X8	1.209e-04	2.815e-05	4.295	1.75e-05 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(6.1416) family taken to be 1)

Null deviance: 84.533 on 26 degrees of freedom

Residual deviance: 27.718 on 19 degrees of freedom

AIC: 449.01

Number of Fisher Scoring iterations: 1

Theta: 6.14

Std. Err.: 1.63

2 x log-likelihood: -431.013

Lampiran 10. Syntax dan Output R Regresi Binomial Negatif dengan 6 Variabel Prediktor

```
> data=read.table("E://enam.txt",header=TRUE)
> library(MASS)
> negbin=glm.nb(Y~X1+X2+X3+X4+X6+X8,data=data)
> summary(negbin)
```

Call:

```
glm.nb(formula = Y ~ X1 + X2 + X3 + X4 + X6 + X8, data =
data, init.theta = 6.135739416, link = log)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.88078	-0.88266	-0.03877	0.42552	2.29531

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	4.923e+00	1.784e+00	2.760	0.00578 **
X1	5.238e-06	9.286e-07	5.640	1.70e-08 ***
X2	3.523e-03	1.848e-03	1.907	0.05655 .
X3	-1.187e-01	5.241e-02	-2.265	0.02349 *
X4	1.682e-02	1.254e-02	1.341	0.17981
X6	-1.173e-03	7.390e-03	-0.159	0.87384
X8	1.211e-04	2.816e-05	4.299	1.71e-05 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(6.1357) family taken to be 1)

Null deviance: 84.453 on 26 degrees of freedom

Residual deviance: 27.719 on 20 degrees of freedom

AIC: 447.04

Number of Fisher Scoring iterations: 1

Theta: 6.14

Std. Err.: 1.63

2 x log-likelihood: -431.041

Lampiran 11. Syntax dan Output R Regresi Binomial Negatif dengan 5 Variabel Prediktor

```

> data=read.table("E://lima.txt",header=TRUE)
> library(MASS)
> negbin=glm.nb(Y~X1+X2+X3+X4+X8,data=data)
> summary(negbin)

Call:
glm.nb(formula = Y ~ X1 + X2 + X3 + X4 + X8, data = data,
init.theta = 6.130985083, link = log)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.90116 -0.86554 -0.04301  0.44936  2.31850

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) 4.892e+00 1.760e+00  2.780 0.00544 **
X1          5.171e-06 8.425e-07  6.138 8.36e-10 ***
X2          3.587e-03 1.821e-03  1.970 0.04878 *
X3         -1.207e-01 5.191e-02 -2.325 0.02005 *
X4          1.573e-02 1.015e-02  1.549 0.12143
X8          1.216e-04 2.788e-05  4.362 1.29e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(6.131) family
taken to be 1)
Null deviance: 84.388  on 26  degrees of freedom
Residual deviance: 27.719  on 21  degrees of freedom
AIC: 445.06
Number of Fisher Scoring iterations: 1
      Theta: 6.13
      Std. Err.: 1.63
2 x log-likelihood: -431.063

```

Lampiran 12. Syntax dan Output R Regresi Binomial Negatif dengan 4 Variabel Prediktor

```
> data=read.table("E://empat.txt",header=TRUE)
> library(MASS)
> negbin=glm.nb(Y~X1+X3+X4+X8,data=data)
> summary(negbin)
```

Call:

```
glm.nb(formula = Y ~ X1 + X3 + X4 + X8, data = data,
init.theta = 4.628486801, link = log)
```

Deviance Residuals:

```
   Min      1Q  Median      3Q      Max
-1.7285 -0.8847 -0.2678  0.4750  2.2787
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	7.071e+00	1.449e+00	4.880	1.06e-06	***
X1	4.494e-06	9.312e-07	4.826	1.39e-06	***
X3	-2.102e-02	2.536e-02	-0.829	0.40728	
X4	1.133e-02	1.126e-02	1.006	0.31430	
X8	7.694e-05	2.445e-05	3.147	0.00165	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(4.6285) family taken to be 1)

Null deviance: 63.762 on 26 degrees of freedom

Residual deviance: 27.963 on 22 degrees of freedom

AIC: 451.1

Number of Fisher Scoring iterations: 1

Theta: 4.63

Std. Err.: 1.22

2 x log-likelihood: -439.104

Lampiran 13. Syntax dan Output R Regresi Binomial Negatif dengan 3 Variabel Prediktor

```
> data=read.table("E://tiga.txt",header=TRUE)
> library(MASS)
> negbin=glm.nb(Y~X1+X3+X8,data=data)
> summary(negbin)
```

Call:

```
glm.nb(formula = Y ~ X1 + X3 + X8, data = data, init.theta =
5.267577604, link = log)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.6153	-1.0513	-0.1463	0.4992	2.2194

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	7.054e+00	1.342e+00	5.255	1.48e-07 ***
X1	5.150e-06	8.997e-07	5.724	1.04e-08 ***
X3	-1.429e-02	2.152e-02	-0.664	0.507
X8	9.876e-05	2.148e-05	4.597	4.28e-06 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(5.2676) family taken to be 1)

Null deviance: 72.540 on 26 degrees of freedom

Residual deviance: 27.833 on 23 degrees of freedom

AIC: 445.38

Number of Fisher Scoring iterations: 1

Theta: 5.27

Std. Err.: 1.39

2 x log-likelihood: -435.378

Lampiran 14. Syntax dan Output R Regresi Binomial Negatif dengan 2 Variabel Prediktor

```
> data=read.table("E://dua.txt",header=TRUE)
> library(MASS)
> negbin=glm.nb(Y~X1+X8,data=data)
> summary(negbin)
```

Call:

```
glm.nb(formula = Y ~ X1 + X8, data = data, init.theta =
5.202570195, link = log)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.80983	-0.99569	-0.04074	0.45603	2.26745

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	6.178e+00	2.206e-01	28.001	< 2e-16 ***
X1	5.385e-06	8.365e-07	6.438	1.21e-10 ***
X8	1.011e-04	2.111e-05	4.790	1.67e-06 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(5.2026) family taken to be 1)

Null deviance: 71.647 on 26 degrees of freedom

Residual deviance: 27.844 on 24 degrees of freedom

AIC: 443.73

Number of Fisher Scoring iterations: 1

Theta: 5.20

Std. Err.: 1.38

2 x log-likelihood: -435.733

Lampiran 15. Syntax dan Output R Regresi Binomial Negatif dengan 1 Variabel Prediktor

```

> data=read.table("E://satu.txt",header=TRUE)
> library(MASS)
> negbin=glm.nb(Y~X8,data=data)
> summary(negbin)

Call:
glm.nb(formula = Y ~ X8, data = data, init.theta =
2.21870508, link = log)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0733 -0.9173 -0.3254  0.1361  2.7024
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  7.556e+00  1.683e-01  44.892 <2e-16 ***
X8           3.327e-05  2.818e-05   1.181  0.238
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(2.2187) family
taken to be 1)
Null deviance: 30.608  on 26  degrees of freedom
Residual deviance: 28.983  on 25  degrees of freedom
AIC: 466.97
Number of Fisher Scoring iterations: 1
      Theta: 2.219
    Std. Err.: 0.565
 2 x log-likelihood: -460.973

```

Lampiran 16. Output Regresi Linier Y dengan 8 Variabel**Regression Analysis: ln(Y) versus X1; X2; X3; X4; X5; X6; X7; X8**

The regression equation is

$$\ln(Y) = 5,46 + 0,000005 X1 + 0,00264 X2 - 0,0982 X3 + 0,0180 X4 - 0,00055 X5 + 0,00056 X6 - 0,00108 X7 + 0,000106 X8$$

Predictor	Coef	SE Coef	T	P
Constant	5,457	2,371	2,30	0,034
X1	0,00000506	0,00000121	4,16	0,001
X2	0,002641	0,002329	1,13	0,272
X3	-0,09823	0,06631	-1,48	0,156
X4	0,01801	0,01679	1,07	0,298
X5	-0,000549	0,007335	-0,07	0,941
X6	0,000561	0,009663	0,06	0,954
X7	-0,001082	0,005221	-0,21	0,838
X8	0,00010550	0,00003688	2,86	0,010

S = 0,509754 R-Sq = 65,0% R-Sq(adj) = 49,5%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	8	8,6941	1,0868	4,18	0,006
Residual Error	18	4,6773	0,2598		
Total	26	13,3714			

Source	DF	Seq SS
X1	1	3,8615
X2	1	1,3428
X3	1	0,2599
X4	1	0,7818
X5	1	0,0147
X6	1	0,0357
X7	1	0,2707
X8	1	2,1270

Lampiran 17. Macro SAS Regresi Non Linier (8 Variabel)

```
data kasus;
input Y X1 X2 X3 X4 X5 X6 X7 X8;
datalines;
9.0205107 446040 1984.97 54.85 57.1 69.58 84.5 58.1 2014
7.9963172 265480 2130.53 57.9 45.7 52.94 63.11 19.1 587
8.0020248 311110 2071.47 54.32 41.7 45.36 59.62 46.1 584
.
.
.
6.4002574 14630 2032.03 57.55 67.3 82.44 78.35 37.8 1599
;
proc nlin data=kasus method=MARQUARDT;
parms b0=5.457 b1=0.00000506 b2=0.002641 b3=-0.09823
b4=0.01801 b5=-0.000549 b6=0.000561 b7=-0.001082
b8=0.00010550;
model
Y=b0*exp(b1*X1+b2*X2+b3*X3+b4*X4+b5*X5+b6*X6+b
7*X7+b8*X8);
run;
```

Lampiran 18. Output SAS Regresi Nonlinier (8 Variabel)

The NLIN Procedure					
Dependent Variable Y					
Method: Marquardt					
Estimation Summary					
Method	Marquardt				
Iterations	7				
Subiterations	4				
Average Subiterations	0.571429				
R	6.398E-6				
PPC(b5)	0.000067				
RPC(b5)	0.003239				
Object	3.328E-7				
Objective	4.736677				
Observations Read	27				
Observations Used	27				
Observations Missing	0				
		Sum of	Mean		Approx
Source	DF	Squares	Square	F Value	Pr > F
Model	9	1503.7	167.1	634.90	<.0001
Error	18	4.7367	0.2631		
Uncorrected Total	27	1508.4			
		Approx	Approximate 95%		
Parameter	Estimate	Std Error	Confidence Limit		
b0	5.5756	1.8609	1.6660	9.4852	
b1	6.708E-7	1.605E-7	3.336E-7	1.008E-6	
b2	0.000368	0.000324	-0.00031	0.00105	
b3	-0.0134	0.00905	-0.0324	0.00561	
b4	0.00257	0.00232	-0.00230	0.00743	
b5	-0.00011	0.00102	-0.00225	0.00203	
b6	0.000083	0.00131	-0.00267	0.00284	
b7	-0.00017	0.000718	-0.00168	0.00133	
b8	0.000014	4.943E-6	3.703E-6	0.000024	

Lampiran 19. Output Regresi Linier Y dengan 2 Variabel Signifikan

Regression Analysis: ln(Y) versus X1; X8

The regression equation is

$$\ln(Y) = 6,14 + 0,000005 X1 + 0,000097 X8$$

Predictor	Coef	SE Coef	T	P
Constant	6,1405	0,2395	25,64	0,000
X1	0,00000514	0,00000091	5,65	0,000
X8	0,00009669	0,00002294	4,22	0,000

S = 0,477170 R-Sq = 59,1% R-Sq(adj) = 55,7%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	7,9068	3,9534	17,36	0,000
Residual Error	24	5,4646	0,2277		
Total	26	13,3714			

Source	DF	Seq SS
X1	1	3,8615
X8	1	4,0453

Lampiran 20. Macro SAS Regresi Nonlinier (2 Variabel Signifikan)

```
data kasus2;
input Y X1 X8;
datalines;
9.0205107 446040 2014
7.996317232 265480 587
8.002024818 311110 584
.
.
.
6.400257445 14630 1599
;
proc nlin data=kasus2 method=MARQUARDT;
parms b0=6.1405 b1=0.00000514 b8=0.00009669;
model Y=b0*exp(b1*X1+b8*X8);
run;
```

Lampiran 21. Output SAS Regresi Nonlinier (2 Variabel Signifikan)

The NLIN Procedure					
Dependent Variable Y					
Method: Marquardt					
Estimation Summary					
Method	Marquardt				
Iterations	6				
R	2.393E-6				
PPC(b1)	7.966E-7				
RPC(b1)	0.000045				
Object	1.894E-8				
Objective	5.538953				
Observations Read	27				
Observations Used	27				
Observations Missing	0				
Source	DF	Sum of Squares	Mean Square	F Value	Approx Pr > F
Model	3	1502.9	501.0	2170.60	<.0001
Error	24	5.5390	0.2308		
Uncorrected Total	27	1508.4			
Parameter	Estimate	Approx Std Error	Approximate 95% Confidence Limit		
b0	6.2455	0.2117	5.8085	6.6825	
b1	6.779E-7	1.212E-7	4.279E-7	9.28E-7	
b8	0.000013	3.053E-6	6.697E-6	0.000019	

Lampiran 22. Surat Pernyataan Legalitas Data

SURAT PERNYATAAN

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

Nama : Dwi Puspita Firdaus
NRP : 1313100023

menyatakan bahwa data yang digunakan dalam Tugas Akhir/Thesis ini merupakan data sekunder yang diambil dari penelitian /-buku/-Tugas Akhir/Thesis/publikasi lainnya yaitu:

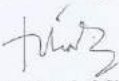
Sumber : Data Publikasi Dinas Kesehatan Provinsi Jawa Barat dan Badan Pusat Statistik (BPS) Provinsi Jawa Barat.

Keterangan : Profil Kesehatan Jawa Barat 2015 dan Jawa Barat dalam Angka 2016 yang dapat di unduh di web Dinas Kesehatan dan BPS Provinsi Jawa Barat.

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

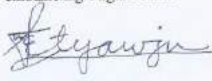
Co-Pembimbing Tugas Akhir

Surabaya, 16 Juli 2016


(Dr. Irhamah, S.Si., M.Si.)
NIP. 19780406 200112 2 002


(Dwi Puspita Firdaus)
NRP. 1313100023

Mengetahui,
Pembimbing Tugas Akhir


(Dra. Wiwiek Setya Winahju, M.S.)
NIP. 19560424 198303 2 001

*(coret yang tidak perlu)

BIODATA PENULIS



Penulis memiliki nama lengkap DWI PUSPITA FIRDAUS yang sering dipanggil Firda. Penulis lahir di Surabaya pada tanggal 19 Desember 1995 dan juga merupakan anak kedua dari empat bersaudara dari pasangan Bapak Syamsul Ma'arif dan Ibu Siti Munawaroh. Pendidikan formal yang telah ditempuh penulis antara lain TK Belia Surabaya, SMP Al-Hikmah Surabaya, dan SMA Al-Hikmah

Surabaya. Pada tahun 2013, penulis diterima di Jurusan Statistika ITS melalui jalur SNMPTN dengan NRP 1313100023. Kemudian penulis lulus pada tahun 2017 dengan menyelesaikan Tugas Akhir yang berjudul **“PEMODELAN JUMLAH KASUS TUBERKULOSIS DI PROVINSI JAWA BARAT MENGGUNAKAN REGRESI BINOMIAL NEGATIF DAN REGRESI NONLINIER”**. Bagi pembaca yang memiliki saran, kritik, atau ingin berdiskusi lebih lanjut terkait dengan Tugas Akhir ini dapat disampaikan melalui email: dpuspitafirdaus@gmail.com.