PEMODELAN POPULASI WERENG HIJAU
PENYEBAB PENYAKIT TUNGRO PADA PADI DI KABUPATEN
BONDOWOSO DENGAN REGRESI *ZERO-INFLATED* POISSON

SKRIPSI

Oleh

Ervin Yulia
NIM 091810101008

JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS JEMBER

2013

PEMODELAN POPULASI WERENG HIJAU PENYEBAB PENYAKIT TUNGRO PADA PADI DI KABUPATEN BONDOWOSO DENGAN REGRESI *ZERO-INFLATED* POISSON

SKRIPSI

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

Oleh

Ervin Yulia
NIM 091810101008

JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS JEMBER
2013

PERSEMBAHAN

Skripsi ini saya persembahkan untuk:

1. Ibunda Hj. Insiyah dan Ayahanda H. Sutrisno, SP. yang tercinta;
2. Rini Sulistyowati dan M. Nuril Huda yang tersayang;
3. guru-guru sejak taman kanak-kanak sampai dengan perguruan tinggi;
4. Almamater Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.

MOTO

Pendidikan merupakan perlengkapan paling baik untuk hari tua. [*)]

Kita sembuh dari penyakit karena kita berobat. Kita bergerak berobat untuk mendapatkan kesehatan bukan karena uang, gaji, juru rawat atau dokter tapi kita bergerak karena diizinkan oleh Allah SWT.[**)]

[*)] Aristoteles. Kumpulan Motto Kehidupan [on line]. http://pristality.wordpress.com/2011/02/23/kumpulan-motto-kehidupan/ [13 Mei 2013]

[**)] Ust.Yusuf Mansyur .2010. *Temukan Penyebabnya Temukan Jawabannya*. Jakarta: PT Bestari Buana Murni

PERNYATAAN

Saya yang bertanda tangan di bawah ini:

nama : Ervin Yulia

NIM : 091810101008

menyatakan dengan sesungguhnya bahwa karya ilmiah yang berjudul “Pemodelan Populasi Wereng Hijau Penyebab Penyakit Tungro pada Padi di Kabupaten Bondowoso dengan Regresi *Zero-Inflated* Poisson” adalah benar-benar hasil karya sendiri, kecuali kutipan yang sudah saya sebutkan sumbernya, belum pernah diajukan pada institusi mana pun, dan bukan karya jiplakan. Saya bertanggung jawab atas keabsahan dan kebenaran isinya sesuai dengan sikap ilmiah yang harus dijunjung tinggi.

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

Jember, 28 Mei 2013

Yang menyatakan,

Ervin Yulia

NIM 091810101008

SKRIPSI

PEMODELAN POPULASI WERENG HIJAU PENYEBAB PENYAKIT TUNGRO PADA PADI DI KABUPATEN BONDOWOSO DENGAN REGRESI *ZERO-INFLATED* POISSON

Oleh

Ervin Yulia
NIM 091810101008

Pembimbing

Dosen Pembimbing Utama : Dr. Alfian Futuhul Hadi, S.Si, M.Si.
Dosen Pembimbing Anggota : Yuliani Setia Dewi, S.Si, M.Si.

PENGESAHAN

Skripsi yang berjudul "Pemodelan Populasi Wereng Hijau Penyebab Penyakit Tungro pada Padi di Kabupaten Bondowoso dengan Regresi *Zero-Inflated* Poisson" telah diuji dan disahkan pada:

hari, tanggal :

tempat : Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember

Tim Penguji:

Dosen Pembimbing Utama,	Dosen Pembimbing Anggota,
Dr. Alfian Futuhul Hadi, S.Si, M.Si. NIP 197407192000121001	Yuliani Setia Dewi, S.Si, M.Si. NIP 197407162000032001
Penguji I,	Penguji II,
Prof. Drs. I Made Tirta,M.Sc.,Ph.D. NIP 195912201985031002	Kusbudiono, S.Si, M.Si. NIP 197704302005011001

Mengesahkan

Dekan,

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

NIP 196101081986021001

RINGKASAN

Pemodelan Populasi Wereng Hijau Penyebab Penyakit Tungro pada Padi di Kabupaten Bondowoso dengan Regresi *Zero-Inflated* Poisson. Ervin Yulia, 091810101008; 2013: 39 halaman; Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.

Penyakit tungro padi, atau disebut tungro merupakan penyakit padi yang berbahaya dan memerlukan penanganan dengan baik. Penyakit tungro disebabkan oleh Virus Tungro Padi (VTP) dan ditularkan oleh hama wereng hijau. Hama wereng kini meresahkan petani di Kabupaten Bondowoso. Keterkaitan faktor-faktor penyebab serangan penyakit tungro dengan banyaknya wereng hijau penular dapat didekati dengan analisis statistika mengenai hubungan variabel takbebas dan variabel bebas, yaitu analisis regresi. Populasi wereng hijau yang didapatkan dalam setiap pengamatan di daerah terserang penyakit tungro relatif jarang sehingga menghasilkan banyak nilai nol dalam beberapa pengamatan. Distribusi Poisson sering digunakan dalam pemodelan kasus yang jarang terjadi. Populasi wereng hijau yang menghasilkan banyak nilai nol dalam pengamatannya dapat dianalisis dengan penggunakan model regresi *Zero-inflated* Poisson. Tujuan penelitian ini adalah untuk mendapatkan model terbaik terhadap banyaknya populasi wereng hijau penyebab penyakit tungro yang banyak mengandung nilai nol dengan beberapa variabel bebas yang mempengaruhi menggunakan *Zero-inflated* Poisson.

Penelitian dilakukan dalam beberapa langkah. Langkah pertama melakukan pengamatan untuk mengetahui populasi wereng hijau. Langkah kedua melakukan pengujian model regresi Poisson dengan menggunakan software program R secara full dan saturated model. Langkah ketiga Melakukan pengujian model regresi Zero-*inflated* Poisson dengan menggunakan software program R secara full dan saturated model. Dan langkah keempat Membandingkan model-

model yang telah didapatkan pada pengujian model regresi Poisson dan *Zero-inflated* Poisson dengan melihat nilai log-likelihood sehingga didapatkan model terbaik.

Berdasarkan kajian yang telah dilakukan, dari empat variabel bebas yaitu luas lahan sawah, umur tanaman terserang, adanya tanaman inang atau gulma, dan curah hujan diperoleh nilai log-likelihood pada model regresi ZIP selalu menghasilkan nilai lebih besar dari pada nilai log-likelihood yang dihasilkan pada model regresi Poisson untuk semua model. Dari pemodelan regresi ZIP didapatkan model terbaik untuk pemodelan populasi wereng hijau penyebab penyakit tungro di Kabupaten Bondowoso adalah model regresi ZIP dengan variabel bebas X_1, X_3 dan X_4, dimana X_1 adalah luas lahan sawah, X_3 adalah adanya tanaman inang atau gulma, dan X_4 merupakan curah hujan.

PRAKATA

Puji syukur ke hadirat Allah SWT atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul "Pemodelan Populasi Wereng Hijau Penyebab Penyakit Tungro pada Padi di Kabupaten Bondowoso dengan Regresi *Zero-Inflated* Poisson". Skripsi ini disusun untuk memenuhi salah satu syarat menyelesaikan pendidikan strata satu (S1) pada Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.

Penulisan skripsi ini tidak lepas dari bantuan berbagai pihak. Oleh karena itu, penulis menyampaikan terima kasih kepada:

1. Dr. Alfian Futuhul Hadi, S.Si, M.Si. selaku Dosen Pembimbing Utama, Yuliani Setia Dewi, S.Si, M.Si. selaku Dosen Pembimbing Anggota yang telah meluangkan waktu, pikiran, dan perhatian dalam penulisan skripsi ini;
2. Prof. Drs. I Made Tirta, M.Sc., PhD. dan Kusbudiono, S.Si, M.Si. selaku dosen penguji yang telah memberi masukan dalam skripsi ini;
3. Kosala Dwidja Purnomo, S.Si, M.Si. selaku Dosen Pembimbing Akademik yang telah membimbing selama penulis menjadi mahasiswa;
4. Ibunda Hj. Insiyah, Ayahanda H. Sutrisno, SP, dan adik-adik saya tersayang yang telah memberikan doa dan dorongannya demi terselesaikannya skripsi ini;
5. Jefri Rieski Triyanto yang sabar dan penuh pengertian dalam menemani serta mendukung segala usaha dalam penyelesaian tugas akhir ini ;
6. teman-teman A FIRE LIFE (Aan, Fendi, Ifa Nur, Rizka, Elna, Lutfi, Ifa Latifatur, Fathur, dan Ervin), sista-sista kosan Belitung II (Pipit, Winda, Nirka, Melani, Syarifa, dan Elita), angkatan 2009 (MALINC), kakak serta adik angkatan Jurusan Matematika MIPA, dan yang lain;
7. semua pihak yang tidak dapat disebutkan satu per satu.

Penulis juga menerima segala kritik dan saran dari semua pihak demi kesempurnaan skripsi ini. Akhirnya penulis berharap, semoga skripsi ini dapat bermanfaat.

Jember, Mei 2013 Penulis

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN

Halaman

BAB. 1 PENDAHULUAN

1.1 Latar Belakang

Indonesia adalah salah satu negara yang sebagian besar penduduknya hidup dari hasil pertanian padi. Setiap tahun produksi padi perlu ditingkatkan untuk memenuhi kebutuhan konsumen yang terus bertambah. Kabupaten Bondowoso adalah salah satu kabupaten yang sebagian besar wilayahnya merupakan lahan pertanian, yaitu seluas 33.682 Ha (Dinas Pertanian Bondowoso, 2012). Seiring dengan pertumbuhan penduduk di Kabupaten Bondowoso dan sekitarnya menuntut suplai beras yang seimbang. Berbagai permasalahan muncul dalam upaya peningkatan produksi padi, diantaranya penyakit tanaman. Penyakit tungro padi, atau disebut tungro merupakan penyakit padi yang berbahaya dan memerlukan penanganan dengan baik (Gallagher, tanpa tahun). Penyakit tungro disebabkan oleh Virus Tungro Padi (VTP) dan ditularkan oleh hama wereng hijau. Hama wereng kini meresahkan petani di Kabupaten Bondowoso. Beberapa areal persawahan telah dirusak oleh hama tersebut (Anonim**,** 2009).

Akhir-akhir ini upaya pemberantasan penyakit tungro dilakukan melalui pemberantasan penyebab tungro (wereng hijau) dan dilanjutkan dengan melakukan pemulihan tanaman padi yang diduga terserang penyakit tungro dengan cara penyemprotan pestisida dan pemberian vitamin padi pada tanaman terserang. Dalam hal ini pemberantasan tungro selain dengan pemulihan langsung juga sering dilakukan dengan jalan pemilihan waktu musim tanam dan penggunaan varietas tanam padi yang tepat untuk menekan penularan virus tungro oleh wereng hijau.

Keterkaitan faktor-faktor penyebab serangan penyakit tungro dengan banyaknya wereng hijau penular dapat didekati dengan analisis statistika mengenai hubungan variabel takbebas dan variabel bebas, yaitu analisis regresi.

Analisis regresi umumnya menggunakan variabel takbebas yang merupakan peubah acak kontinu berdistribusi normal. Namun ada juga variabel takbebas yang diamati merupakan peubah acak diskrit yang berdistribusi Poisson. Apabila terdapat variabel takbebas yang akan diamati merupakan peubah acak diskrit yang berdistribusi Poisson, maka hubungan antara variabel takbebas dan variabel bebas dapat diketahui dengan analisis regresi Poisson (Myers, 1990).

Distribusi Poisson sering digunakan dalam pemodelan kasus yang jarang terjadi (*rare event*) (Lam *et al*, 2006). Populasi wereng hijau yang didapatkan dalam setiap pengamatan di daerah terserang penyakit tungro relatif jarang sehingga menghasilkan banyak nilai nol dalam beberapa pengamatan. Hal ini dapat diketahui dari sekitar 33.000 Ha luas lahan sawah yang tersebar di 250 desa di Kabupaten Bondowoso terdapat rata-rata 120 wereng hijau setiap bulannya. Meskipun demikian populasi wereng hijau tidak dapat dianggap remeh karena dapat menularkan virus tungro dengan cepat dan dapat merusak tanaman padi di areal persawahan sampai habis. Nilai nol dalam setiap pengamatan didapatkan dari dua keadaan. Keadaan yang pertama yaitu, populasi wereng hijau bernilai nol karena tidak dijumpai wereng hijau pada saat pengamatan. Sedangkan keadaan yang kedua yaitu, populasi wereng hijau bernilai nol karena tidak ada wereng hijau di areal persawahan padi. Populasi wereng hijau tersebut selaku variabel takbebas dapat diasumsikan mengikuti distribusi Poisson. Hubungan antara populasi wereng hijau dengan faktor-faktor yang mempengaruhinya di Kabupaten Bondowoso dapat dicari dengan menggunakan analisis regresi Poisson.

1.2 Rumusan Masalah

Dari latar belakang di atas, nilai nol yang diperoleh dari dua keadaan tersebut dapat dianalisis dengan regresi Poisson. Dalam regresi Poisson terdapat dua perlakuan terhadap nilai nol. Yang pertama dengan mengabaikan nilai nol sehingga hanya menganalisis banyaknya populasi wereng hijau bernilai bilangan bulat positif yang disebut model regresi *Zero-truncated* Poisson. Yang kedua, yaitu dengan mempertimbangan nilai nol. Nilai nol yang diperoleh dari dua keadaan seperti pada penjelasan diatas dapat dianalisis dengan menggunakan model regresi *Zero-inflated* Poisson.

Setelah mengetahui beberapa uraian diatas dapat dirumuskan suatu permasalahan yang akan dibahas dalam skripsi ini yaitu bagaimana pemodelan terbaik terhadap banyaknya populasi wereng hijau penyebab penyakit tungro yang banyak mengandung nilai nol dengan beberapa variabel bebas yang mempengaruhi menggunakan *Zero-Inflated* Poisson.

1.3 Tujuan

Adapun tujuan dari penyusunan skripsi ini yaitu untuk mendapatkan model terbaik terhadap banyaknya populasi wereng hijau penyebab penyakit tungro yang banyak mengandung nilai nol dengan beberapa variabel bebas yang mempengaruhi menggunakan *Zero-inflated* Poisson.

1.4 Manfaat

Manfaat dari penyusunan skripsi ini selain untuk mengetahui model terbaik terhadap populasi wereng hijau penyebab penyakit tungro, juga untuk mengetahui faktor-faktor apa saja yang paling mempengarui persebaran populasi wereng hijau sehingga dapat dilakukan penanganan lebih awal terhadap hama tersebut.

BAB. 2 TINJAUAN PUSTAKA

2.1 Penyakit Tungro

Penyakit tungro telah disebut-sebut sejak pertengahan abad XIX dan peranannya sebagai faktor pembatas hasil tanaman padi pada waktu itu masih belum banyak dibicarakan (Departemen Pertanian, 1985). Tetapi dengan bertambahnya kemajuan teknologi dibidang pertanian memungkinkan terjadinya perubahan lingkungan, sehingga penangananannya mulai diperbincangkan karena kehilangan hasil panen akibat serangan penyakit ini sangat memprihatinkan.

Penyakit tungro disebabkan oleh Virus Tungro Padi (VTP) dan ditularkan oleh wereng hijau, tetapi peranan wereng hijau tersebut secara langsung tidak menimbulkan kerusakan yang berarti pada tanaman (Departemen Pertanian, 1985). Penularan penyakit tungro dilakukan oleh serangga-serangga penular dengan jalan menghisap cairan tanaman sakit yang mengandung virus tungro dalam waktu minimal 5 menit, dan segera memindahkan ke tanaman sehat dengan cara dan waktu yang sama. Setelah 7 – 10 hari kemudian tanaman tersebut menunjukkan gejala serangan penyakit tugro (Gallagher, tanpa tahun).

2.1.1 Faktor-faktor Yang Mendukung Panyebaran Penyakit Tungro

Menurut Departemen Pertanian (1985) penyebaran penyakit tungro sangat dipengaruhi oleh serangga penular, sumber penyakit, tanaman inang, dan lingkungan.

a. Serangga Penular

Penyebaran penyakit tungro pada tanaman padi dapat terbawa dan ditularkan oleh serangga. Serangga tersebut di Indonesia adalah wereng hijau (*Nephotettix sp.*) dan wereng loreng (*Recilia Dorsalis*), yang biasanya terdapat dan banyak dijumpai pada daun padi. *Nephottetix sp*. dikenal sebagai wereng hijau, karena warnanya hijau dan menyerang bagian daun tanaman padi. *Recilia Dorsalis*, serangga ini lebih dikenal dengan nama wereng loreng, karena serangga yang dewasa bersayap putih dengan pita berwarna coklat muda berbentuk huruf W (Departemen Pertanian, 1985).

b. Sumber Penyebab Penyakit

Penyakit tungro disebabkan oleh virus yang disebut dengan Virus Tungro Padi (VTP), virus ini mempunyai sifat *non persinten*, artinya virus tersebut hanya dapat menyerang tanaman dalam masa yang pendek saya.

c. Tanaman Inang dan Penularannya

Penyakit tungro dan serangan penular tidak hanya hidup pada tanaman padi saja, tetapi dapat juga berkembang pada tanaman gulma seperti rerumputan, hanya saja penularan tidak terbawa atau melalui benih, tanah maupun secara mekanik. Rerumputan yang dapat digunakan sebagai tanamaan inang selain padi adalah *Echinochloa Colonum* (Tuton), *Echinochloa Crusgali* (Jawan), dan *Paspalum Vaginatum*.

d. Lingkungan

Populasi serangga penular (wereng hijau) dan serangan penyakit tungro bervariasi dari bulan kebulan dan dari musim kemusim. Hal ini dapat dipengaruhi oleh keadaan lingkungan, seperti curah hujan dan jamur parasit (*Entomophora sp*. dan *Metarrhizium sp.*) yang dapat menyerang serangga. Curah hujan yang tinggi berpengaruh langsung terhadap aktifitas serangga, sedangkan yang tidak langsung menyebabkan kelembaban udara meningkat sehingga dapat memacu perkembangan jamur untuk tumbuh lebih baik pada tubuh serangga, akibatnya perkembangan serangga terhambat dan populasi serangga menurun.

2.1.2 Penilaian Kerusakan Organisme Pengganggu Tanaman

Penilaian terhadap kerusakan tanaman padi akibat serangan penyakit tungro menurut Direktorat Jenderal Tanaman Pangan (2007) dilakukan melalui dua cara, yaitu:

a. Perhitungan intensitas serangan penyakit tungro

Untuk melakukan penilaian terhadap intensitas serangan penyakit tungro digunakan rumus:

$$I = \frac{a}{a+b} \times 100\%$$

dimana

I = intensitas serangan (%)

a = banyaknya contoh tanaman yang dianggap rusak atau terserang penyakit

b = banyaknya contoh tanaman yang tidak rusak (tidak menunjukkan gejala serangan penyakit)

b. Perhitungan populasi wereng hijau penular penyakit tungro

Perhitungan populasi wereng hijau penular penyakit tungro adalah sebagai berikut:

$$\text{Pop} = 10 \times \text{ayunan tunggal jaring} \qquad (2.1)$$

dimana ayunan jaring yang dilakukan mengikuti arah diagonal pada petak sawah tanaman padi yang terserang penyakit tungro sedangkan banyaknya populasi wereng hijau yang didapatkan diketahui dengan cara menghitung banyaknya wereng hijau yang tertangkap jaring.

2.1.3 Cara Mengatasi Penyakit Tungro

Menurut Balai Proteksi Tanaman Pangan Wilayah VI Surabaya (1985) terdapat beberapa cara mengatasi Penyakit tungro, antara lain:

a. Sanitasi (Pemusnahan Sumber Penyakit)

Untuk menghilangkan sumber penyakit, tanaman sakit yang berumur kurang dari 2 bulan, sisa-sisa tanaman sakit dan rerumputan dimusnahkan. Pada tanaman yang terserang pada waktu sudah berbunga sanitasinya dilakukan secara selektif, kecuali tanaman yang mengalami serangan total. Cara pemusnaannya adalah dengan cara dibenamkan seluruh bagian tanaman kedalam lumpur, kemudian diikuti dengan pengolahan tanah dan dibiarkan dalam keadaan terolah sampai dengan saat mulai bertanam secara bersamaan.

b. Penerapan Pola Tanam dan Pergiliran Varietas Padi

Didalam pola penanaman padi setahun dikenal pertanaman padi sekali, dua kali, bahkan dilakukan secara terus menerus di lahan yang beririgasi baik. Penanaman secara terus-menerus tersebut akan mendorong serangga penular maupun virus tungro menyesuaikan diri dengan lingkungannya, sehingga mampu berkembang dan berhasil menghancurkan tanaman yang semula dikatakan tahan. Disamping itu juga penanaman secara tidak serempak, memungkinkan tersedianya makanan dan tempat berlindung bagi serangga penular maupun virus tungro sepanjang tahun. Untuk mengatasi masalah tersebut perlu dilaksanakan:

1) Penerapan Pola Tanam

Penerapan pola tanam dilaksanakan dengan pergiliran tanaman bukan padi. Penanaman padi hanya dilakukan paling banyak dua kali dalam setahun (misalnya dengan pola tanam padi – padi – palawija, padi – palawija – padi, atau palawija – padi – padi). Disamping itu diusahakan juga pada waktu pengolahan lahan minimal ada jarak satu bulan diantara dua masa tanam yang berurutan, dengan harapan dapat memutuskan daur hidup atau rantai makanan serangga penular.

2) Pergiliran Varietas Padi

Dalam hal mengatasi penyakit tungro di daerah yang menanam padi dua kali setahun, disarankan adanya pergiliran varietas padi yang ada, dimulai dari musin hujan kemusim kemarau. Dalam hal ini tidak menanam satu golongan varietas yang sama secara terus-menerus selama periode dan tempat tertentu.

c. Penggunaan Insektisida

Walaupun penanaman padi telah menggunakan varietas tahan penyakit, tetapi sifat ketahanan suatu varietas tidaklah mutlak. Oleh karena itu, walaupun sudah menanam varietas tersebut dalam menghadapi serangan penyakit tungro masih memerlukan cara pengendalian lain seperti insektisida, terutama pada daerah serangan tungro. Hal ini bertujuan menekan perkembangan wareng hijau selaku serangga penular.

2.2 Analisis Regresi

Salah satu metode peramalan yang sering digunaan saat ini adalah analisis regresi. Analisis regresi adalah metode statistika yang popular digunakan untuk menyatakan hubungan antara variabel takbebas dan variabel bebas. Variabel takbebas dinyatakan dengan Y sedangakan variabel bebas dinyatakan dengan $X_1, X_2, \ldots, X_k$ (Kleinbaum *et al*. 1998).

Analisis regresi menentukan hubungan fungsional yang berlaku untuk populasi berdasarkan data sample yang diambil dari populasi yang bersangkutan. Hubungan fungsional ini dituliskan dalam bentuk persamaan matematis sebagai berikut:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ \ldots + \ \beta_r X_r + \epsilon \,, \quad r = 1,2, \ldots, n$$

dengan Y = variabel takbebas

X_r= variabel bebas ke-r

β_r = parameter regresi (nilai yang akan ditaksir) ke-r

ϵ = galat atau kesalahan

2.3 Distribusi Poisson

Andaikan kegiatan menghitung banyaknya objek atau peristiwa dilakukan terus menerus dalam kurun waktu yang sangat lama. Jika $X(t)$ menyatakan banyaknya peristiwa sampai waktu t, maka barisan variable acak $\{X(t), t > 0\}$ disebut proses menghitung (proses mencacah). Proses mencacah mempunyai beberapa jenis, diantaranya proses Poisson. Indeks pada proses Poisson tidak harus berupa waktu, tetapi bisa juga berupa lokasi ataupun banyaknya suatu populasi (Nurhayati, 2012).

Variabel bebas X dikatakan mempunyai distribusi Poisson jika fungsi kepadatan peluangnya berbentuk:

$$P(Y_i = y_i) = \begin{cases} \frac{e^{-\lambda}\lambda^{y}}{y_i!}, & \text{untuk } y = 0, 1, 2, 3, \ldots \\ 0, & \text{untuk yang lainnya} \end{cases}$$

dimana $y = 0, 1, 2, \ldots,$ sedangkan e merupakan sebuah bilangan konstan yang jika dihitung dengan pembulatan empat desimal sama dengan 2,7183 yang diperoleh dari pendekatan $\lim_{x \to \sim} \left[1 + \frac{1}{x}\right]^{x}$, x merupakan bilangan asli.

Distribusi Poisson dapat diperoleh berdasarkan pendekatan distribusi binomial, dengan $n\theta = \lambda$ dengan θ sangat kecil $(\theta \to 0)$ dan n besar $(n \to \infty)$ (Nasoetion dan Rambe, 1983).

Rata-rata dan varians variable bebas X yang mengikuti distribusi Poisson dengan parameter λ masing-masing adalah sebagai berikut:

$$E(X) = \lambda$$

$$Var\ (X) = \lambda$$

distribusi Poisson sering digunakan untuk menentukan peluang sebuah peristiwa atau kejadian yang dalam keadaan tertentu diharapkan terjadinya sangat jarang (Kleinbaum *et al*, 1998).

2.4 Model Regresi Poisson

Regesi Poisson merupakan suatu bentuk analisis regresi yang digunakan untuk memodelkan data yang berbentuk jumlah, misalkan data tersebut dilambangkan dengan Y, yaitu banyaknya kejadian yang terjadi dalam suatu periode waktu dan atau wilayah tertentu. Model regesi Poisson merupakan model standar untuk data diskrit dan termasuk dalam model regresi linear (Cameron & Trivedi, 1998). Poisson adalah suatu bentuk model linier umum dimana variable takbebas dimodelkan sebagai distribusi Poisson.

Model regresi Poisson ditulis sebagai berikut (Myers, 1990);

$$y_i = \mu_i + \varepsilon_i\ , \quad i = 1,\ 2,\ \ldots\ ,\ n$$

dimana y_i adalah jumlah kejadian dan μ_i adalah rata-rata jumlah kejadian yang berdistribusi Poisson . μ_i bergantung pada unit tertentu atau periode tertentu, jarak, luas, volume, dan lain-lain. μ_i diasumsikan tidak berubah dari data ke data. Model-model regresi Poisson merupakan *Generalized Linear Model* (GLM) dengan data responnya (komponen random) diasumsikan berdistribusi Poisson (McCullagh dan Nelder, 1989). Dalam GLM terdapat sebuah fungsi g yang linier dan menghubungkan mean dari variable takbebas dengan sebuah variable bebas, yaitu:

$$g(\mu_i) = \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik}$$

$$= \sum_{j=1}^{k} \beta_j x_{ij}, \qquad i = 1,2,\ldots,n$$

fungsi g disebut fungsi penghubung (*link function*).

Pada model regresi Poisson, biasanya *link function* yang digunakan adalah *log*, sehingga $\log \mu_i = \eta_i$. Dengan demikian model regresi Poisson dapat ditulis sebagai berikut:

$$\log \mu_i = \sum_{j=1}^{k} \beta_j x_{ij}, \qquad i = 1,2,\ldots,n$$

dimana $\mu_i = \mu_i(x_i) = \exp(\sum_{j=1}^{k} \beta_j x_{ij})$.

sehingga persamaan distribusi Poisson dinyatakan sebagai berikut:

$$P(y_{i;}\widehat{\boldsymbol{\beta}}) = \frac{e^{-t_i[\lambda(X_i;\widehat{\boldsymbol{\beta}})]} t_i[\lambda(X_i;\widehat{\boldsymbol{\beta}})]^{y_i}}{y_i!} \qquad (2.2)$$

dimana

$$\mu_i = t_i \mu(X_i;\widehat{\boldsymbol{\beta}}) = t_i \exp(\boldsymbol{x}_i^T \widehat{\boldsymbol{\beta}})$$

dan

$$var\ (y_i) = t_i \mu(X_i;\widehat{\boldsymbol{\beta}}) = t_i \exp(\boldsymbol{x}_i^T \widehat{\boldsymbol{\beta}})$$

dimana y_i adalah jumlah kejadian dan μ_i adalah rata-rata jumlah kejadian dalam periode t_i. μ_i diasumsikan tidak berubah dari data ke data. $\mu(X_i;\widehat{\boldsymbol{\beta}})$ adalah rata-rata Poisson dan vektor $\widehat{\boldsymbol{\beta}}$ menunjukkan parameter yang ditaksir. selanjutnya model regresi Poisson dapat dituliskan sebagai berikut (Myers, 1990)

$$y_i = t_i \exp(\boldsymbol{x}_i^T \widehat{\boldsymbol{\beta}}) + \varepsilon_i$$

Permasalahan yang terjadi pada regresi Poisson adalah jika terdapat banyak data yang bermilai nol, sehingga lebih banyak data nol-nya dibandingkan regresi Poisson yang akan diprediksi. Hal tersebut akan menyebabkan regresi Poisson menjadi tidak tepat menggambarkan data yang sebenarnya.

2.5 Distribusi *Zero-Inflated* Poisson (ZIP)

Lambert (1992) menjelaskan bahwa *Zero-inflated* Poisson adalah model campuran yang sederhana untuk data diskrit dengan banyak peristiwa nol. Jika Y_i adalah variabel acak takbebas yang mempuyai distribusi *Zero-inflated* Poisson, maka penelitian nol dapat dikembangkan dalam dua langkah, yaitu:

$$Y_i \sim \begin{cases} 0\,, & \text{dengan Peluang [illegible]} \\ Poisson\,(\mu_i), & \text{dengan Peluang } (1 - \text{[illegible]} \end{cases} \qquad (2.3)$$

mean dan variannya diberikan sebagai berikut:

$$\text{mean} = E(Y_i) = \mu_i$$

dan

$$\text{var}\,(Y_i) = \mu_i + \frac{w}{(1-w)}\,\mu_i{}^2$$

dari mean dan variannya dapat dilihat bahwa distribusi dari Y_i menunjukkan fenomena overdispersi jika $w > 0$, karena varian > mean.

2.6 Model Regresi Zero-*Inflated* Poisson (ZIP)

Jika data yang bernilai nol atau kosong dijumpai pada data jenis cacah dan proporsinya besar (*zero inflation*), maka disarankan menggunakan model regresi *Zero-inflated* Poisson (ZIP) (Lambert, 1992). Pada masa-masa sekarang, model ZIP banyak diterapkan dalam berbagai bidang. Untuk peramalan, model ZIP yang diperkenalkan oleh Lambert lebih tepat dibandingkan dengan regresi Poisson ketika data mengadung lebih banyak kejadian nol. Sesuai persamaan (2.3) maka Lambert (1992) mendefinisikan model regresi *Zero-inflated* Poisson sebagai

$$P(Y_i = y_i) = \begin{cases} w + (1-w)e^{-\lambda} & \text{untuk } y_i = 0 \\ (1-w)\dfrac{e^{-\lambda}\lambda^{y_i}}{y_i!}, & \text{untuk } y_i = 1,2,\ldots \end{cases}$$

yang dinotasikan dengan $Y_i \sim ZIP\ (\lambda, w)$.

Berdasarkan Lambert (1992) untuk memodelkan w sudah umum digunakan model logit, yaitu:

$$w = \frac{\exp(\boldsymbol{z}_i^T \gamma)}{1 + \exp(\boldsymbol{z}_i^T \gamma)}$$

dimana $\boldsymbol{z}_i$ adalah sebuah vektor $(1 \times p)$ dari kovarian peneitian ke-i, dan γ adalah vektor $(p \times 1)$ dari parameter tambahan. Elemen $\boldsymbol{z}_i$ mungkin saja mengandung elemen x_i.

Untuk mengaplikasikan model *Zero-inflated* Poisson dalam model yang lebih praktis, Lambert (1992) menyarankan hubungan model untuk λ dan w adalah sebagai berikut:

$$\log(\boldsymbol{\lambda}) = \boldsymbol{X\beta} \text{ dan } logit\ (w) = \log\left(\frac{w}{1-w}\right) = \boldsymbol{X_\gamma}$$

dimana $\boldsymbol{X}$ matriks kovarian sedangkan $\boldsymbol{\beta}$ dan $\boldsymbol{\gamma}$ adalah matriks berukuran $(p + 1) \times 1$ dan $(q + 1) \times 1$ dari parameter yang akan ditaksir.

2.7 Penaksiran Parameter Analisis Regresi Poisson dan Zero-Inflated Poisson

Metode MLE (*Maximum Likelihood Estimator*) adalah salah satu metode penaksiran parameter yang dapat digunakan untuk menaksir parameter suatu model yang diketahui distribusinya. Sebagaimana diketahui bahwa taksiran parameter melalui metode MLE adalah melakukan turunan parsial fungsi kemungkinan terhadap parameter yang ditaksir. Berdasarkan persamaan (2.2), maka fungsi kemungkinan Analisis regresi Poisson adalah sebagai berikut (Kleinbaum *et al*. 1998):

$$L(y;\widehat{\boldsymbol{\beta}}) = \prod_{i=1}^{n} P(y_{i;}\widehat{\boldsymbol{\beta}})$$

$$= \prod_{i=1}^{n} \frac{t_i\left[\lambda(X_i;\widehat{\boldsymbol{\beta}})\right]^{y_i} e^{-t_i[\lambda(X_i;\widehat{\boldsymbol{\beta}})]}}{y_i!}$$

$$= \frac{\left\{\prod_{i=1}^{n}[t_i[\lambda(X_i;\widehat{\boldsymbol{\beta}})]^{y_i} e^{-\sum_{i=1}^{n} t_i[\lambda(X_i;\widehat{\boldsymbol{\beta}})]}]\right\}}{\prod_{i=1}^{n} y_i!}$$

persamaan di atas akan dimaksimumkan dengan menggunakan teknik *iterative* yang menghasilkan penaksir kemungkinan *likelihood* untuk koefisien regresi dalam $\widehat{\boldsymbol{\beta}}$. Myers (1990) menyarankan menggunakan prosedur pendekatan *Iteratively Reweighted Least Square* (IRWLS) untuk menentukan penaksir kemungkinan maksimum. *Iteratively Reweighted Least Square* (IRWLS) menggunakan *Newton-Raphson*, biasanya pada iterasi ke-s, metode *Newton-Raphson* memperbaiki taksiran $\hat{\beta}_s$ yang biasa dipakai dengan rumus sebagai berikut:

$$\widehat{\boldsymbol{\beta}}_{s+1} = \widehat{\boldsymbol{\beta}}_s - \widehat{H}_s\widehat{\boldsymbol{g}}_s$$

dimana $\boldsymbol{g} = \frac{\partial \log L((y;\widehat{\boldsymbol{\beta}})}{\partial\widehat{\boldsymbol{\beta}}.}$ dan $H = \frac{\partial^2 log\widehat{\boldsymbol{\beta}}}{\partial(\widehat{\boldsymbol{\beta}})^2}$

metode *Newton-Raphson* digunakan untuk menyelesaikan persamaan berikut:

$$\frac{\partial \log L((y;\widehat{\boldsymbol{\beta}})}{\partial\widehat{\boldsymbol{\beta}}.} = 0$$

dimana

$$\log L(y;\widehat{\boldsymbol{\beta}}) = \sum_{i=1}^{n} y_i \log t_i\lambda(x_I;\widehat{\boldsymbol{\beta}}) - \sum_{i=1}^{n} t_i\lambda(x_i;\widehat{\boldsymbol{\beta}}) - \sum_{i=1}^{n} \log(y_i)!$$

persamaan *likelihood* untuk mencari $\widehat{\boldsymbol{\beta}}$ adalah sebagai berikut:

$$\sum_{i=1}^{n}\left[\frac{y_i}{\lambda(x_i;\widehat{\boldsymbol{\beta}})}\right]\left[\frac{\partial\lambda(x_i;\widehat{\boldsymbol{\beta}})}{\partial\widehat{\boldsymbol{\beta}}}\right] - \sum_{i=1}^{n}[t_i]\left[\frac{\partial\lambda(x_i;\widehat{\boldsymbol{\beta}})}{\partial\widehat{\boldsymbol{\beta}}}\right] = 0$$

$$\sum_{i=1}^{n}\left[\frac{y_i}{\lambda(x_i;\widehat{\boldsymbol{\beta}})} - t_i\right]\left[\frac{\partial\lambda(x_i;\widehat{\boldsymbol{\beta}})}{\partial\widehat{\boldsymbol{\beta}}}\right] = 0$$

Sedangkan fungsi kemungkinan ZIP adalah sebagai berikut (Lambert,1992):

$$L(\beta,\gamma|y_i,x_i) = \begin{cases} \prod_{i=1}^{n} \frac{\exp(\boldsymbol{x}_i^T\gamma) + \exp(-\exp(\boldsymbol{x}_i^T\beta))}{1+\exp(\boldsymbol{x}_i^T\gamma)}, & y_i = 0 \\ \prod_{i=1}^{n} \frac{\frac{1}{1+\exp(\boldsymbol{x}_i^T\gamma)} (\exp\left((1+\exp(\boldsymbol{x}_i^T\beta) + (\boldsymbol{x}_i^T\beta)y_i)\right))}{y_i!} & \end{cases}$$

dimana x_i adalah variable bebas, dan y_i adalah variabel takbebas, serta β dan γ adalah parameter yang akan ditaksir.

Log fungsi *likelihood* dari model regresi ZIP diberikan oleh:

$$\begin{aligned} \log L(\beta,\gamma|y_i,x_i) \\ &= \sum_{i=1}^{n} \log\left(\exp(\boldsymbol{x}_i^T\gamma) + \exp(-\exp(\boldsymbol{x}_i^T\beta))\right) \\ &- \sum_{i=1}^{n} \log(1+\exp(\boldsymbol{x}_i^T\gamma)) + \sum_{i=1}^{n}((y_i\boldsymbol{x}_i^T\beta) \\ &- \exp(\boldsymbol{x}_i^T\beta)) - \sum_{i=1}^{n} \log(y_i)! \end{aligned}$$

2.8 Pengujian Kesesuaian Analisis Regresi Poisson

Pengujian kesesuaian model dengan menggunakan *goodness of fit* disebut devians (Kleinbaun *et al*, 1988). Perumusan hipotesisi pengujian kesesuaian model regresi poisson adalah sebagai berikut.

$$H_0: \lambda_i = t_i\lambda(x_i;\widehat{\boldsymbol{\beta}}), \ \ i = 1,2,3,\dots,n$$

$$H_0: \lambda_i \neq t_i\lambda(x_i;\widehat{\boldsymbol{\beta}})$$

statistik uji yang digunakan sebagai berikut:

$$g = -2\log\left[\frac{L(y;\widehat{\boldsymbol{\beta}})}{L(y;\lambda)}\right]$$

Nilai $\lambda_i = t_i\lambda(x_i;\widehat{\boldsymbol{\beta}})$, sehingga persamaan (2.2) dapat dituliskan menjadi persamaan berikut:

$$\log L(y;\widehat{\boldsymbol{\beta}}) = \sum_{i=1}^{n} y_i \log \hat{y}_i - \sum_{i=1}^{n} \hat{y}_i - \sum_{i=1}^{n} \log(y_i)!$$

$$\log L(y;\widehat{\boldsymbol{\beta}}) = \sum_{i=1}^{n} y_i \log \lambda_i - \sum_{i=1}^{n} \lambda_i - \sum_{i=1}^{n} \log(y_i)!$$

Sedangkan nilai $L(y;\lambda)$ dapat ditulis dalam persamaan berikut:

$$L(y;\lambda) = \prod_{i=1}^{n} \frac{\lambda_i^{y} e^{-\lambda i}}{y_i!} = \frac{\prod_{i=1}^{n} \lambda i^{y_i} \exp(-\sum_{i=1}^{n} \lambda_i)}{\prod_{i=1}^{n} y_i!} \quad (2.4)$$

nilai $\hat{\lambda}_i = y_i$ sehingga persamaan (2.4) dapat ditulis sebagai berikut

$$L(y,\hat{\lambda}) = \frac{(\prod_{i=1}^{n} y_i^{y_i}) \exp(-\sum_{i-1}^{n} y_i)}{\prod_{i=1}^{n} y_i!}$$

nilai $\log L(y,\hat{\lambda})$ ditulis dalam persamaan berikut:

$$\begin{aligned} G &= -2\log[L(y;\hat{\beta}) - L(y,\hat{\lambda})] \\ &= 2\log[L(y,\hat{\lambda}) - L(y;\hat{\beta})] \\ &= 2\sum_{i=1}^{n}[(y_i \log y_i - y_i - \log((y_i)!) - (y_i \log \hat{y}_i - \hat{y}_i - \log(y_i)!)] \\ &= 2\sum_{i=1}^{n}\left[y_i \log\left(\frac{y_i}{\hat{y}_i}\right) - (y_i - \hat{y}_i)\right] \end{aligned}$$

nilai G tersebut disebut *devians* untuk model regresi Poisson. Menurut Ismail dan Jermain (2005) untuk model yang sesuai, *devians* mendekati distribusi *Chi-Kuadrat* dengan derajat kebebasan $= (n-k-1)$, dimana n adalah banyak pengamatan dan $k+1$ adalah banyak parameter. Kriteria untuk pengujian ini adalah tolak H_0 pada taraf signifikansi α, jika $G > \mathcal{X}^2_{(n-k-1)\alpha}$.

Kleinbaum *et all* (1988) menyatakan bahwa *devians* seperti *sum Square Error* pada analisis regresi linier berganda. Apabila nilai data pengamatan sama dengan prediksi ($y_i = \hat{y}_i$), maka nilai $G = 0$. Semakin besar selisih antara respon pengamatan dan respon taksiran, maka semakin besar pula nilai *devians*. Taksiran respon diharapkan mendekati pengamatan atau tingkat kesalahan kecil sehingga nilai *devians* kecil sesuai dengan yang diharapkan.

Parameter model regresi Poisson yang telah dihasilkan dari taksiran parameter belum tentu mempunyai pengaruh yang signifikan terhadap model. Untuk itu perlu dilakukan pengujian terhadap parameter model regresi Poisson secara individu. Perumusan hipotesis yang dilakukan adalah sebagai berikut:

$H_o: \beta_r = 0,\ 0 < r < k$ (pengaruh variable ke-r tidak signifikan)

$H_1: \beta_r \neq 0$ (Pengaruh variable ke-r signifikan)

dimana $k+1$ adalah banyak parameter. Statistik uji yang digunakan adalah sebagai berikut:

$$t_i = \frac{\hat{\boldsymbol{\beta}}_i}{SE(\hat{\boldsymbol{\beta}}_i)}$$

kriteria pengujian adalah H_o ditolak jika $|t_{hit}| > t_{\alpha/2,v}$ dimana α adalah tingkat signifikansi dan v adalah derajat kebebasan.

2.9 Pengujian Kesesuaian Analisis Regresi Zero-Inflated Poisson

Pengujian kesesuaian model regresi ZIP dilakukan menggunakan LR (*Likelihood Ratio*) test. Hipotesis untuk pengujian kesesuaian model adalah sebagai berikut:

H_0: $\beta_1 = \beta_2 = \cdots = \beta_k = \gamma_1 = \gamma_1 = \cdots = \gamma_k = 0$

H_0: paling sedikit ada satu $\beta_r \neq 0$ atau $\gamma_r \neq 0$, $0 < r < k$

dimana $k+1$ adalah jumlah parameter, β_r adalah parameter model log ke-r, dan γ_r adalah parameter model logit ke-r. Hall dan Shen (2009) telah melakukan perhitungan statistik uji untuk pengujian kesesuaian model sebagai berikut:

$$G = -2\log\left[\frac{L(y;\hat{w})}{L(y;\hat{\Omega})}\right] = (2\sum_{i=1}^{n} z_i x_i^T\hat{\gamma} - \log(1+\exp(x_i^T\hat{\gamma})) + 2\sum_{i=1}^{n}(1 - z_i)(y_i x_i^T\hat{\beta} - \exp(x_i^T\hat{\beta})) - (2\sum_{i=1}^{n} z_i\hat{y}_0 - \log(1+x_i^T\hat{y}_0) + 2\sum_{i=1}^{n}(1-z_i)(y_i\hat{\beta}_0 - \exp(\hat{\beta}_0)))$$

Sedangakan pengujian parameter secara individu ada dua, yaitu pengujian parameter model log dan pengujian parameter logit. Berikt ini adalah perumusan hipotesis untuk pengujian parameter model log:

$H_o: \beta_r = 0,\ 0 < r < k$

$H_1: \beta_r \neq 0$

dimana $k + 1$ adalah banyak parameter.

Statistik uji untuk pegujian parameter model log secara individu adalah sebagai berikut (Hall & Shen, 2009):

$$G = -2\log\left[\frac{L(y;\widehat{w})}{L(y;\widehat{\Omega})}\right]$$

$$= \left(2\sum_{i=1}^{n} z_i \boldsymbol{x}_i^T \hat{\gamma} - \log(1 + \exp(\boldsymbol{x}_i^T \hat{\gamma})\right) + 2\sum_{i=1}^{n}(1 - z_i)(y_i \boldsymbol{x}_i^T \widehat{\boldsymbol{\beta}}$$

$$- \exp(\boldsymbol{x}_i^T \hat{\beta})) - 2\sum_{i=1}^{n}(1 - z_i)(y_i \boldsymbol{x}_i^T \hat{\beta}_i)$$

Perumusan hipotesis untuk pengujian parameter model logit secara individu adalah sebagai berikut:

$H_o: \gamma_r = 0,\ 0 < r < k$

$H_1: \gamma_r \neq 0$

dimana $k + 1$ adalah banyak parameter.

Statistik uji yang digunakan untuk pengujian parameter model logit adalah sebagai berikut (Hall & Shen, 2009):

$$G = -2\log\left[\frac{L(y;\widehat{w})}{L(y;\widehat{\Omega})}\right]$$

$$= \left(2\sum_{i=1}^{n} z_i x_i^T \hat{\gamma} - \log(1 + \exp(x_i^T \hat{\gamma})\right) + 2\sum_{i=1}^{n}(1 - z_i)(y_i x_i^T \hat{\beta}$$

$$- \exp(x_i^T \hat{\beta}))$$

$$- 2\sum_{i=1}^{n}(1 - z_i)\log\ (y_i)! - 2\sum_{i=1}^{n} z_i \hat{y}_0$$

$$- \log(1 + x_i^T \hat{y}_0) + \exp \hat{\gamma}_0$$

Kriteria pengujian untuk ketiga pengujian diatas adalah H_o ditolak pada taraf signifikan α , jika $G_{hitung} > \mathcal{X}^2_{(v,\alpha)}$, dimana v adalah derajat bebas. Untuk *goodness of fit* dari suatu model regresi yang diperoleh dapat digunakan nilai log-*likelihoood* dimana model dengan nilai log-*likelihood* yang lebih besar menunjukkan model yang lebih baik (Femoye *et all*, 2004).

2.10 Overdispersi

Data cacah untuk regresi Poisson dikatakan mengandung overdispersi apabila nilai variansnya lebih besar dari nilai meannya. apabila pada data cacah terjadi overdispersi dan tetap menggunakan regresi Poisson, maka dugaan dari parameter koefisien regresinya tetap konsisten namun tidak efisien. Hal ini berdampak pada nilai standar error yang menjadi *under estimate*, sehingga kesimpulan yang dihasilkan tidak valid. Fenomena overdispersi oleh McCullagh & Nelder (1989) dinyatakan dengan:

$$Var\,(Y)\ >\ E\,(Y)$$

Overdispersi dapat diindikasikan dengan nilai *pearson Chi-square* dan *residual deviance* yang dibagi dengan derajat bebasnya. Jika kedua nilai ini lebih besar dari satu maka dikatakan terjadi overdispersi pada data. Overdispersi dapat pula terjadi karena adanya pengamatan *missing* pada variable bebas, adanya pencilan pada data, variable bebas perlu ditransformasi atau kesalahan spesifikasi *link function* (Hardin dan Hilbe, 2007).

BAB. 3 METODE PENELITIAN

3.1 Data

Data yang digunakan dalam penelitian ini adalah data sekunder. Data tersebut antara lain: data populasi wereng hijau penular penyakit tungro, luas lahan sawah di Kabupaten Bondowoso, umur tanaman padi terserang penyakit tungro, adanya tanaman inang atau gulma, dan data curah hujan perbulan di seluruh kecamatan di Kabupaten Bondowoso tahun 2012 yang diperoleh dari Dinas Pertanian Tanaman Pangan dan Hortikultura Kabupaten Bondowoso. Data-data tersebut merupakan data-data perdesa diseluruh kecamatan di Bondowoso. Pada kasus ini variabel-variabel yang digunakan adalah sebagai berikut.

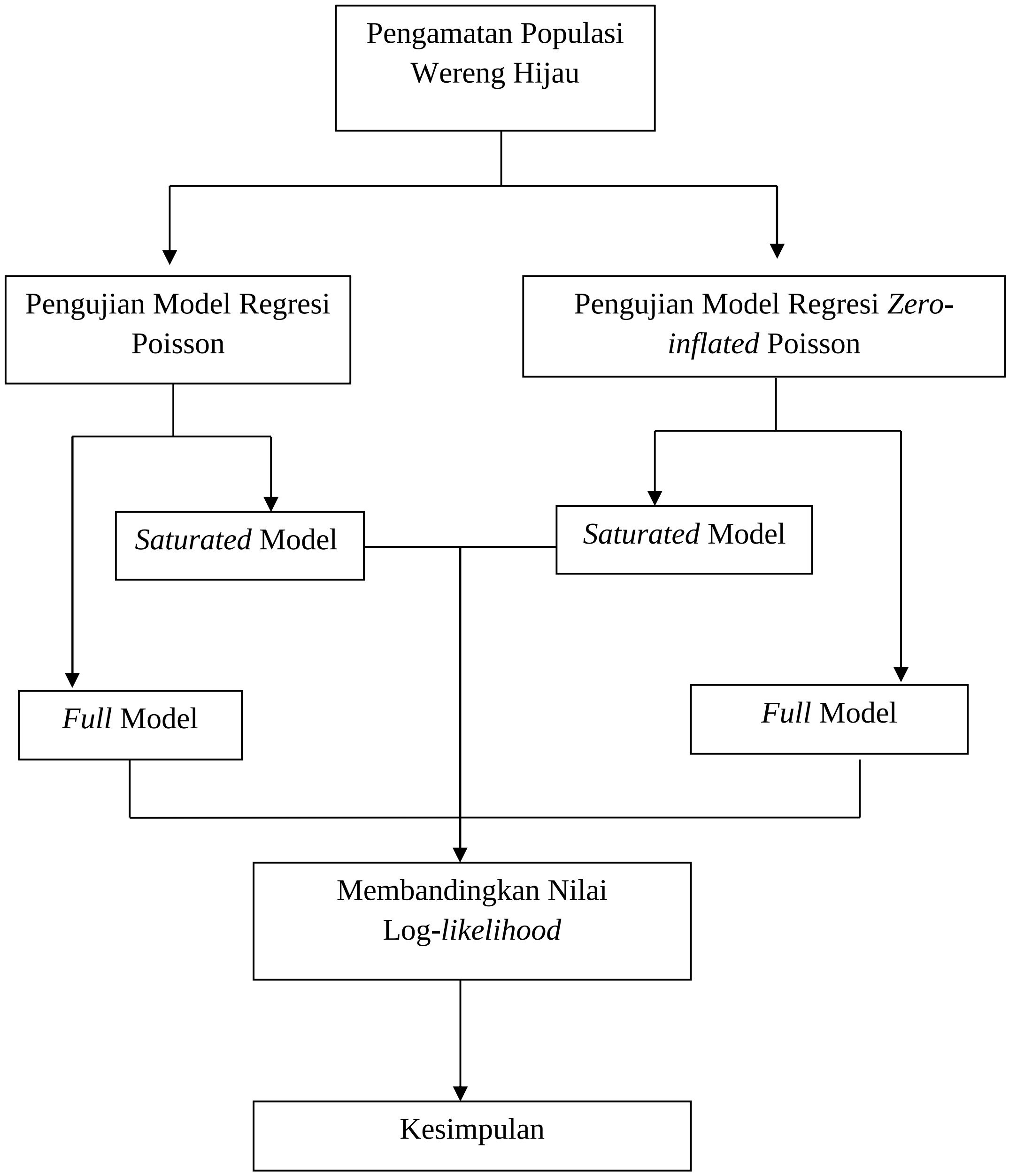
a. Variabel takbebas (Y) adalah populasi Wereng Hijau.
b. Variabel bebas (X) yang akan digunakan dalam penelitian ini disajikan pada Table 3.1 berikut:

Table 3.1 Variabel bebas (X) yang akan digunakan

Variabel Bebas	**Jenis**
Luas lahan sawah (X_1)	numerik
Umur tanaman terserang (X_2)	numerik
Adanya tanaman inang atau gulma (X_3)	presentase
Curah Hujan (X_4)	numerik

3.2 Metode Pengolahan dan Analisis Data

Secara skematik, langkah-langkah yang dilakukan dalam penelitian tentang Pemodelan Populasi Wereng Hijau Penyebab Penyakit Tungro pada Padi di Kabupaten Bondowoso dengan Regresi *Zero-Inflated* Poisson diberikan dalam Gambar 3.1.



Gambar 3.1 Diagram metode penelitian

Untuk mendapatkan suatu model dari data yang telah didapatkan dengan variabel-variabel yang mempengaruhinya, diperlukan suatu langkah-langkah sebagai berikut:

a. Melakukan pengamatan untuk mengetahui populasi wereng hijau sesuai dengan persamaan (2.1).
b. Melakukan pengujian model regresi Poisson dengan menggunakan software program R secara *full* dan *saturated* model.
c. Melakukan pengujian model regresi *Zero-inflated* Poisson dengan menggunakan software program R secara *full* dan *saturated* model.
d. Membandingkan model-model yang telah didapatkan pada pengujian model regresi Poisson dan *Zero-inflated* Poisson dengan melihat nilai log-*likelihood* sehingga didapatkan model terbaik.

Pengolahan data dilakukan dengan menggunakan program R untuk mencari model Regresi Poisson dan *Zero-inflated* Poisson serta mencari model terbaik. Program statistik R (paket R) merupakan paket open source yang dapat diperoleh secara cuma-cuma dari situs http:://www.r-project.org/. Analisis Populasi Wereng Hijau Penyebab Penyakit Tungro pada Padi di Kabupaten Bondowoso dengan Regresi *Zero-Inflated* Poisson dalam program R telah mempunyai paket tersendiri yaitu *pscl*. Untuk fungsi R nya sebagai berikut:

```
zeroinfl(formula, data, subset, na.action, weights, offset,
  dist = c("poisson", "negbin", "geometric"),
  link = c("logit", "probit", "cloglog", "cauchit", "log"),
  control = zeroinfl.control(...),
  model = TRUE, y = TRUE, x = FALSE, ...)
```

BAB.4 HASIL DAN PEMBAHASAN

4.1 Hasil

4.1.1 Model Regresi Poisson

Model regresi Poisson dapat dibentuk dari variabel takbebas dan beberapa variabel bebas yang mempengaruhi. Pemodelan populasi wereng hijau mempunyai beberapa kombinasi model. Model-model yang dibentuk dapat diperoleh dari *full* dan *saturated* model. Berdasarkan hasil olahan program R pada lampiran B diperoleh *full* model regresi Poisson mengenai faktor-faktor yang mempengaruhi populasi wereng hijau di Kabupaten Bondowoso sebagai berikut:

$$\mu = \exp(-0{,}068768 + 0{,}001143x_{1i} - 0{,}092209x_{2i} + 0{,}102258x_{3i} - 0{,}003781x_{4i})$$

Berikut ini disajikan Table 4.1 yang menunjukkan hasil estimasi parameter dari model regresi Poisson.

Tabel 4.1 Estimasi Parameter Model Regresi Poisson

Parameter	Estimasi	SE	*P-value*	Sig	Log-*likelihood*
β_0	-0,068768	0,850978	0,93559	ns	
β_1	0,001143	0,000942	0,22484	ns	
β_2	-0,092209	0,033309	0,00563	**	-95,49355
β_3	0,102258	0,011399	2×10^{-16}	***	
β_4	-0,003781	0,006357	0,55199	ns	

Berdasarkan Tabel 4.1 dapat diketahui variabel yang signifikan terhadap model yaitu parameter β_2 dan β_3. Parameter model regresi Poisson yang dihasilkan dari estimasi parameter belum tentu mempunyai pengaruh yang signifikan terhadap model. Oleh karena itu perlu dilakukan pengujian terhadap kesesuaian model dan parameter model regresi Poisson.

Pengujian parameter secara serentak dapat digunakan untuk mengetahui kesesuaian model. Hipotesis yang digunakan adalah sebagai berikut:

$H_0: \beta_0 = \beta_1 = \cdots = \beta_4 = 0$

H_1: minimal ada satu dari $\beta_r \neq 0$ dengan $r = 0,\ 1, \dots ,\ 4$

taraf signifikansi: $\alpha = 0{,}05$

statistik uji:

$G = -2 \log\left[L\left(y;\hat{\beta}\right) - L\left(y,\hat{\lambda}\right)\right]$

kriteria uji : tolak H_0 jika $G_{Hitung} > X^2_{(5,\alpha)}$ atau *P-value* $< \alpha$

keputusan:

dari tabel distribusi χ^2 diperoleh nilai $X^2_{(5,\ 0{,}05)} = 11{,}070$

karena nilai $G_{Hitung} = 58{,}869 > X^2_{(5,\ 0{,}05)} = 11{,}070$ maka H_0 ditolak.

Karena H_0 ditolak maka disimpulkan bahwa pemodelan secara keseluruhan adalah signifikan atau sesuai, maka paling tidak ada satu varibel dari persamaan mempunyai kontribusi yang signifikan terhadap populasi wereng hijau. Setelah dilakukan pengujian serentak maka dilakukan uji parsial terhadap model awal untuk mengetahui kesignifikansi masing-masing variabel bebasnya dengan menggunakan hipotesis:

$H_0: \beta_r = 0$, untuk suatu $r = 0,\ 1,\ \dots ,\ 4$

$H_1: \beta_r \neq 0$ untuk suatu $r = 0,\ 1,\ \dots ,\ 4$

taraf signifikansi: $\alpha = 0{,}05$

statistik uji: $t_i = \frac{\hat{\boldsymbol{\beta}}_i}{SE(\hat{\boldsymbol{\beta}}_i)}$

kriteria uji: tolak H_0 jika $|t_{hit}| > t_{\alpha/2,v}$ atau *P-value* < α

berdasarkan keluaran program R pada lampiran untuk hasil estimasi parameter dapat dilihat dalam Tabel 4.2 dan $t_{\frac{\alpha}{2},\ v} = 2.571$ yang diperoleh dari tabel sebaran *t*.

Tabel 4.2 Estimasi Parameter Model Regresi Poisson untuk Kesignifikanan Variabel

Variabel	$\hat{\beta}_r$	$SE(\hat{\beta}_r)$	$\lvert Z_{hit} \rvert$	*P-value*	sig	Keputusan
X_1	0,001143	0,000942	1,2133	0,22484	ns	Terima H_0
X_2	-0,092209	0,033309	2,76829	0,00563	**	Tolak H_0
X_3	0,102258	0,011399	8,9707	2×10^{-16}	***	Tolak H_0
X_4	-0,003781	0,006357	0,59478	0,55199	ns	Terima H_0

Berdasarkan hasil keputusan pada Tabel 4.2 dapat diketahui bahwa X_1 dan X_4 tidak signifikan. Hal ini berarti dari empat variabel bebas terdapat dua parameter yaitu X_2 dan X_3 yang signifikan dalam model awal yang diperoleh sebelumnya. Dengan kata lain dapat dinyatakan bahwa faktor-faktor yang mempengaruhi populasi wereng hijau penyebab penyakit tungro di Kabupaten Bondowoso adalah umur tanaman padi terserang penyakit dan adanya tanaman inang atau gulma.

Selain *full* model yang telah didapatkan pada uraian diatas, model regresi Poisson juga dapat diperoleh dari *saturated* model yang merupakan kombinasi dari empat variabel bebas yang mempengaruhi populasi wereng hijau penyebab penyakit tungro. Model-model tersebut adalah model dengan kombinasi dua variabel bebas dan model dengan tiga variabel bebas. Model-model yang akan didapatkan bertujuan untuk mendapatkan model terbaik dengan cara seleksi model menggunakan nilai log-*likelihood*.

A. Model Regresi Poisson Terbaik dengan Dua Variabel Bebas

Untuk mengetahui model terbaik regresi Poisson dengan dua variabel bebas langkah pertama yang dilakukan yaitu melakukan uji kesesuaian model seperti yang telah dilakukan untuk *full* model regresi Poisson sebelumnya. Hipotesis yang digunakan yaitu:

$H_0: \beta_1 = \beta_2 = 0$

H_1: minimal ada satu dari $\beta_r \neq 0$ dengan $r = 1,\ 2$

taraf signifikansi: $\alpha = 0{,}05$

statistik uji:

$G = -2 \log\left[L(y; \hat{\beta}) - L(y, \hat{\lambda})\right]$

kriteria uji : tolak H_0 jika $G_{Hitung} > X^2_{(v,\ \alpha)}$ atau P-*Value* $< \alpha$

Langkah kedua yang dilakukan yaitu melakukan uji parsial terhadap model yang didapatkan untuk mengetahui kesignifikansi masing-masing variabel bebasnya, hipotesis yang digunakan:

$H_0: \beta_r = 0$, untuk suatu $r = 1,\ 2$

$H_1: \beta_r \neq 0$ untuk suatu $r = 1,\ 2$

taraf signifikansi: $\alpha = 0{,}05$

statistik uji: $t_i = \dfrac{\widehat{\boldsymbol{\beta}}_i}{SE(\widehat{\boldsymbol{\beta}}_i)}$

kriteria uji: tolak H_0 jika $|t_{hit}| > t_{\alpha/2,v}$ atau P-*Value* < α

Berdasarkan hasil olahan program R pada lampiran C didapatkan rangkuman hasil regresi Poisson dengan dua variabel bebas pada Tabel 4.3

Tabel 4.3 Rangkuman Model Regresi Poisson dengan Dua Variabel Bebas

No.	Variabel	Log-*Likelihood*	G_{Hitung}	df	χ^2 [illegible]	Keputusan	P-*Value*	sig
1	X_1 X_2	-151,9633	112,9395	3	7,815	Tolak H_0	0,1527 0,1085	ns ns
2	X_1 X_3	-99,9066	8,8261			Tolak H_0	0,0211 2×10^{-16}	* ***
3	X_1 X_4	-141,48	91,9729			Tolak H_0	0.312 99×10^{-7}	ns ***
4	X_2 X_3	-97,17323	3,35936.			Terima H_0	0,00147 2×10^{-16}	** ***
5	X_2 X_4	-140,6751	90,3631			Tolak H_0	0,109 99×10^{-7}	ns ***
6	X_3 X_4	-100,5921	10,1917			Tolak H_0	2×10^{-16} 0,0522	*** ns

Berdasarkan Tabel 4.3 diatas dapat diketahui bahwa dari enam model regresi Poisson dengan dua variabel pembentuk terdapat lima model yang berbeda dengan *full* model karena $G_{Hitung} > X^2_{(v,\ \alpha)}$ sehingga tolak H_0. Pada rangkuman tersebut juga didapatkan satu model regresi dengan variabel bebas X_2 dan X_3 yang sama dengan *full* model karena $G_{Hitung} < X^2_{(v,\ \alpha)}$ dan terima H_0 . Model regresi dengan variabel bebas X_2 dan X_3 dikatakan sama dengan *full* model regresi Poisson karena pada *full* model Poisson variabel yang signifikan hanya X_2 dan X_3. Dari keenam model regresi Poisson yang diperoleh dapat diperoleh model terbaik yaitu dengan nilai log-*likelihood* terbesar dan tingkat signifikan yang tinggi. Model terbaik regresi Poisson dengan dua variabel adalah model dengan variabel bebas X_2 dan X_3 dengan log-*likelihood* = -97,17323 dan tingkat signifikan masing-masing variabel tinggi.

B. Model Regresi Poisson Terbaik dengan Tiga Variabel bebas

Untuk mengetahui model terbaik regresi Poisson dengan tiga variabel bebas langkah pertama yang dilakukan yaitu melakukan uji kesesuaian model seperti yang telah dilakukan untuk *full* model dan model dengan dua variabel bebas regresi Poisson sebelumnya. Hipotesis yang digunakan yaitu:

$H_0: \beta_1 = \beta_2 = \beta_3 = 0$

H_1: minimal ada satu dari $\beta_r \neq 0$ dengan $r = 1,\ 2,\ 3$

taraf signifikansi: $\alpha = 0{,}05$

statistik uji:

$G = -2 \log\left[L\left(y; \hat{\beta}\right) - L\left(y, \hat{\lambda}\right)\right]$

kriteria uji : tolak H_0 jika $G_{Hitung} > X^2_{(v,\ \alpha)}$ atau P-*Value* $< \alpha$

Langkah kedua yang dilakukan yaitu melakukan uji parsial terhadap model yang didapatkan untuk mengetahui kesignifikansi masing-masing variabel bebasnya dengan menggunakan hipotesis:

$H_0: \beta_r = 0$, untuk suatu $r = 1,\ 2,\ 3$

$H_1: \beta_r \neq 0$ untuk suatu $r = 1,\ 2,\ 3$

taraf signifikansi: $\alpha = 0{,}05$

statistik uji: $t_i \approx \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)}$

kriteria uji: tolak H_0 jika $|t_{hit}| > t_{\frac{\alpha}{2},\ v}$ atau P-*Value* < α

Berdasarkan hasil olahan program R pada lampiran D didapatkan rangkuman hasil regresi Poisson dengan tiga variabel bebas pada Tabel 4.4.

Tabel 4.4 Rangkuman Model Regresi Poisson dengan Tiga Variabel Bebas

No	Variabel	Log-*Likelihood*	G_{Hitung}	df	χ^2_{df}	Keputusan	P-*Value*	Signifikan
1	X_1 X_2 X_3	-95,66976	0,35242	4	9,488	Terima H_0	0,07684 0,00447 2×10^{-16}	ns ** ***
2	X_1 X_2 X_4	-140,0786	89,1701			Tolak H_0	0.2784 0,0988 5×10^{-7}	ns ns ***
3	X_2 X_3 X_4	-96,21738	1,44766			Terima H_0	0,0038 2×10^{-16} 0,1666	** *** ns
4	X_1 X_3 X_4	-101,1888	101,549			Tolak H_0	0.179 2×10^{-16} 0,263	ns *** ns

Berdasarkan Tabel 4.4 diatas dapat diketahui bahwa dari empat model regresi Poisson dengan tiga variabel pembentuk terdapat dua model yang dapat dikatakan sama dengan *full* model Poisson karena terima H_0 dan variabel bebas X_2 dan X_3 memberikan pengaruh signifikan terhadap model-model tersebut, model tersebut yaitu model regresi dengan kombinasi variabel-variabel bebas X_1, X_2, X_3 dan X_2,X_3, X_4. Sedangkan model regresi Poisson dengn tiga variabel bebas X_1, X_2 , X_4 dan X_1 ,X_3, X_4 tolak H_0 yang berarti model tersebut berbeda dengan *full* model regresi Poisson. Dari keempat model regresi Poisson dengan tiga variabel yang diperoleh dapat dipilih model terbaik yaitu dengan nilai log-*likelihood* yang terbesar dan tingkat signifikan yang tinggi. Model terbaik regresi

Poisson dengan tiga variabel adalah model dengan variabel bebas X_1, X_2 dan X_3 dengan log-*likelihood* = -95,66976 dan tingkat signifikan masing-masing variabel cukup tinggi walaupun ada satu variabel yang tidak signifikan.

4.1.2 Overdispersi pada Regresi Poisson

Ciri dari regresi Poisson adalah *equidispersi*. Apabila variabel respon mengalami overdispersi maka model regresi Poisson tidak sesuai. Taksiran dispersi dapat diukur dengan nilai *Residual deviance* yang dibagi derajat bebas. Apabila nilai taksiran lebih dari 1 maka ada indikasi overdispesi. Berikut akan disajikan Tabel 4.5 tentang nilai taksiran *disperse* untuk model-model regresi Poisson yang telah diperoleh.

Tabel 4.5 Taksiran *Disperse* Model-model Regresi Poisson

No	Model Regresi Poisson dengan Variabel Bebas	*Residual deviance*	derajat bebas	Taksiran *dispersi*	Keterangan
1	Null Model	206,67	66	3,131364	
2	Full Model	88,934	62	1,434419	
3	x_1 dan x_2	201,87	64	3,154219	
4	x_1 dan x_3	97,76	64	1,5275	
5	x_1 dan x_4	180,91	64	2,826719	
6	x_2 dan x_3	92,293	64	1,442078	Overdispersi
7	x_2 dan x_4	179,30	64	2,801563	
8	x_3 dan x_4	99,131	64	1,548922	
9	x_1, x_2, dan x_3	89,286	63	1,417238	
10	x_1, x_2, dan x_4	178,10	63	2,26984	
11	x_2, x_3, dan x_4	90,381	63	1,434619	
12	x_1, x_3, dan x_4	100,32	63	1,592380	

Dari Tabel 4.5 dapat diketahui bahwa model regresi Poisson yang didapatkan mengalami overdispersi yang mengakibatkan model regresi Poisson menjadi tidak sesuai. Sehingga perlu dilakukan analisis dengan model regresi yang lain yaitu ZIP. Selanjutnya akan dilakukan analisis untuk mendapatkan model regresi ZIP.

4.1.3 Model Regresi *Zero-inflated* Poisson (ZIP)

Model regresi *Zero-inflated* Poisson bertujuan untuk memperbaiki model regresi Poisson karena pada model regresi Poisson dimungkinkan adanya overdipersi pada variabel takbebas. Pemodelan *Zero-inflated* Poisson melibatkan variabel bebas yang sama dengan model regresi Poisson sebelumnya. Seperti dalam model regresi Poisson, akan dilakukan estimasi parameter model regresi ZIP. Hasil olahan program R pada lampiran F memperoleh estimasi parameter regresi ZIP pada Table 4.6.

Tabel 4.6 Estimasi Parameter Model Regresi ZIP

Parameter	Estimasi	SE	*P-value*	Sig	Log-*likelihood*
β_0	1,7787463	0,8751331	0,0421	ns	
β_1	-0,0003873	0,0010370	0,7088	ns	
β_2	-0,0426830	0,0321336	0,1841	ns	
β_3	0,0307344	0,0142562	0,0311	*	
β_4	-0,0061117	0,0062061	0,3247	ns	-67,37
γ_0	-2,172188	7,921490	0,78392	ns	
γ_1	-0,003285	0,006338	60422	ns	
γ_2	0,064570	0,292944	0,82555	ns	
γ_3	-0,750561	0,245305	0,00222	**	
γ_4	0,089653	0,068681	0,19177	ns	

Berdasarkan Tabel 4.6 diatas didapatkan *full* model ZIP untuk model log dan logit sebagai berikut:

$$\log(\mu_i) = 1{,}7787463 - 0{,}0003873x_{1i} - 0{,}0426830x_{2i} + 0{,}0307344x_{3i} \\ -0{,}0061117x_{4i}$$

dan

$$\text{logit}(p_i) = -2{,}172188 - 0{,}003285x_{1i} + 0{,}064570x_{2i} - 0{,}750561x_{3i} \\ + 0{,}089653x_{4i}$$

Parameter model regresi ZIP yang dihasilkan dari estimasi parameter belum tentu mempunyai pengaruh yang signifikan terhadap model. Oleh karena itu perlu dilakukan pengujian terhadap kesesuaian model dan parameter model regresi ZIP seperti yang telah dilakukan pada model regresi Poisson..

Pengujian parameter secara serentak dapat digunakan untuk mengetahui kesesuaian model. Hipotesis yang digunakan adalah sebagai berikut:

$H_0: \beta_0 = \beta_1 = \cdots = \beta_4 = 0$

H_1: minimal ada satu dari $\beta_r \neq 0$ dengan $r = 0,\ 1,\ \ldots,\ 4$

dan

$H_0: \gamma_0 = \gamma_1 = \cdots = \gamma_4 = 0,$

H_1: minimal ada satu dari $\gamma_r \neq 0$ dengan $r = 0,\ 1,\ \ldots,\ 4$

taraf signifikansi: $\alpha = 0{,}05$

statistik uji:

$G = -2 \log\left[L(y;\hat{\beta}) - L(y,\hat{\lambda})\right]$

kriteria uji : tolak H_0 jika $G_{Hitung} > X^2_{(10,\ \alpha)}$ atau *P-value* $< \alpha$

keputusan:

dari tabel distribusi χ^2 diperoleh nilai $X^2_{(10,\ 0{,}05)} = 18{,}307$

karena nilai $G_{Hitung} = 77{,}86 > X^2_{(10,\ 0{,}05)} = 18{,}307$ maka H_0 ditolak.

Karena H_0 ditolak maka disimpulkan bahwa pemodelan secara keseluruhan adalah signifikan atau sesuai maka paling tidak ada satu varibel dari persamaan mempunyai kontribusi yang signifikan terhadap populasi wereng hijau. Setelah dilakukan pengujian serentak maka dilakukan uji parsial terhadap model awal untuk mengetahui kesignifikansi masing-masing variabel bebasnya.

Pengujian parameter secara parsial untuk model log dengan hipotesis sebagai berikut:

$H_0: \beta_r = 0$, untuk suatu $r = 0,\ 1,\ \dots,\ 4$

$H_1: \beta_r \neq 0$ untuk suatu $r = 0,\ 1,\ \dots,\ 4$

dan pengujian parameter secara parsial untuk model logit dengan hipotesis sebagai berikut:

$H_0: \gamma_r = 0$, untuk suatu $r = 0,\ 1,\ \dots,\ 4$

$H_1: \gamma_r \neq 0$ untuk suatu $r = 0,\ 1,\ \dots,\ 4$

taraf signifikansi: $\alpha = 0{,}05$

statistik uji: $t_i = \frac{\hat{\boldsymbol{\beta}}_i}{SE(\hat{\boldsymbol{\beta}}_i)}$

kriteria uji: tolak H_0 jika $|t_{hit}| > t_{\frac{\alpha}{2},\ v}$ atau *P-value* < α

berdasarkan keluaran program R pada lampiran untuk hasil estimasi parameter dapat dilihat dalam Tabel 4.7 dan $t_{\frac{\alpha}{2},\ v} = 2{,}228$ yang diperoleh dari tabel sebaran *t*.

Tabel 4.7 Estimasi Parameter Model Regresi ZIP untuk Kesignifikanan Variabel

Variabel	$\hat{\beta}_r$	$SE(\hat{\beta}_r)$	$\lvert t_{hit} \rvert$	*P-value*	sig	Keputusan
			Model Log			
x_1	-0,0003873	0,0010370	0,373481	0,7088	ns	Terima H_0
x_2	-0,0426830	0,0321336	1,328298	0,1841	ns	Terima H_0
x_3	0,0307344	0,0142562	2,551862	0,0311	*	Tolak H_0
x_4	-0,0061117	0,0062061	0,984789	0,3247	ns	Terima H_0
			Model Logit			
x_1	-0,003285	0,006338	0,518302	0,60422	ns	Terima H_0
x_2	0,064570	0,292944	0,220418	0,82555	ns	Terima H_0
x_3	-0,750561	0,245305	3,059705	0,00222	**	Tolak H_0
x_4	0,089653	0,068681	1,305354	0,19177	ns	Terima H_0

Berdasarkan hasil keputusan pada Tabel 4.7 dapat diketahui bahwa X_1 , X_2 , dan X_4 tidak signifikan untuk model log dan logit. Hal ini berarti dari empat variabel bebas terdapat satu parameter yaitu X_3 yang signifikan dalam model awal. Dengan kata lain dapat dinyatakan bahwa faktor-faktor mempengaruhi populasi wereng hijau penyebab penyakit tungro di Kabupaten Bondowoso adalah adanya tanaman inang atau gulma.

Seperti halnya pada model regresi Poisson, selain *full* model yang telah didapatkan juga dapat diperoleh *saturated* model yang merupakan kombinasi dari empat variabel bebas yang mempengaruhi populasi wereng hijau penyebab penyakit tungro. Model-model tersebut adalah model dengan kombinasi dua variabel bebas dan model dengan tiga variabel bebas. Model-model yang akan didapatkan bertujuan untuk mendapatkan model terbaik dengan cara seleksi model menggunakan nilai log-*likelihood*.

A. Model Regresi ZIP Terbaik dengan Dua Variabel bebas

Untuk mengetahui model terbaik regresi ZIP dengan dua variabel bebas langkah pertama yang dilakukan yaitu melakukan uji kesesuaian model seperti yang telah dilakukan untuk *full* model regresi ZIP sebelumnya. Hipotesis yang digunakan yaitu:

$H_0: \beta_1 = \beta_2 = 0$

H_1: minimal ada satu dari $\beta_r \neq 0$ dengan $r = 1,\ 2$

dan

$H_0: \gamma_1 = \gamma_2 = 0$

H_1: minimal ada satu dari $\gamma_r \neq 0$ dengan $r = 1,\ 2$

taraf signifikansi: $\alpha = 0{,}05$

statistik uji:

$G = -2 \log\left[L(y;\hat{\beta}) - L(y,\hat{\lambda})\right]$

kriteria uji : tolak H_0 jika $G_{Hitung} > X^2_{(6,\ \alpha)}$ atau *P-value* $< \alpha$

Setelah dilakukan pengujian serentak maka dilakukan uji parsial terhadap model regresi ZIP dengan dua variabel bebas untuk mengetahui kesignifikansi masing-masing variabel bebasnya.

Pengujian parameter secara parsial untuk model log dengan hipotesis sebagai berikut:

$H_0: \beta_r = 0$, untuk suatu $r = 1,\ 2$

$H_1: \beta_r \neq 0$ untuk suatu $r = 1,\ 2$

dan pengujian parameter secara parsial untuk model logit dengan hipotesis sebagai berikut:

$H_0: \gamma_r = 0$, untuk suatu $r = 1,\ 2$

$H_1: \gamma_r \neq 0$ untuk suatu $r = 1,\ 2$

taraf signifikansi: $\alpha = 0{,}05$

statistik uji: $t_i = \frac{\widehat{\boldsymbol{\beta}}_i}{SE(\widehat{\boldsymbol{\beta}}_i)}$

kriteria uji: tolak H_0 jika $|t_{hit}| > t_{\frac{\alpha}{2},\ v}$ atau *P-value* < α

berdasarkan keluaran program R pada lampiran G untuk hasil estimasi parameter dapat dilihat dalam Tabel 4.8 dan $t_{\frac{\alpha}{2},v} = 2{,}447$ yang diperoleh dari tabel sebaran *t*.

Tabel 4.8 Rangkuman Model Regresi ZIP dengan dua Variabel Bebas

No.	Variabel	Log-*Likelihood*	GHitung	df	X^2_{α}	Keputusan	sig Log	sig Logit
1	X_1 X_2	-104.7	74.66	6	18.5	Tolak H_0	ns ns	ns ns
2	X_1 X_3	-70.47	6.2	6	18.5	Terima H_0	ns *	** ***
3	X_1 X_4	-100.8	66.86	6	18.5	Tolak H_0	ns ns	ns *
4	X_2 X_3	-70.9	7.06	6	18.5	Terima H_0	ns **	* ***
5	X_2 X_4	-101.3	67.86	6	18.5	Tolak H_0	ns ns	* *
6	X_3 X_4	-68.49	2.24	6	18.5	Terima H_0	* ns	** ns

Berdasarkan Tabel 4.8 diatas dapat diketahui bahwa dari enam model regresi Poisson dengan dua variabel pembentuk terdapat tiga model yang berbeda dengan *full* model karena GHitung > $X^2_{(v,\alpha)}$ sehingga tolak H_0. Selain itu didapatkan tiga model regresi yang sama dengan *full* model karena GHitung < $X^2_{(v,\alpha)}$ dan terima H_0 . Model regresi dengan variabel bebas X_1, X_3; X_2, X_3; dan X_3, X_4 dikatakan sama dengan *full* model regresi ZIP karena pada *full* model variabel yang signifikan hanya X_3 dan variabel X_3 menunjukkan tingkat signifikansi yang paling tinggi. Dari keenam model regresi ZIP yang diperoleh dapat dipilih model terbaik yaitu dengan nilai log-*likelihood* yang terbesar dan tingkat signifikan yang tinggi. Model terbaik regresi ZIP dengan dua variabel adalah model dengan variabel bebas X_3 dan X_4 dengan log-*likelihood* = -68,49 .

B. Model Regresi ZIP Terbaik dengan Tiga Variabel bebas

Untuk mengetahui model terbaik regresi ZIP dengan tiga variabel bebas langkah pertama yang dilakukan yaitu melakukan uji kesesuaian model seperti yang telah dilakukan untuk *full* model regresi ZIP sebelumnya. Hipotesis yang digunakan yaitu:

$H_0: \beta_1 = \beta_2 = \beta_3 = 0$

H_1: minimal ada satu dari $\beta_r \neq 0$ dengan $r = 1, 2, 3$

dan

$H_0: \gamma_1 = \gamma_2 = \gamma_3 = 0$

H_1: minimal ada satu dari $\gamma_r \neq 0$ dengan $r = 1, 2, 3$

taraf signifikansi: $\alpha = 0{,}05$

statistik uji:

$G = -2\log\left[L(y;\hat{\beta}) - L(y,\hat{\lambda})\right]$

kriteria uji : tolak H_0 jika $G_{Hitung} > X^2_{(8,\ \alpha)}$ atau *P-value* $< \alpha$

Setelah dilakukan pengujian serentak maka dilakukan uji parsial terhadap model regresi ZIP dengan dua variabel bebas untuk mengetahui kesignifikansi masing-masing variabel bebasnya.

Pengujian parameter secara parsial untuk model log dengan hipotesis sebagai berikut:

$H_0: \beta_r = 0$, untuk suatu $r = 1, 2, 3$

$H_1: \beta_r \neq 0$ untuk suatu $r = 1, 2, 3$

dan pengujian parameter secara parsial untuk model logit dengan hipotesis sebagai berikut:

$H_0: \gamma_r = 0$, untuk suatu $r = 1, 2, 3$

$H_1: \gamma_r \neq 0$ untuk suatu $r = 1, 2, 3$

taraf signifikansi: $\alpha = 0{,}05$

statistik uji: $t_i \coloneqq \frac{\widehat{\boldsymbol{\beta}}_i}{SE(\widehat{\boldsymbol{\beta}}_i)}$

kriteria uji: tolak H_0 jika $|t_{hit}| > t_{\frac{\alpha}{2},\ v}$ atau *P-value* $< \alpha$

berdasarkan keluaran program R pada lampiran untuk hasil estimasi parameter dapat dilihat dalam Tabel 4.9 dan $t_{\frac{\alpha}{2},\ v} = 2{,}306$ yang diperoleh dari tabel sebaran *t*.

Tabel 4.9 Rangkuman Model Regresi ZIP dengan Tiga Variabel Bebas

No.	Variabel	Log-*Likelihood*	G_{Hitung}	df	χ^2_{tabel}	Keputusan	sig	
							Log	**Logit**
1	X_1 X_2 X_3	-69,52	204,26	8	15,507	Tolak H_0	ns ns *	ns ns **
2	X_1 X_2 X_4	-100,3	235,04			Tolak H_0	ns ns ns	ns ns *
3	X_2 X_3 X_4	-67,52	202,26			Tolak H_0	ns * ns	ns * ns
4	X_1 X_3 X_4	-63,98	81,44			Tolak H_0	ns * ns	ns ns ns

Berdasarkan Tabel 4.9 diatas dapat diketahui bahwa dari empat model regresi ZIP dengan tiga variabel yang dibentuk menghasilkan empat model yang berbeda dengan *full* model karena $G_{Hitung} > X^2_{(v,\ \alpha)}$ sehingga tolak H_0. Dari keempat model regresi ZIP yang diperoleh dapat dipilih model terbaik yaitu dengan nilai log-*likelihood* yang terbesar dan tingkat signifikan yang tinggi. Model terbaik regresi ZIP dengan tiga variabel adalah model dengan variabel bebas X_1, X_3 dan X_4 dengan log-*likelihood* = -63,98.

Berdasarkan paparan hasil penelitian sebelumnya dapat diperoleh model-model terbaik untuk regresi Poisson dan ZIP dengan kombinasi variabel-variabel bebas yang ada. Berikut disajikan tabel *full* dan model terbaik yang telah didapatkan pada pembahasan sebelumnya.

Tabel 4.10 Model Terbaik untuk Regresi Poisson dan ZIP

No	Df	Variabel Poisson	Log-*likelihood* Poisson	Log-*likelihood* ZIP	Variabel ZIP
1	5	X_1, X_2, X_3, X_4	-95,49355	-67,37	X_1, X_2, X_3, X_4
2	3	X_2 dan X_3	-97,17323	-68,49	X_3 dan X_4
3	4	X_1, X_2, X_3	-95,66976	-63,98	X_1, X_3 dan X_4

Telah diketahui pada Tabel 4.2 bahwa untuk *full* model regresi Poisson, variabel yang signifikan adalah variabel X_2 dan X_3. Model terbaik Poisson untuk dua dan tiga variabel bebas pembentuk diperoleh dengan hasil uji kesesuaian model yaitu terima H_0 yang berarti model regresi Poisson terbaik untuk dua dan tiga variabel penyusun sama dengan *full* model Poisson awal yang telah didapatkan. Dari penjelsan tersebut dapat dianalisis untuk menentukan model terbaik regresi Poisson secara keseluruhan. Pada Tabel 4.3 dapat diketahui model Poisson terbaik untuk dua variabel bebas adalah model dengan variabel X_2 dan X_3 dimana kedua variabel tersebut signifikan dengan nilai log-*likelihood* -97,17323. Begitu juga dengan model terbaik dengan tiga variabel pembentuk yaitu X_1, X_2, dan X_3, pada Tabel 4.4 dapat diketahui bahwa variabel X_2 dan X_3 signifikan dan mempengaruhi model regresi Poisson yang diperoleh.

Berikutnya akan ditentukan model terbaik regresi ZIP, dari penjelasan hasil penelitian sebelumnya, pada ketiga model ZIP pada tabel 4.10 variabel yang memberikan pengaruh terhadap model yang didapat hanya variabel X_3 . Berdasarkan Tabel 4.10 model ZIP dengan variabel bebas X_1, X_3 dan X_4 menghasilkan nilai log-*likelihood* yang terbesar yaitu senilai -63,98.

4.2 Pembahasan

Berdasarkan hasil penelitian pada subbab sebelumnya dapat diketahui pemodelan terbaik untuk populasi wereng hijau melalui model terbaik Poisson dan ZIP. Dari kedua model terbaik Poisson dan ZIP pada Tabel 4.10 dapat diketahui bahwa model regresi ZIP selalu memberikan nilai log-*likelihood* yang lebih besar dibandingkan dengan model regresi Poisson untuk tiap-tiap *saturated* dan *full* model yang diperoleh. Sehingga dapat diketahui bahwa pemodelan populasi wereng hijau penyebab penyakit tungro di Kabupaten Bondowoso lebih baik didekati dengan model regresi ZIP. Sehingga diperoleh model terbaik dalam penelitian ini yaitu model regresi ZIP dengan dengan tiga variabel bebas X_1, X_3 dan X_4, model regresinya yaitu:

$$\log(\mu_i) = 1.3925955 - 0.0001230X_{1i} + 0.0305248X_{3i} \\ -0.0064362X_{4i}$$

dan

$$\text{logit}(p_i) = -265.75854 + 0.09628X_{1i} - 28.74401X_{3i} + 4.76248X_{4i}$$

Dimana X_1 adalah luas lahan sawah, X_3 adalah adanya tanaman inang atau gulma, dan X_4 merupakan curah hujan. Model logit regresi ZIP menjelaskan bahwa probabilitas populasi wereng hijau di wilayah Kabupaten Bondowoso tidak dipengaruhi oleh luas lahan sawah, jumlah tanaman inang atau gulma, dan curah hujan.

Selain itu dapat diketahui juga bahwa model regresi ZIP lebih baik dari model regresi Poisson berdasarkan nilai residu pada Lampiran K. Pada Lampiran K dapat diketahui bahwa nilai residu untuk regresi Poisson menunjukkan angka yang lebih besar dari pada regresi ZIP. Perbedaan yang cukup besar antara nilai residu regresi Poisson dan ZIP menunjukkan bahwa model regresi ZIP lebih tepat digunakan untuk memodelkan populasi wereng hijau penyebab penyakit tungro dengan banyak mengandung nilai nol.

BAB 5. PENUTUP

5.1 Kesimpulan

Berdasarkan hasil analisis dan pembahasan dapat diambil kesimpulan sebagai berikut:

a. Nilai log-*likelihood* pada model regresi ZIP selalu menghasilkan nilai lebih besar dari pada nilai log-*likelihood* yang dihasilkan pada model regresi Poisson untuk semua model.
b. Nilai residu untuk regresi Poisson menunjukkan angka yang lebih besar dari pada regresi ZIP, hal ini berarti model regresi ZIP lebih baik dari pada Poisson.
c. Pemodelan populasi wereng hijau penyebab penyakit tungro di Kabupaten Bondowoso lebih baik didekati dengan pemodelan regresi ZIP.
d. Dari pemodelan regresi Poisson dan ZIP didapatkan model terbaik untuk pemodelan populasi wereng hijau penyebab penyakit tungro di Kabupaten Bondowoso yaitu:

   $$\log(\mu_i) = 1.3925955 - 0.0001230X_{1i} + 0.0305248X_{3i} - 0.0064362X_{4i}$$

   dan

   $$\text{logit}(p_i) = -265.75854 + 0.09628x_{1i} - 28.74401x_{3i} + 4.76248x_{4i}$$

   dimana X_1 adalah luas lahan sawah, X_3 adalah adanya tanaman inang atau gulma, dan X_4 merupakan curah hujan.

5.2 Saran

Penelitian ini dapat dikembangkan lebih lanjut dengan metode lain, seperti *Zero-truncated* Poisson dan *Zero-inflated* Negatif Binomial.

DAFTAR PUSTAKA

Anonim. 2009. *Wereng Menyerang Sawah di Bondowoso* [on line]. http://www.kabarbisnis.com/read/283725. [19 Desember 2012]

Cameron, A.C, dan Trivedi, P.K. 1998. *Regression Analysis of count data*. Cambridge: Cambridge University Press.

Departmen Pertanian. 1985. *Hama Tungro dan cara Mengatasinya*. Surabaya: Proyek Pengembangan Penyuluhan Pertaniam Pusat/ NAEP.

Dinas Pertanian. 2012. *Penyusunan Data Base Potensi Produk Pangan Kabupaten Bondowoso*. Bondowoso: Dinas Pertanian.

Direktorat Jenderal Tanaman Pangan. 2007. *Pedoman Rekomendasi Pengendalian Hama Terpadu Pada Tanaman Padi*. Jakarta: Direktorat Jenderal Bina Produksi Tanaman Pangan.

Famoye, F., Wulu, J.T. dan Singh, K.P. 2004. On The Generalize Poisson Regression Model with an Application to Accident Data. *Journal of Data Science* **2**(2004) 287-298.

Gallagher, K. *Pengendalian Hama Terpadu Untuk Padi*. Jakarta: Direktorat Jenderal Bina Produksi Tanaman Pangan.

Hall, D.B. & Shen. J. 2009. Robust Estimation for Zero-Imflated Poisson Regession. *Scandinavian Journal of Statistics* **10.**1111/j/1467-9469**.**2009.00657.x.

Hardin, J.W dan Hilbe, J.M. 2007. *Generalized Linier Models and Extensions*. Texas: Stata Press.

Ismail, N. & Jemain, A. A. 2005. Generalized Poisson Regression: An Alternative For Risk Classification. *Jurnal Teknologi*,**43**(C):39-54.

Kleinbaum, Kupper, Muller, dan Nizam.1998. *Applied Regression Analysis and Other Multivariable Methods 3rd Editon.* London: Brooks/Cole Publishing Comp.

Lam, K.F., Xue, H, dan Cheung, Y. B. 2006. Semiparametric Analysis of Zero-Inflated Count Data. *Biometric*s, **62**: (996-1003).

Lambert, D. 1992. Zero-Inflated Poisson Regression, With an Application to Defects in Manufacturing. *Technometrics*, **34**(1):1-14.

McCullagh & Nelder. 1989. *Generalize Linear Models*. Chicago: University of Chicago.

Myers, R. H. 1990. *Classical and Modern Regression with Applications*. Boston: PWS-KENT Publising Company.

Nasoetion, A. H. & Rambe, A. 1983. *Teori Statistika Untuk Ilmu-ilmu Kuantitatif*. Jakarta: Bhratara Karya Aksara.

Nurhayati, N. 2012. *Teori Peluang* (PAM 2231)-Unsoed [on line]. http://nunung.blog.unsoed.ac.id/files/2012/04/Slide-140512_Distribusi_Poisson.pdf. [6 Maret 2012]

LAMPIRAN

A. Data Populasi Wereng Hijau yang diambil dari Dinas Tanaman Pangan dan Hortikultura Kabupaten Bondowoso Tahun 2012 sebagai berikut.

Kecamatan	Desa	Populasi WH	Luas Tanam	Umur Padi Terserang	Gulma	Curah Hujan
Maesan	Sucolor	2.00	429.50	7.00	15.00	90.00
Maesan	PujerBaru	3.00	320.99	6.00	10.00	90.00
Maesan	Tanah Wulan	5.00	399.96	8.00	10.00	90.00
Maesan	Maesan	4.00	92.44	14.00	30.00	90.00
Maesan	Gambangan	4.00	186.66	9.00	10.00	90.00
Maesan	Suger Lor	8.00	192.00	4.00	15.00	90.00
Maesan	Sumber Pakem	2.00	246.79	5.00	20.00	90.00
Maesan	Sumber Sari	0.00	356.23	6.00	5.00	90.00
Maesan	Sumber Anyar	4.00	376.62	12.00	25.00	90.00
Maesan	Penanggungan	6.00	133.36	10.00	30.00	90.00
Maesan	Pakuniran	6.00	246.40	5.00	20.00	90.00
Maesan	Gunung Sari	3.00	390.67	8.00	10.00	90.00
Grujugan	Sumber Pandan	0.00	148.39	7.00	0.00	107.00
Grujugan	Pekauman	0.00	202.08	7.00	4.00	107.00
Grujugan	Wanisodo	0.00	277.47	5.00	8.00	107.00
Grujugan	Dawuhan	0.00	149.81	5.00	0.00	107.00
Grujugan	Kabuaran	0.00	153.84	10.00	5.00	107.00
Grujugan	Wonosari	0.00	469.09	10.00	4.00	107.00
Grujugan	Dadapan	0.00	187.02	9.00	7.00	107.00
Grujugan	Taman	6.00	225.57	4.00	15.00	107.00
Grujugan	Tegal Mijin	0.00	258.48	5.00	10.00	107.00
Grujugan	Grujugan Kidul	0.00	152.37	7.00	5.00	107.00
Grujugan	Kejawan	0.00	126.96	7.00	6.00	107.00
Tenggarang	Koncer Kidul	2.00	223.76	5.00	10.00	110.00
Tenggarang	Sumber Salam	3.00	227.31	4.00	20.00	110.00
Tenggarang	Pekalangan	4.00	115.80	10.00	20.00	110.00
Tenggarang	Kasemek	3.00	180.59	14.00	15.00	110.00
Tenggarang	Lojajar	3.00	145.76	10.00	20.00	110.00
Tenggarang	Kajar	5.00	114.77	10.00	30.00	110.00
Tenggarang	Bataan	2.00	253.00	9.00	10.00	110.00
Tenggarang	Gebang	2.00	81.54	5.00	10.00	110.00

Tenggarang	Dawuhan	4.00	121.23	5.00	15.00	110.00
Tenggarang	Tenggarang	3.00	99.22	5.00	15.00	110.00
Tenggarang	Tangsil Kulon	4.00	236.70	5.00	20.00	110.00
Tenggarang	Koncer Darul Aman	0.00	153.10	10.00	3.00	110.00
Bondowoso	Pancoran	0.00	250.56	10.00	5.00	115.00
Bondowoso	Sukowiryo	0.00	197.03	9.00	4.00	115.00
Bondowoso	Kembang	0.00	255.44	7.00	5.00	115.00
Bondowoso	Nangkaan	0.00	141.84	7.00	10.00	115.00
Bondowoso	Tamansari	0.00	250.56	6.00	0.00	115.00
Bondowoso	Dabasah	0.00	27.83	9.00	3.00	115.00
Bondowoso	Badean	0.00	78.98	10.00	3.00	115.00
Bondowoso	Kotakulon	0.00	43.96	9.00	0.00	115.00
Bondowoso	Blindungan	0.00	18.91	7.00	6.00	115.00
Bondowoso	Kademangan	0.00	24.91	7.00	5.00	115.00
Bondowoso	Pejaten	0.00	198.80	5.00	0.00	115.00
Curahdami	Jetis	0.00	366.89	10.00	5.00	85.00
Curahdami	Paku Wesi	0.00	344.16	9.00	0.00	85.00
Curahdami	Kupang	0.00	468.17	7.00	5.00	85.00
Curahdami	Petung	0.00	124.55	7.00	10.00	85.00
Curahdami	Penambangan	0.00	136.98	7.00	5.00	85.00
Curahdami	Curahpoh	0.00	224.04	5.00	3.00	85.00
Curahdami	Curahdami	0.00	197.08	9.00	0.00	85.00
Curahdami	Poncogati	0.00	100.43	12.00	5.00	85.00
Curahdami	Sumbersuko	0.00	392.70	10.00	6.00	85.00
Curahdami	Silolembu	0.00	206.42	10.00	5.00	85.00
Curahdami	Locare	0.00	223.22	6.00	0.00	85.00
Curahdami	Sumber Salak	0.00	267.68	7.00	3.00	85.00
Tamanan	Sukosari	5.00	208.55	7.00	25.00	67.00
Tamanan	Karang Melok	5.00	134.62	8.00	25.00	67.00
Tamanan	Mengen	8.00	265.48	5.00	30.00	67.00
Tamanan	Kemirian	4.00	502.26	5.00	20.00	67.00
Tamanan	Tamanan	4.00	404.90	12.00	20.00	67.00
Tamanan	Wonosuko	3.00	350.85	9.00	10.00	67.00
Tamanan	Kalianyar	2.00	320.26	10.00	10.00	67.00
Tamanan	Sumber Kemuning	4.00	295.37	4.00	10.00	67.00
Tamanan	Sumber Anom	4.00	278.37	5.00	10.00	67.00

B. Skrip dan Output Program R untuk Full Model Regresi Poisson

```
> summary(glm(Y ~ X1 +X2 +X3 +X4,family=poisson, data=jan1))

Call:
glm(formula = Y ~ X1 + X2 + X3 + X4, family = poisson, data = jan1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.7133  -1.1340  -0.8799   0.6144   2.6397

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.068768   0.850978  -0.081  0.93559
X1           0.001143   0.000942   1.214  0.22484
X2          -0.092209   0.033309  -2.768  0.00563 **
X3           0.102258   0.011399   8.971  < 2e-16 ***
X4          -0.003781   0.006357  -0.595  0.55199
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 206.672  on 66  degrees of freedom
Residual deviance:  88.934  on 62  degrees of freedom
AIC: 200.99

Number of Fisher Scoring iterations: 6

> logLik(glm(Y ~ X1 +X2 +X3 +X4,family=poisson, data=jan1))
'log Lik.' -95.49355 (df=5)
```

C. Skrip dan Output Program R untuk Dua Variabel Bebas Regresi Poisson

Program R untuk regresi Poisson dengan dua variabel bebas X_1 dan X_2

```
> summary(glm(Y ~ X1 +X2,family=poisson, data=jan1))

Call:
glm(formula = Y ~ X1 + X2, family = poisson, data = jan1)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-2.252  -1.888  -1.657   1.087   2.976

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.8347064  0.3451977   2.418   0.0156 *
X1           0.0011112  0.0007771   1.430   0.1527
X2          -0.0605828  0.0377478  -1.605   0.1085
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 206.67  on 66  degrees of freedom
Residual deviance: 201.87  on 64  degrees of freedom
AIC: 309.93

Number of Fisher Scoring iterations: 6


> logLik(glm(Y ~ X1 +X2,family=poisson, data=jan1))
'log Lik.' -151.9633 (df=3)
```

Program R untuk regresi Poisson dengan dua variabel bebas X_1 dan X_3

```
> summary(glm(Y ~ X1 +X3,family=poisson, data=jan1))

Call:
glm(formula = Y ~ X1 + X3, family = poisson, data = jan1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.5931  -1.1314  -0.9149   0.5357   3.3272

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.2203663  0.2987682  -4.085 4.41e-05 ***
X1           0.0017765  0.0007705   2.306   0.0211 *
X3           0.0999449  0.0098768  10.119  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 206.67  on 66  degrees of freedom
Residual deviance:  97.76  on 64  degrees of freedom
AIC: 205.81

Number of Fisher Scoring iterations: 6


> logLik(glm(Y ~ X1 +X3,family=poisson, data=jan1))
'log Lik.' -99.9066 (df=3)
```

Program R untuk regresi Poisson dengan dua variabel bebas X_1 dan X_4

```
> summary(glm(Y ~ X1 +X4,family=poisson, data=jan1))

Call:
glm(formula = Y ~ X1 + X4, family = poisson, data = jan1)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-2.319  -1.642  -1.398   0.904   3.063

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.5653038  0.6599141   5.403 6.57e-08 ***
X1          -0.0008963  0.0008871  -1.010    0.312
X4          -0.0292508  0.0059264  -4.936 7.99e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 206.67  on 66  degrees of freedom
Residual deviance: 180.91  on 64  degrees of freedom
AIC: 288.96

Number of Fisher Scoring iterations: 6


> logLik(glm(Y ~ X1 +X4,family=poisson, data=jan1))
'log Lik.' -141.48 (df=3)
```

Program R untuk regresi Poisson dengan tiga variabel bebas X_2, dan X_3

```
> summary(glm(Y ~ X2 +X3,family=poisson, data=jan1))

Call:
glm(formula = Y ~ X2 + X3, family = poisson, data = jan1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.7545  -1.1287  -0.9387   0.3549   2.6779

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.08166    0.28496  -0.287  0.77445
X2          -0.10468    0.03292  -3.180  0.00147 **
X3           0.10363    0.01001  10.353  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 206.672  on 66  degrees of freedom
Residual deviance:  92.293  on 64  degrees of freedom
AIC: 200.35

Number of Fisher Scoring iterations: 6


> logLik(glm(Y ~ X2 +X3,family=poisson, data=jan1))
'log Lik.' -97.17323 (df=3)
```

Program R untuk regresi Poisson dengan dua variabel bebas X_2, dan X_4

```
> summary(glm(Y ~ X2 +X4,family=poisson, data=jan1))

Call:
glm(formula = Y ~ X2 + X4, family = poisson, data = jan1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.3256  -1.6420  -1.3521   0.9242   2.7866

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.521752   0.541118   6.508 7.60e-11 ***
X2          -0.059557   0.037126  -1.604    0.109
X4          -0.026226   0.005254  -4.992 5.99e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 206.67  on 66  degrees of freedom
Residual deviance: 179.30  on 64  degrees of freedom
AIC: 287.35

Number of Fisher Scoring iterations: 6


> logLik(glm(Y ~ X2 +X4,family=poisson, data=jan1))
'log Lik.' -140.6751 (df=3)
```

Program R untuk regresi Poisson dengan dua variabel bebas X_3, dan X_4

```
> summary(glm(Y ~ X3 +X4,family=poisson, data=jan1))

Call:
glm(formula = Y ~ X3 + X4, family = poisson, data = jan1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6661  -1.1864  -0.9068   0.7877   3.1300

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.322963   0.574633   0.562   0.5741
X3           0.090043   0.009896   9.099   <2e-16 ***
X4          -0.010536   0.005427  -1.941   0.0522 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 206.672  on 66  degrees of freedom
Residual deviance:  99.131  on 64  degrees of freedom
AIC: 207.18

Number of Fisher Scoring iterations: 6


> logLik(glm(Y ~ X3 +X4,family=poisson, data=jan1))
'log Lik.' -100.5921 (df=3)
```

D. Skrip dan Output Program R untuk Tiga Variabel Bebas Regresi Poisson

Program R untuk regresi Poisson dengan tiga variabel bebas X_1, X_2, dan X_3

```
> summary(glm(Y ~ X1 +X2 +X3,family=poisson, data=jan1))

Call:
glm(formula = Y ~ X1 + X2 + X3, family = poisson, data = jan1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.7552  -1.1220  -0.8668   0.6294   2.6751

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.5217125  0.3854241  -1.354  0.17586
X1           0.0014307  0.0008086   1.769  0.07684 .
X2          -0.0940085  0.0330697  -2.843  0.00447 **
X3           0.1054032  0.0101737  10.360  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 206.672  on 66  degrees of freedom
Residual deviance:  89.286  on 63  degrees of freedom
AIC: 199.34

Number of Fisher Scoring iterations: 6


> logLik(glm(Y ~ X1 +X2 +X3,family=poisson, data=jan1))
'log Lik.' -95.66976 (df=4)
```

Program R untuk regresi Poisson dengan tiga variabel bebas X_1, X_2, dan X_4

```
> summary(glm(Y ~ X1 +X2 +X4,family=poisson, data=jan1))

Call:
glm(formula = Y ~ X1 + X2 + X4, family = poisson, data = jan1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.3674  -1.6739  -1.3054   0.9716   2.7020

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  4.041533   0.717339   5.634 1.76e-08 ***
X1          -0.000983   0.000907  -1.084   0.2784
X2          -0.061329   0.037151  -1.651   0.0988 .
X4          -0.029226   0.005892  -4.960 7.05e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 206.67  on 66  degrees of freedom
Residual deviance: 178.10  on 63  degrees of freedom
AIC: 288.16

Number of Fisher Scoring iterations: 6


> logLik(glm(Y ~ X1 +X2 +X4,family=poisson, data=jan1))
'log Lik.' -140.0786 (df=4)
```

Program R untuk regresi Poisson dengan tiga variabel bebas X_2, X_3, dan X_4

```
> summary(glm(Y ~ X2 +X3 +X4,family=poisson, data=jan1))

Call:
glm(formula = Y ~ X2 + X3 + X4, family = poisson, data = jan1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6721  -1.1525  -0.9172   0.5064   2.5962

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.648760   0.596296   1.088   0.2766
X2          -0.096817   0.033447  -2.895   0.0038 **
X3           0.098024   0.010698   9.163   <2e-16 ***
X4          -0.007569   0.005472  -1.383   0.1666
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 206.672  on 66  degrees of freedom
Residual deviance:  90.381  on 63  degrees of freedom
AIC: 200.43

Number of Fisher Scoring iterations: 6


> logLik(glm(Y ~ X2 +X3 +X4,family=poisson, data=jan1))
'log Lik.' -96.21738 (df=4)
```

Program R untuk regresi Poisson dengan tiga variabel bebas $X_1, X_3,$ dan X_4

```
> summary(glm(Y ~ X1 +X3 +X4,family=poisson, data=jan1))

Call:
glm(formula = Y ~ X1 + X3 + X4, family = poisson, data = jan1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.5469  -1.1885  -0.9142   0.6681   3.2155

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.2928025  0.8153710  -0.359    0.720
X1           0.0012111  0.0009005   1.345    0.179
X3           0.0920365  0.0106631   8.631   <2e-16 ***
X4          -0.0071145  0.0063590  -1.119    0.263
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 206.67  on 66  degrees of freedom
Residual deviance: 100.32  on 63  degrees of freedom
AIC: 210.38

Number of Fisher Scoring iterations: 6

> logLik(glm(Y ~ X1 +X3 +X4,family=poisson, data=jan1))
'log Lik.' -101.1888 (df=4)
```

E. Skrip dan Output Program R untuk Null Model Regresi Poisson

```
> summary(glm(Y ~ NULL,family=poisson, data=jan1))

Call:
glm(formula = Y ~ NULL, family = poisson, data = jan1)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-1.947  -1.947  -1.947   1.329   3.291

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.63949    0.08874   7.207 5.73e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 206.67  on 66  degrees of freedom
Residual deviance: 206.67  on 66  degrees of freedom
AIC: 310.73

Number of Fisher Scoring iterations: 6

> logLik(glm(Y ~ NULL,family=poisson, data=jan1))
'log Lik.' -154.3629 (df=1)
```

F. Skrip dan Output Program R untuk Full Model Regresi *Zero-inflated* Poisson

```
> summary(zeroinfl(Y ~ X1 +X2 +X3 +X4 | X1 +X2 +X3+X4, data = jan1))

Call:
zeroinfl(formula = Y ~ X1 + X2 + X3 + X4 | X1 + X2 + X3 + X4, data = jan1)

Pearson residuals:
     Min       1Q   Median       3Q      Max 
-1.46154 -0.27248 -0.07403  0.10572  4.04227 

Count model coefficients (poisson with log link):
              Estimate Std. Error z value Pr(>|z|)  
(Intercept)  1.7787463  0.8751331   2.033   0.0421 *
X1          -0.0003873  0.0010370  -0.374   0.7088  
X2          -0.0426830  0.0321336  -1.328   0.1841  
X3           0.0307344  0.0142562   2.156   0.0311 *
X4          -0.0061117  0.0062061  -0.985   0.3247  

Zero-inflation model coefficients (binomial with logit link):
             Estimate Std. Error z value Pr(>|z|)   
(Intercept) -2.172188   7.921490  -0.274  0.78392   
X1          -0.003285   0.006338  -0.518  0.60422   
X2           0.064570   0.292944   0.220  0.82555   
X3          -0.750561   0.245305  -3.060  0.00222 **
X4           0.089653   0.068681   1.305  0.19177   
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 

Number of iterations in BFGS optimization: 22 
Log-likelihood: -67.37 on 10 Df
```

G. Skrip dan Output Program R untuk Dua Variabel Bebas Regresi ZIP

Program R untuk regresi ZIP dengan dua variabel bebas X_1 dan X_2

```
> summary(zeroinfl(Y ~ X1 +X2 | X1 +X2, data = jan1))

Call:
zeroinfl(formula = Y ~ X1 + X2 | X1 + X2, data = jan1)

Pearson residuals:
    Min      1Q  Median      3Q     Max 
-1.0672 -0.7584 -0.6046  0.7029  2.2890 

Count model coefficients (poisson with log link):
              Estimate Std. Error z value Pr(>|z|)    
(Intercept)  1.6728625  0.3282848   5.096 3.47e-07 ***
X1          -0.0005611  0.0008524  -0.658    0.510    
X2          -0.0245898  0.0314968  -0.781    0.435    

Zero-inflation model coefficients (binomial with logit link):
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.421877   0.996523   0.423    0.672
X1          -0.003279   0.002373  -1.382    0.167
X2           0.046833   0.103733   0.451    0.652
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 

Number of iterations in BFGS optimization: 12 
Log-likelihood: -104.7 on 6 Df
```

Program R untuk regresi ZIP dengan dua variabel bebas X_1 dan X_3

```
> summary(zeroinfl(Y ~ X1 +X3 | X1 +X3, data = jan1))

Call:
zeroinfl(formula = Y ~ X1 + X3 | X1 + X3, data = jan1)

Pearson residuals:
     Min       1Q   Median       3Q      Max 
-1.07315 -0.25226 -0.13940  0.01239 11.16690 

Count model coefficients (poisson with log link):
             Estimate Std. Error z value Pr(>|z|)  
(Intercept) 0.8230985  0.3705496   2.221   0.0263 *
X1          0.0001443  0.0008370   0.172   0.8631  
X3          0.0288348  0.0128429   2.245   0.0248 *

Zero-inflation model coefficients (binomial with logit link):
             Estimate Std. Error z value Pr(>|z|)    
(Intercept)  8.011353   2.499638   3.205 0.001351 ** 
X1          -0.007745   0.004909  -1.578 0.114677    
X3          -0.726625   0.210864  -3.446 0.000569 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 

Number of iterations in BFGS optimization: 22 
Log-likelihood: -70.47 on 6 Df
```

Program R untuk regresi ZIP dengan dua variabel bebas X_1 dan X_4

```
> summary(zeroinfl(Y ~ X1 +X4 | X1 +X4, data = jan1))

Call:
zeroinfl(formula = Y ~ X1 + X4 | X1 + X4, data = jan1)

Pearson residuals:
    Min      1Q  Median      3Q     Max 
-0.9961 -0.6440 -0.5417  0.5652  2.4081 

Count model coefficients (poisson with log link):
              Estimate Std. Error z value Pr(>|z|)    
(Intercept)  2.6197748  0.6868803   3.814 0.000137 ***
X1          -0.0013553  0.0009761  -1.389 0.164971    
X4          -0.0104197  0.0060163  -1.732 0.083288 .  

Zero-inflation model coefficients (binomial with logit link):
              Estimate Std. Error z value Pr(>|z|)  
(Intercept) -4.0040201  2.3126952  -1.731   0.0834 .
X1          -0.0005059  0.0027938  -0.181   0.8563  
X4           0.0429914  0.0200013   2.149   0.0316 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 

Number of iterations in BFGS optimization: 13 
Log-likelihood: -100.8 on 6 Df
```

Program R untuk regresi ZIP dengan dua variabel bebas X_2 dan X_3

```
> summary(zeroinfl(Y ~ X2 +X3 | X2 +X3, data = jan1))

Call:
zeroinfl(formula = Y ~ X2 + X3 | X2 + X3, data = jan1)

Pearson residuals:
     Min       1Q   Median       3Q      Max 
-1.24056 -0.26536 -0.12863  0.04979 15.36148 
```

```
Count model coefficients (poisson with log link):
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.08475    0.29818   3.638 0.000275 ***
X2          -0.04388    0.03200  -1.371 0.170281
X3           0.03400    0.01310   2.594 0.009473 **

Zero-inflation model coefficients (binomial with logit link):
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  6.29683    3.11598   2.021 0.043298 *
X2          -0.05053    0.27041  -0.187 0.851774
X3          -0.70721    0.21259  -3.327 0.000879 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Number of iterations in BFGS optimization: 16
Log-likelihood: -70.9 on 6 Df
```

Program R untuk regresi ZIP dengan dua variabel bebas X_2 dan X_4

```
> summary(zeroinfl(Y ~ X2 +X4 | X2 +X4, data = jan1))

Call:
zeroinfl(formula = Y ~ X2 + X4 | X2 + X4, data = jan1)

Pearson residuals:
    Min      1Q  Median      3Q     Max
-1.0678 -0.6544 -0.4998  0.6442  2.2311

Count model coefficients (poisson with log link):
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.092990   0.530441   3.946 7.95e-05 ***
X2          -0.023809   0.031858  -0.747    0.455
X4          -0.006232   0.005258  -1.185    0.236

Zero-inflation model coefficients (binomial with logit link):
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -4.82583    1.97933  -2.438   0.0148 *
X2           0.06509    0.10819   0.602   0.5474
X4           0.04522    0.01770   2.555   0.0106 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Number of iterations in BFGS optimization: 14
Log-likelihood: -101.3 on 6 Df
```

Program R untuk regresi ZIP dengan dua variabel bebas X_3 dan X_4

```
> summary(zeroinfl(Y ~ X3 +X4 | X3 +X4, data = jan1))


Call:
zeroinfl(formula = Y ~ X3 + X4 | X3 + X4, data = jan1)

Pearson residuals:
     Min       1Q   Median       3Q      Max
-1.58889 -0.31080 -0.09137  0.09406  3.32428

Count model coefficients (poisson with log link):
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.324826   0.543887   2.436   0.0149 *
X3           0.028452   0.012248   2.323   0.0202 *
X4          -0.005194   0.005212  -0.997   0.3189

Zero-inflation model coefficients (binomial with logit link):
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.90058    5.22591  -0.746   0.4554
```

```
X3           -0.79988    0.29394   -2.721    0.0065 **
X4            0.10781    0.06597    1.634    0.1022
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Number of iterations in BFGS optimization: 14
Log-likelihood: -68.49 on 6 Df
```

H. Skrip dan Output Program R untuk Tiga Variabel Bebas Regresi ZIP

Program R untuk regresi ZIP dengan tiga variabel bebas X_1, X_2, dan X_3

```
> summary(zeroinfl(Y ~ X1 +X2 +X3 | X1 +X2 +X3, data = jan1))

Call:
zeroinfl(formula = Y ~ X1 + X2 + X3 | X1 + X2 + X3, data = jan1)

Pearson residuals:
    Min       1Q  Median      3Q     Max 
-1.2453 -0.2514 -0.1054  0.1015 12.9590 

Count model coefficients (poisson with log link):
              Estimate Std. Error z value Pr(>|z|)   
(Intercept)  1.0605203  0.4070297   2.606  0.00917 **
X1           0.0001045  0.0008421   0.124  0.90129   
X2          -0.0434689  0.0319605  -1.360  0.17380   
X3           0.0339588  0.0135246   2.511  0.01204 * 

Zero-inflation model coefficients (binomial with logit link):
             Estimate Std. Error z value Pr(>|z|)   
(Intercept)  7.847781   3.502208   2.241  0.02504 * 
X1          -0.007816   0.004995  -1.565  0.11768   
X2           0.012925   0.282272   0.046  0.96348   
X3          -0.717897   0.221781  -3.237  0.00121 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Number of iterations in BFGS optimization: 18
Log-likelihood: -69.52 on 8 Df
```

Program R untuk regresi ZIP dengan tiga variabel bebas X_1, X_2, dan X_4

```
> summary(zeroinfl(Y ~ X1 +X2 +X4 | X1 +X2 +X4, data = jan1))

Call:
zeroinfl(formula = Y ~ X1 + X2 + X4 | X1 + X2 + X4, data = jan1)

Pearson residuals:
    Min       1Q  Median      3Q     Max 
-1.0741 -0.6571 -0.5024  0.6375  2.0496 

Count model coefficients (poisson with log link):
              Estimate Std. Error z value Pr(>|z|)    
(Intercept)  2.8559158  0.7401966   3.858 0.000114 ***
X1          -0.0014311  0.0009906  -1.445 0.148535    
X2          -0.0271854  0.0318003  -0.855 0.392618    
X4          -0.0106202  0.0059996  -1.770 0.076702 .  

Zero-inflation model coefficients (binomial with logit link):
              Estimate Std. Error z value Pr(>|z|)  
(Intercept) -4.6012478  2.5783645  -1.785   0.0743 .
X1          -0.0004124  0.0028221  -0.146   0.8838  
X2           0.0631369  0.1088073   0.580   0.5617  
X4           0.0439303  0.0202973   2.164   0.0304 *
```

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

Number of iterations in BFGS optimization: 16
Log-likelihood: -100.3 on 8 Df
```

Type equation here.

Program R untuk regresi ZIP dengan tiga variabel bebas X_2, X_3, dan X_4

```
> summary(zeroinfl(Y ~ X2 +X3 +X4 | X2 +X3 +X4, data = jan1))

Call:
zeroinfl(formula = Y ~ X2 + X3 + X4 | X2 + X3 + X4, data = jan1)

Pearson residuals:
     Min       1Q   Median       3Q      Max 
-1.61356 -0.23263 -0.06978  0.09344  3.82681 

Count model coefficients (poisson with log link):
             Estimate Std. Error z value Pr(>|z|)   
(Intercept)  1.500382   0.572846   2.619  0.00881 **
X2          -0.043244   0.032251  -1.341  0.17997   
X3           0.033853   0.013284   2.548  0.01082 * 
X4          -0.004714   0.005218  -0.904  0.36625   

Zero-inflation model coefficients (binomial with logit link):
            Estimate Std. Error z value Pr(>|z|)  
(Intercept) -5.77760    8.31886  -0.695   0.4874  
X2           0.09062    0.31462   0.288   0.7733  
X3          -0.81503    0.34672  -2.351   0.0187 *
X4           0.12056    0.08737   1.380   0.1677  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Number of iterations in BFGS optimization: 18
Log-likelihood: -67.52 on 8 Df
```

Program R untuk regresi ZIP dengan tiga variabel bebas X_1, X_3, dan X_4

```
> summary(zeroinfl(Y ~ X1 +X3 +X4 | X1 +X3 +X4, data = jan1))

Call:
zeroinfl(formula = Y ~ X1 + X3 + X4 | X1 + X3 + X4, data = jan1)

Pearson residuals:
       Min         1Q     Median         3Q        Max 
-1.764e+00 -1.472e-01 -1.245e-08 -1.205e-08  2.421e+00 

Count model coefficients (poisson with log link):
              Estimate Std. Error z value Pr(>|z|)  
(Intercept)  1.3925955  0.7941836   1.753   0.0795 .
X1          -0.0001230  0.0009496  -0.130   0.8970  
X3           0.0305248  0.0127420   2.396   0.0166 *
X4          -0.0064362  0.0060179  -1.070   0.2848  

Zero-inflation model coefficients (binomial with logit link):
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -265.75854  294.40218  -0.903    0.367
X1             0.09628    0.11752   0.819    0.413
X3           -28.74401   31.72699  -0.906    0.365
X4             4.76248    5.25448   0.906    0.365
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Number of iterations in BFGS optimization: 356
Log-likelihood: -63.98 on 8 Df
```

I. Program R untuk regresi ZIP Null Model

```
> summary(zeroinfl(Y ~ NULL , data = jan1))

Call:
zeroinfl(formula = Y ~ NULL, data = jan1)

Pearson residuals:
    Min      1Q  Median      3Q     Max 
-0.7960 -0.7960 -0.7960  0.8837  2.5634 

Count model coefficients (poisson with log link):
            Estimate Std. Error z value Pr(>|z|)    
(Intercept)  1.35774    0.09258   14.66   <2e-16 ***

Zero-inflation model coefficients (binomial with logit link):
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.04958    0.24979   0.198    0.843
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 

Number of iterations in BFGS optimization: 8 
Log-likelihood: -106.3 on 2 Df
```

J. Pembuktian Mendapatkan Mean dan Varian Distribusi ZIP

Zero-inflated Poisson adalah model campuran yang sederhana untuk data diskrit dengan banyak peristiwa nol. Dalam ZIP terdapat banyak penelitian nol yang dikembangkan dalam dua langkah, yaitu:

$$Y_i \sim \begin{cases} 0, & \text{dengan Peluang [illegible]} \\ Poisson\ (\mu_i), & \text{dengan Peluang } (1 - \text{[illegible]} \end{cases} \qquad (2.3)$$

dimana fungsi kepadatan peluang Poisson adalah sebagai berikut:

$$\mathrm{P}(Y_i = y_i) = \begin{cases} \frac{e^{-\lambda}\lambda^y}{y_i!}, & \text{untuk [illegible]} 0, 1, 2, 3, \ldots \\ 0, & \text{untuk yang lainnya} \end{cases}$$

Karena ZIP merupakan model campuran untuk data diskrit yang terdapat banyak nilai nol dan terdiri dari dua parameter inflasi nol dan Poisson sehingga untuk fungsi kepadatan peluangnya diperoleh dari fungsi kepadatan peluang Poisson,yaitu:

a. Untuk $y = 0$ dengan parameter w dan termasuk kedalam fungsi kepadatan peluang Poisson sehingga FKP ZIP adalah $w + f(y = 0) = w + \frac{(1-w)e^{-\lambda}\lambda^0}{0!} = w + (1 - w)e^{-\lambda}$

b. Untuk y selain nol yaitu dengan parameter $(1 - w)$, FKP = $(1 - w)\frac{e^{-\lambda}\lambda^y}{y}$

Sehingga FKP untuk model regresi ZIP menurut Lambert (1992) sebagai berikut:

$$P(Y_i = y_i) = \begin{cases} w + (1-w)e^{-\lambda}, & \text{untuk } y_i = 0 \\ (1-w)\dfrac{e^{-\lambda}\lambda^{y}}{y_i!}, & \text{untuk } y_i = 1,2,\ldots \end{cases}$$

Dari FKP ZIP dapat diperoleh mean dan variannya sebagai berikut:

$$E(Y_i) = \sum_{y=0}^{n} y_i[I_{y=0}\,.\left(w + (1-w)e^{-\lambda}\right) + I_{y>0}.\,(1-w)\frac{e^{-\lambda}\lambda^{y}}{y!}]$$

$$E(Y_i) = \sum_{y=1}^{n} y_i\,(1-w)\frac{e^{-\lambda}\lambda^{y}}{y!}$$

$$= \sum_{y=1}^{n} \frac{(1-w)e^{-\lambda}\lambda^{y}}{(y-1)!}$$

$$= \sum_{y=1}^{n} \frac{(1-w)e^{-\lambda}\lambda^{(y-1)}\lambda}{(y-1)!}$$

$$= \lambda(1-w)\sum_{y=1}^{n} \frac{e^{-\lambda}\lambda^{(y-1)}}{(y-1)!}$$

$$= \lambda(1-w)$$

$$= \mu_i$$

$$E\left(Y^{\;2}\right) = E(y_i^{\,2}) - E(y_i) + E(y_i)$$

$$= E\big(y(y-1)\big) + E(y)$$

$$= \sum_{y=0}^{n}\big(y(y-1)\big)\,(1-w)\frac{e^{-\lambda}\lambda^{y}}{y!} + \lambda(1-w)$$

$$= \sum_{y=0}^{n}\big(y(y-1)\big)\,(1-w)\frac{e^{-\lambda}\lambda^{y}}{y(y-1)(y-2)!} + \lambda(1-w)$$

$$= \sum_{y=0}^{n}(1-w)\frac{e^{-\lambda}\lambda^{2}\,\lambda^{(y-2)}}{(y-2)!} + \lambda(1-w)$$

$$= (1-w)\lambda^{2} + \lambda(1-w)$$

$$= (1-w)(\lambda^{2} + \lambda)$$

$$Var(Y_i) = E(Y_i^2) - \left(E(Y_i^2)\right)^2$$
$$= (1-w)(\lambda+\lambda^2) - \left(\lambda(1-w)\right)^2$$
$$= (1-w)(\lambda+\lambda^2) - \lambda^2(1-w)^2$$
$$= (1-w)\,\lambda\,(\lambda+1) - \lambda(1-w)\,\lambda(1-w)$$
$$= (1-w)\,\lambda\left(\lambda+1-\lambda(1-w)\right)$$
$$= (1-w)\,\lambda(\lambda+1-\lambda+\lambda w)$$
$$= (1-w)\,\lambda(1+\lambda w)$$
$$= \mu_i(1+\lambda w)$$
$$= \mu_i\left(1+\frac{\mu_i}{(1-w)}w\right)$$
$$= \mu_i\left(1+\frac{w}{(1-w)}\mu_i\right)$$
$$= \mu_i + \frac{w}{(1-w)}\mu_i^2$$

Dapat diperoleh mean $= \mu_i$ dan varian $= \mu_i + (\frac{w}{1-w})\mu_i^{\,2}$ sehingga dapat dilihat bahwa pada distribusi ZIP terjadi overdispersi untuk $w > 0$ karena varian > mean.

K. *Fitted Value* untuk Model ZIP dan Model Poisson

Berikut adalah nilai *fitted value* untuk model regresi terbaik ZIP dan model regresi Poisson sebagai pembandingnya.

Fitted Value Model Regresi ZIP:

```
> fitted.values(zeroinfl(Y ~ X1 +X3 +X4 | X1 +X3 +X4, data = jan1))
           1            2            3            4            5            6 
3.381733e+00 2.942053e+00 2.913621e+00 5.571757e+00 2.991057e+00 3.481960e+00 
           7            8            9           10           11           12 
4.028849e+00 5.583735e-16 4.618817e+00 5.543788e+00 4.029045e+00 2.916952e+00 
          13           14           15           16           17           18 
4.407809e-16 4.947458e-16 5.538379e-16 4.407038e-16 5.131161e-16 4.787644e-16 
          19           20           21           22           23           24 
5.431966e-16 3.108219e+00 2.657482e+00 5.132085e-16 5.307717e-16 2.616725e+00 
          25           26           27           28           29           30 
3.550726e+00 3.599750e+00 3.065693e+00 3.586513e+00 4.885344e+00 2.590404e+00 
          31           32           33           34           35           36 
2.664003e+00 3.088154e+00 3.096522e+00 3.546627e+00 4.735387e-16 4.816017e-16 
          37           38           39           40           41           42 
4.702078e-16 4.813126e-16 7.449367e-04 4.134321e-16 4.656612e-16 4.627411e-16 
          43           44           45           46           47           48 
```

```
4.240707e-16 5.108774e-16 4.382407e-16 4.160720e-16 5.758792e-16 4.957488e-16
          49           50           51           52           53           54
5.687516e-16 3.112565e+00 5.319345e-04 5.513753e-16 5.047976e-16 1.791254e-02
          55           56           57           58           59           60
3.244778e-02 6.591870e-07 5.031773e-16 5.484247e-16 5.467609e+00 5.517543e+00

   61           62           63           64           65           66
6.324714e+00 4.527171e+00 4.581702e+00 3.398959e+00 3.411768e+00 2.732478e+00
          67
3.429390e+00
```

Fitted Value Model Regresi Poisson:

```
> fitted.values(glm(Y ~ X1 +X3 +X4,family=poisson, data=jan1))
         1          2          3          4          5          6          7
 2.6315770  1.4564406  1.6026023  6.9583426  1.2377786  1.9737950  3.3417726
         8          9         10         11         12         13         14
 0.9593403  6.1960828  7.3118772  3.3401743  1.5846691  0.4171338  0.6432795
        15         16         17         18         19         20         21
 1.0184418  0.4178538  0.6652671  0.8888573  0.8325092  1.8215320  1.1964337
        22         23         24         25         26         27         28
 0.6640878  0.7060443  1.1229442  2.8309628  2.4733650  1.6885187  2.5647451
        29         30         31         32         33         34         35
 6.2008934  1.1634187  0.9452736  1.5713945  1.5300712  2.8633466  0.5412524
        36         37         38         39         40         41         42
 0.7065601  0.6039869  0.7107508  0.9813494  0.4459585  0.4488106  0.4774927
        43         44         45         46         47         48         49
 0.3472434  0.5851732  0.3720342  0.4188636  1.0070021  0.6183291  1.1383983
        50         51         52         53         54         55         56
 1.1896540  0.7622663  0.7046208  0.5174385  0.7292630  1.1391348  0.8291361
        57         58         59         60         61         62         63
 0.5340830  0.7428547  5.9536379  5.4437889 10.1059717  5.3629993  4.7665025
        64         65         66         67
 1.7785359  1.7138619  0.8731404  1.6290786
```

Dari nilai fitted value yang diperoleh dapat diketahui nilai residu untuk masing-masing data pada variabel takbebas (Y) yang terlampir pada Lampiran A yaitu dengan rumus residu $= Y - \hat{Y}$, dimana $\hat{Y}$ merupakan nilai fitted values untuk masing-masing data, residu dapat dirangkum pada tabel dibawah ini:

No	Residu ZIP	Residu Poisson
1	-1.381733	-0.6315770
2	$5.794680\times 10^{\substack{33\\-2}}$	1.5435594
3	2.086379	3.3973977

4	-1.571757	-2.9583426
5	1.008943	2.7622214
6	4.518040	6.0262050
7	-2.028849	-1.3417726
8	-5.583735×10^{-16}	-0.9593403
9	-6.188166×10^{-1}	-2.1960828
10	4.562119×10^{-1}	-1.3118772
11	1.970955	2.6598257
12	8.304796×10^{-2}	1.4153309
13	-4.407809×10^{-16}	-0.4171338
14	-4.947458×10^{-16}	-0.6432795
15	-5.538379×10^{-16}	-1.0184418
16	-4.407038×10^{-16}	-0.4178538
17	-5.131161×10^{-16}	-0.6652671
18	-4.787644×10^{-16}	-0.8888573
19	-5.431966×10^{-16}	-0.8325092
20	2.891781	4.1784680
21	-2.657482	-1.1964337
22	-5.132085×10^{-16}	-0.6640878
23	-5.307717×10^{-16}	-0.7060443
24	-6.167251×10^{-1}	0.8770558
25	-5.507256×10^{-1}	0.1690372
26	4.002498×10^{-1}	1.5266350
27	-6.569279×10^{-2}	1.3114813
28	-5.865130×10^{-1}	0.4352549
29	1.146557×10^{-1}	-1.2008934
30	-5.904037×10^{-1}	0.8365813

31	-6.640029×10^{-1}	1.0547264
32	9.118459×10^{-1}	2.4286055
33	-9.652228×10^{-2}	1.4699288
34	4.533732×10^{-1}	1.1366534
35	-4.735387×10^{-16}	-0.5412524
36	-4.816017×10^{-16}	-0.7065601
37	-4.702078×10^{-16}	-0.6039869
38	-4.813126×10^{-16}	-0.7107508
39	-7.449367×10^{-4}	-0.9813494
40	-4.134321×10^{-16}	-0.4459585
41	-4.656612×10^{-16}	-0.4488106
42	-4.627411×10^{-16}	-0.4774927
43	-4.240707×10^{-16}	-0.3472434
45	-5.108774×10^{-16}	-0.5851732
46	-4.382407×10^{-16}	-0.3720342
47	-4.160720×10^{-16}	-0.4188636
48	-5.758792×10^{-16}	-1.0070021
49	-4.957488×10^{-16}	-0.6183291
50	-5.687516×10^{-16}	-1.1383983
51	-3.112565	-1.1896540
52	-5.319345×10^{-4}	-0.7622663
53	-5.513753×10^{-16}	-0.7046208
54	-5.047976×10^{-16}	-0.5174385
55	-1.791254×10^{-2}	-0.7292630
56	-3.244778×10^{-2}	-1.1391348
57	-6.591870×10^{-7}	-0.8291361
58	-5.031773×10^{-16}	-0.5340830

59	$-5.484247\times 10_{-16}$	-0.7428547
60	$-4.676093\times 10^{\substack{16\\-1}}$	-0.9536379
61	$-5.175426\times 10^{\substack{-1\\-1}}$	-0.4437889
62	1.675286	-2.1059717
63	$-5.271705\times 10^{\substack{86\\-1}}$	-1.3629993
64	$-5.817016\times 10^{\substack{-1\\-1}}$	-0.7665025
65	$-3.989587\times 10^{\substack{-1\\-1}}$	1.2214641
66	-1.411768	0.2861381
67	1.267522	3.1268596