

SKRIPSI

**ANALISIS FREKUENSI KLAIM ASURANSI KENDARAAN
BERMOTOR MENGGUNAKAN MODEL REGRESI
POISSON-LOGNORMAL**



ANGELA GRACIA

NPM: 6162001148

**PROGRAM STUDI MATEMATIKA
FAKULTAS TEKNOLOGI INFORMASI DAN SAINS
UNIVERSITAS KATOLIK PARAHYANGAN
2024**

FINAL PROJECT

**ANALYZING THE FREQUENCY OF MOTOR VEHICLE
INSURANCE CLAIMS USING THE POISSON-LOGNORMAL
REGRESSION MODEL**



ANGELA GRACIA

NPM: 6162001148

**DEPARTMENT OF MATHEMATICS
FACULTY OF INFORMATION TECHNOLOGY AND SCIENCES
PARAHYANGAN CATHOLIC UNIVERSITY
2024**

LEMBAR PENGESAHAN

ANALISIS FREKUENSI KLAIM ASURANSI KENDARAAN BERMOTOR MENGGUNAKAN MODEL REGRESI POISSON-LOGNORMAL

Angela Gracia

NPM: 6162001148

Telah lulus ujian skripsi pada 23 Januari 2024 dengan penguji:
Dr. Erwinna Chendra dan Farah Kristiani, Ph.D.

Bandung, 2 Februari 2024

Menyetujui,

Pembimbing 1

Pembimbing 2

Liem Chin, M.Si.

Rizky Reza Fauzi, D.Phil.Math.

Mengetahui,

Ketua Program Studi

Jonathan Hoseana, Ph.D.

PERNYATAAN

Dengan ini saya yang bertandatangan di bawah ini menyatakan bahwa skripsi dengan judul:

ANALISIS FREKUENSI KLAIM ASURANSI KENDARAAN BERMOTOR MENGUNAKAN MODEL REGRESI POISSON-LOGNORMAL

adalah benar-benar karya saya sendiri, dan saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan.

Atas pernyataan ini, saya siap menanggung segala risiko dan sanksi yang dijatuhkan kepada saya, apabila di kemudian hari ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya, atau jika ada tuntutan formal atau non-formal dari pihak lain berkaitan dengan keaslian karya saya ini.

Dinyatakan di Bandung,
2 Februari 2024



Angela Gracia
NPM: 6162001148

ABSTRAK

Perusahaan asuransi kendaraan bermotor perlu melakukan prediksi terhadap frekuensi klaim yang mungkin terjadi di masa mendatang untuk menetapkan harga premi yang sesuai untuk polis asuransi. Akan tetapi, seringkali terjadi overdispersi pada data asuransi kendaraan bermotor, yaitu kondisi di mana nilai variansi lebih besar dibanding rata-ratanya, akibat faktor-faktor tersembunyi yang tidak tercatat dalam data, seperti kemampuan dan pengalaman pengendara, yang tidak dapat direpresentasikan oleh distribusi pada umumnya. Oleh karena itu, penelitian ini bertujuan menyajikan model regresi Poisson-lognormal untuk memprediksi frekuensi klaim asuransi kendaraan bermotor, karena model ini cocok digunakan untuk menggambarkan data overdispersi. Penaksiran parameter model regresi Poisson-lognormal akan dilakukan melalui algoritma ekspektasi-maksimasi. Setelah diperoleh parameter model, dilakukan analisis terhadap pengaruh variabel-variabel yang terlibat dalam penelitian terhadap frekuensi klaim asuransi kendaraan bermotor. Hasil analisis menunjukkan bahwa penambahan umur, pemegang polis laki-laki, penggunaan mobil untuk keperluan pribadi di daerah pedesaan memiliki kecenderungan untuk mengajukan klaim asuransi dengan frekuensi yang lebih rendah. Sementara itu, penambahan umur, pemegang polis laki-laki, penggunaan mobil untuk keperluan komersial di daerah perkotaan cenderung memiliki tingkat variasi klaim yang lebih rendah. Model regresi yang dihasilkan diharapkan dapat memberikan manfaat signifikan bagi perusahaan asuransi dalam melakukan perhitungan prediktif terkait frekuensi klaim asuransi kendaraan bermotor berdasarkan karakteristik individu dan penggunaan kendaraan yang dimiliki oleh pemegang polis.

Kata-kata kunci: Poisson-lognormal; lognormal; asuransi kendaraan bermotor; frekuensi; algoritma ekspektasi-maksimasi.

ABSTRACT

An insurance company for motor vehicles needs to predict the frequency of claims that may occur in the future to set appropriate premium prices for insurance policies. However, there is often overdispersion in motor vehicle insurance data, where the variance is greater than the mean, due to hidden factors not recorded in the data, such as drivers' abilities and experiences, which cannot be represented by typical distributions. Therefore, this study aims to present a Poisson-lognormal regression model to predict the frequency of motor vehicle insurance claims, as this model is suitable for describing overdispersed data. Parameter estimation of the Poisson-lognormal regression model will be carried out using the expectation-maximization algorithm. After obtaining the model parameters, an analysis will be conducted on the effects of the variables involved in the study on the frequency of motor vehicle insurance claims. The analysis results show that increasing age, male policyholders, and using cars for personal purposes in rural areas tend to file insurance claims with lower frequencies. Meanwhile, increasing age, male policyholders, and using cars for commercial purposes in urban areas tend to have lower claim variation rates. The resulting regression model is expected to provide significant benefits to insurance companies in predicting insurance claim frequencies for motor vehicles based on individual characteristics and the use of vehicles held by policyholders.

Keywords: Poisson-lognormal; lognormal; motor vehicle insurance; frequency; expectation-maximization algorithm.

I have fought the good fight, I have finished the race, I have kept the faith.



KATA PENGANTAR

Puji syukur kepada Tuhan Yang Maha Esa atas berkat dan pertolongan-Nya sehingga penulis dapat menyelesaikan skripsi ini dengan tepat waktu. Skripsi dengan judul "Analisis Frekuensi Klaim Asuransi Kendaraan Bermotor Menggunakan Model Regresi Poisson-Lognormal" ini disusun guna memenuhi salah satu syarat kelulusan dari Program Studi Matematika, Fakultas Teknologi Informasi dan Sains, Universitas Katolik Parahyangan, Bandung. Dalam penyusunan skripsi ini, penulis mendapat banyak bantuan, ilmu, dan dukungan dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan ucapan terima kasih kepada:

1. Kedua orang tua, kakek, nenek, kakak perempuan, dan seluruh keluarga besar yang selalu mendukung penulis dalam segala keadaan dan mendoakan penulis.
2. Bapak Liem Chin, M.Si., dan Bapak Rizky Reza Fauzi, D.Phil.Math., selaku dosen pembimbing yang senantiasa membimbing, memberikan ilmu, saran, bantuan, dan arahan selama perkuliahan dan penyusunan skripsi ini.
3. Ibu Dr. Erwinna Chendra dan Ibu Farah Kristiani, Ph.D., selaku dosen penguji yang memberikan kritik dan saran untuk pengembangan skripsi ini.
4. Bapak Dr. Daniel Salim selaku dosen koordinator skripsi yang memberikan informasi, panduan, dan saran selama pembuatan skripsi ini.
5. Bapak Taufik Limansyah, M.T., selaku dosen wali penulis yang memberikan informasi dan dukungan selama perkuliahan.
6. Seluruh dosen Matematika UNPAR yang telah memberikan ilmu yang berguna kepada penulis.
7. Seluruh staf Tata Usaha FTIS yang membantu penulis dalam mengumpulkan syarat kelengkapan wisuda dan semua proses kelengkapan data dalam perkuliahan penulis.
8. Teman-teman penulis selama masa perkuliahan, yaitu Cella, Lia, Catherine, dan Wristopher, yang senantiasa memberikan ilmu, membantu, menyemangati, dan menemani penulis dalam pembuatan skripsi ini.
9. Teman-teman Matematika UNPAR angkatan 2020.
10. Dan semua pihak yang tidak dapat disebutkan satu-satu.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna. Oleh karena itu, kritik dan saran yang membangun selalu penulis harapkan. Penulis berharap agar skripsi ini dapat bermanfaat bagi para pembaca.

Bandung, 2 Februari 2024

Penulis

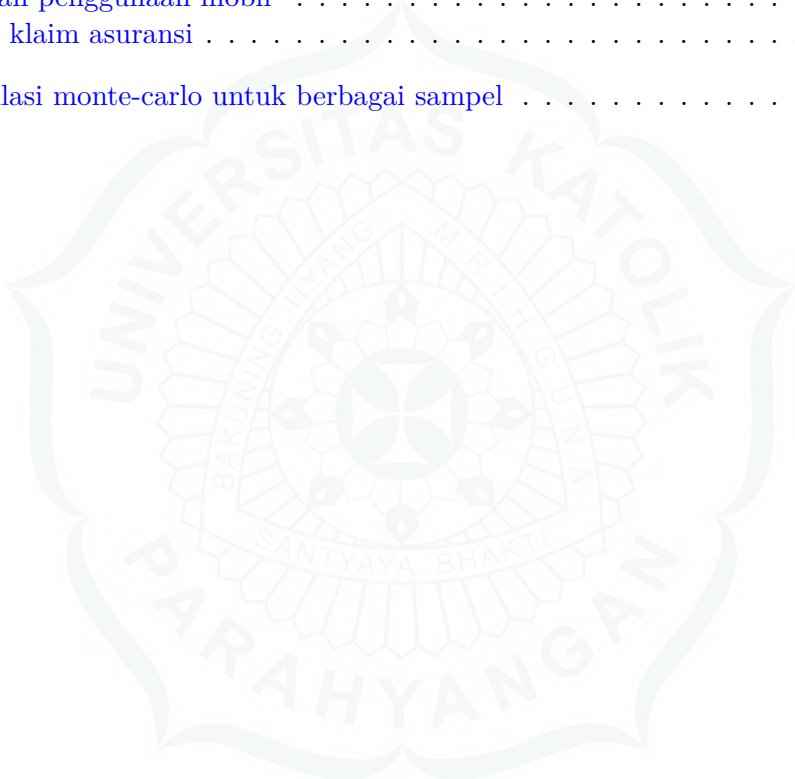
DAFTAR ISI

KATA PENGANTAR	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL	xii
DAFTAR NOTASI	xiii
1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	3
1.4 <i>State of the Art</i>	3
1.5 Sistematika Pembahasan	3
2 LANDASAN TEORI	5
2.1 Distribusi Poisson	5
2.2 Distribusi Lognormal	6
2.3 Distribusi Campuran (<i>Mixture Distribution</i>)	6
2.4 Distribusi Campuran Poisson (<i>Poisson Mixture Distribution</i>)	7
2.5 Algoritma Ekspektasi-Maksimasi	8
2.6 Simulasi Monte-Carlo	9
2.7 Metode Newton-Raphson	10
3 MODEL REGRESI POISSON-LOGNORMAL	12
3.1 Model Regresi Poisson-Lognormal	12
3.2 Algoritma Ekspektasi-Maksimasi untuk Model Regresi Poisson-Lognormal dengan Dispersi yang Berbeda	14
4 ANALISIS DATA FREKUENSI KLAIM ASURANSI KENDARAAN BERMOTOR	21
4.1 Analisis Data Eksploratif	21
4.2 Pemilihan Variabel Bebas dan Variabel Tidak Bebas	27
4.3 Hasil dan Interpretasi Model Regresi	29
4.4 Contoh Prediksi	33
5 KESIMPULAN DAN SARAN	37
5.1 Kesimpulan	37
5.2 Saran	38
DAFTAR REFERENSI	39
A PENURUNAN RUMUS PERSAMAAN (3.31)	40



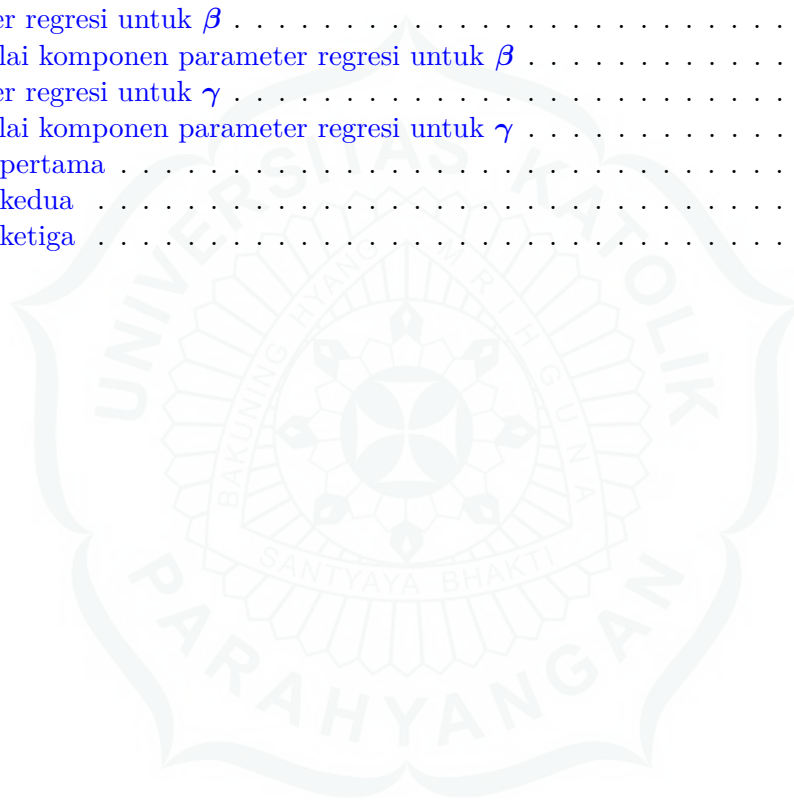
DAFTAR GAMBAR

- 3.1 *Flow chart* algoritma ekspektasi-maksimasi 15
- 4.1 Jenis kelamin pemegang polis 23
- 4.2 Penggunaan mobil pemegang polis 23
- 4.3 Tipe mobil pemegang polis 24
- 4.4 Lingkungan penggunaan mobil 25
- 4.5 Frekuensi klaim asuransi 26
- B.1 Plot simulasi monte-carlo untuk berbagai sampel 42



DAFTAR TABEL

4.1	Informasi variabel pada data	21
4.2	Usia pemegang polis	22
4.3	Frekuensi klaim	25
4.4	Total besar klaim	26
4.5	Nilai variabel bebas pada pemrograman	29
4.6	Parameter regresi untuk β	29
4.7	Selisih nilai komponen parameter regresi untuk β	29
4.8	Parameter regresi untuk γ	30
4.9	Selisih nilai komponen parameter regresi untuk γ	30
4.10	Simulasi pertama	33
4.11	Simulasi kedua	34
4.12	Simulasi ketiga	35



DAFTAR NOTASI

$H_1(\beta)$	Turunan parsial kedua dari fungsi Q terhadap β
$H_2(\gamma)$	Turunan parsial kedua dari fungsi Q terhadap γ
$P(k_i z_i)$	Fungsi massa peluang dari $k_i z_i$
$P(k_i)$	Fungsi massa peluang dari frekuensi klaim pemegang polis ke- i
Z_i	Faktor-faktor tersembunyi yang tidak tercatat pada data perusahaan asuransi pada pemegang polis ke- i yang diasumsikan mengikuti distribusi lognormal
\mathbf{K}	Vektor yang berisi data frekuensi klaim
β	Parameter regresi dari rata-rata frekuensi klaim
γ	Parameter regresi dari dispersi frekuensi klaim
$\theta^{(r)}$	Vektor yang terdiri dari β dan γ pada iterasi ke- i
\mathbf{x}_i	Vektor variabel bebas yang memengaruhi rata-rata frekuensi klaim pemegang polis ke- i
\mathbf{y}_i	Vektor variabel bebas yang memengaruhi dispersi frekuensi klaim pemegang polis ke- i
$E(z_i)$	Ekspektasi dari Z_i
$E(k_i z_i)$	Ekspektasi dari $k_i z_i$
$E(k_i)$	Ekspektasi dari frekuensi klaim pemegang polis ke- i
$\text{Var}(z_i)$	Variansi dari Z_i
$\text{Var}(k_i z_i)$	Variansi dari $k_i z_i$
$\text{Var}(k_i)$	Variansi dari frekuensi klaim pemegang polis ke- i
μ_i	Rata-rata dari frekuensi klaim pemegang polis ke- i
ϕ_i	Dispersi dari frekuensi klaim pemegang polis ke- i
a_{i_m}	Pembilang dari w_{1_i} yang berisi sampel acak z_{i_m}
b_{i_m}	Pembilang dari w_{2_i} yang berisi sampel acak z_{i_m}
c_{i_m}	Penyebut dari w_{1_i} dan w_{2_i} yang berisi sampel acak z_{i_m}
$g(z_i; \phi_i)$	Fungsi kepadatan peluang dari Z_i
$h_1(\beta)$	Turunan parsial pertama dari fungsi Q terhadap β
$h_2(\gamma)$	Turunan parsial pertama dari fungsi Q terhadap γ
k_i	Frekuensi klaim pemegang polis ke- i
$l_c(\theta)$	Fungsi <i>log-likelihood</i> data lengkap
m	Banyak sampel acak yang dibangkitkan pada simulasi Monte-Carlo untuk setiap i
n	Banyak pemegang polis pada data
w_{1_i}	Ekspektasi dari z_i bersyarat k_i dan $\theta^{(r)}$
w_{2_i}	Ekspektasi dari $(\log(z_i))^2$ bersyarat k_i dan $\theta^{(r)}$
z_{i_m}	Sampel acak yang dibangkitkan pada simulasi Monte-Carlo untuk setiap i

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Mobilitas merupakan aspek yang sangat penting bagi kehidupan manusia, di mana kendaraan bermotor menjadi salah satu sarana penunjang aktivitas tersebut. Oleh karena itu, tidak heran apabila berdasarkan catatan Kepolisian Republik Indonesia (Polri) tertulis bahwa jumlah kendaraan bermotor di negeri ini mencapai 152,51 juta unit pada tahun 2022.¹ Akan tetapi, tidak dapat dipungkiri bahwa manusia sering dihadapkan dengan berbagai peristiwa tidak terduga, salah satunya adalah kecelakaan lalu lintas. Berdasarkan data yang dikeluarkan oleh Kementerian Perhubungan, jumlah korban kecelakaan lalu lintas di Indonesia pada tahun 2022 mencapai 204,4 ribu orang. Angka ini mengalami peningkatan sebesar 33% dibandingkan dengan tahun sebelumnya.² Oleh sebab itu, berbagai upaya antisipasi kerap dilakukan untuk mengurangi dampak kerugian akibat peristiwa tersebut. Salah satu cara yang dapat dilakukan adalah mengalihkan risiko ke perusahaan asuransi kendaraan bermotor.

Asuransi kendaraan bermotor adalah produk asuransi yang menawarkan perlindungan kepada pemegang polis dari risiko kerugian sehubungan dengan kepemilikan dan pemakaian kendaraan bermotor.³ Dalam perhitungan premi asuransi, perusahaan harus dapat memperkirakan banyak klaim yang mungkin terjadi berdasarkan data masa lampau. Frekuensi klaim asuransi sering diasumsikan mengikuti distribusi Poisson karena distribusi tersebut dapat menggambarkan banyaknya kejadian dalam suatu interval waktu tertentu. Pengembangan dari distribusi Poisson ini adalah regresi Poisson [1] yang dapat memprediksi banyaknya klaim yang mungkin terjadi untuk suatu nilai variabel bebas tertentu.

Pada model regresi Poisson, terdapat salah satu asumsi yang harus dipenuhi, yaitu kesamaan nilai rata-rata dan variansi. Akan tetapi, nilai variansi pada data klaim asuransi kendaraan bermotor kerap ditemukan lebih besar dibanding nilai rata-ratanya. Hal ini menunjukkan terjadinya overdispersi, yaitu kondisi di mana nilai variansinya lebih besar dibanding nilai rata-ratanya.⁴ Overdispersi ini dapat disebabkan oleh faktor-faktor tersembunyi (*hidden factors*) yang tidak tercatat dalam data asuransi kendaraan bermotor, seperti kemampuan dan pengalaman berkendara pemegang polis.

¹<https://dataindonesia.id/sektor-riil/detail/polri-catat-15251-juta-kendaraan-di-indonesia-pada-2022>, diakses pada tanggal 1 April 2023.

²<https://databoks.katadata.co.id/datapublish/2023/09/15/jumlah-korban-kecelakaan-lalu-lintas-pada-2022-meningkat-tajam-ini-trennya>, diakses pada tanggal 31 Desember 2023.

³<https://www.cermati.com/artikel/asuransi-motor>, diakses pada tanggal 3 April 2023.

⁴<https://www.neliti.com/id/publications/359227/model-model-regresi-untuk-mengatasi-masalah-overdispersi-pada-regresi-poisson>, diakses pada tanggal 9 Desember 2023.

Overdispersi pada data dapat menyebabkan data tersebut tidak cocok untuk dianalisis dengan regresi Poisson. Penggunaan regresi ini dapat menyebabkan kekeliruan interpretasi model karena estimasi *standard error* yang terlalu rendah dan signifikansi parameter regresi yang terlalu tinggi [2]. Oleh karena itu, diperlukan metode lain untuk menggambarkan frekuensi klaim yang memperhitungkan keacakan faktor-faktor tersembunyi. Salah satu cara yang dapat digunakan adalah dengan menggunakan distribusi campuran Poisson [3], yang menggabungkan distribusi Poisson dengan suatu distribusi lain yang dapat mencerminkan faktor-faktor tersembunyi tersebut.

Konsep regresi Poisson dan distribusi campuran Poisson dapat digabungkan menjadi regresi campuran Poisson [4] yang dapat digunakan untuk memprediksi frekuensi klaim yang mungkin diajukan dari pemegang polis. Model regresi campuran Poisson ini baik digunakan untuk data asuransi kendaraan bermotor karena dapat menggambarkan jumlah klaim overdispersi yang timbul akibat faktor-faktor risiko yang sulit untuk diamati. Pada umumnya, model regresi campuran Poisson yang digunakan adalah model regresi Poisson-gamma [5], di mana menurut [6], model tersebut ekuivalen dengan model regresi binomial negatif. Akan tetapi, dikarenakan sifat distribusi gamma yang sulit untuk memodelkan nilai-nilai yang relatif ekstrem [3], model regresi Poisson-gamma kurang akurat digunakan apabila terdapat faktor-faktor tersembunyi yang bernilai ekstrem. Oleh karena itu, dalam penelitian ini, akan dibentuk model regresi Poisson-lognormal (PLN) yang diharapkan dapat menggambarkan faktor-faktor tersembunyi bernilai ekstrem secara lebih akurat. Hal tersebut sesuai dengan sifat distribusi lognormal yang sering digunakan untuk menganalisis nilai-nilai ekstrem [3].

Pada dasarnya, berbagai faktor tersembunyi yang tidak tercatat pada data menyulitkan penentuan penaksiran parameter apabila digunakan teknik penaksiran biasa, seperti estimasi maksimum *likelihood* (MLE). Masalah ini dapat diatasi jika algoritma ekspektasi-maksimasi (EM) diterapkan pada penaksiran maksimum *likelihood* [6]. Algoritma EM merupakan teknik iteratif yang terdiri dari tahap ekspektasi (*E step*) dan maksimisasi (*M step*) untuk menentukan estimasi kemungkinan maksimum atau *maximum likelihood estimation* (MLE) untuk *missing data*. Oleh karena itu, metode ini cocok untuk digunakan pada asuransi kendaraan bermotor yang cenderung memiliki faktor yang tersembunyi.

Dengan demikian, tujuan dari skripsi ini adalah untuk menganalisis dan memprediksi frekuensi klaim asuransi kendaraan bermotor menggunakan model regresi Poisson-lognormal. Untuk mengatasi adanya faktor-faktor tersembunyi yang tidak tercatat dalam data perusahaan asuransi kendaraan bermotor, penaksiran parameter pada model akan dilakukan melalui penggunaan algoritma ekspektasi-maksimasi.

1.2 Rumusan Masalah

Berdasarkan penjabaran pada latar belakang permasalahan, maka rumusan masalah pada penelitian ini adalah:

1. Bagaimana cara membangun model regresi Poisson-lognormal berdasarkan konsep distribusi campuran Poisson?
2. Bagaimana cara kerja algoritma ekspektasi-maksimasi dalam penaksiran parameter model regresi Poisson-lognormal?

3. Bagaimana pengaruh variabel umur, jenis kelamin, tujuan penggunaan mobil, dan area tempat mobil digunakan terhadap frekuensi klaim asuransi kendaraan bermotor?

1.3 Tujuan

Tujuan yang ingin dicapai dari penelitian ini adalah:

1. Membangun model regresi Poisson-lognormal berdasarkan konsep distribusi campuran Poisson.
2. Mengevaluasi cara kerja algoritma ekspektasi-maksimasi dalam penaksiran parameter model regresi Poisson-lognormal.
3. Menganalisis pengaruh variabel umur, jenis kelamin, tujuan penggunaan mobil, dan area tempat mobil digunakan terhadap frekuensi klaim asuransi kendaraan bermotor.

1.4 *State of the Art*

Model regresi campuran Poisson memiliki keunggulan dalam meramalkan frekuensi klaim asuransi kendaraan bermotor. Keunggulan ini muncul karena nilai variansi pada distribusi campuran Poisson selalu lebih besar daripada nilai rata-ratanya. Hal ini membuat model ini lebih sesuai untuk memodelkan frekuensi klaim asuransi kendaraan bermotor yang seringkali bervariasi dikarenakan faktor-faktor tersembunyi yang tidak tercatat dalam data asuransi. Model yang umum digunakan untuk mengatasi kasus di mana nilai variansi lebih besar dibanding rata-ratanya adalah regresi Poisson-gamma [5], di mana makalah tersebut memberikan hasil bahwa model regresi Poisson-gamma merupakan model terbaik untuk menggambarkan angka kematian ibu di Kota Bandung. Akan tetapi, dikarenakan faktor-faktor tersembunyi tidak akan pernah diketahui, sehingga apabila faktor-faktor tersembunyi memiliki nilai yang ekstrem, maka distribusi gamma tidak cocok untuk memodelkannya. Penelitian ini memberikan alternatif, yaitu model regresi Poisson-lognormal. Kelebihan ini terkait dengan kemampuan distribusi lognormal untuk menggambarkan nilai-nilai yang relatif ekstrem [3]. Untuk mengilustrasikan kinerja regresi Poisson-lognormal, skripsi ini akan menyajikan analisis data riil menggunakan model tersebut.

1.5 Sistematika Pembahasan

BAB 2: Landasan Teori

Pada bab ini, akan dibahas teori-teori mengenai distribusi Poisson, distribusi lognormal, distribusi campuran (*mixture distribution*), distribusi campuran Poisson (*Poisson mixture distribution*), langkah-langkah dalam algoritma ekspektasi-maksimasi, simulasi Monte-Carlo, dan metode Newton-Raphson.

BAB 3: Model Regresi Poisson-Lognormal

Pada bab ini, akan dijelaskan penyusunan model regresi Poisson-lognormal untuk memprediksi frekuensi klaim asuransi kendaraan bermotor di masa mendatang. Kemudian, akan dijelaskan penerapan algoritma ekspektasi-maksimasi dalam menaksir parameter model regresi Poisson-lognormal.

BAB 4: Analisis Data Frekuensi Klaim Asuransi Kendaraan Bermotor

Pada bab ini, dilakukan analisis terhadap data yang digunakan dalam penelitian. Selain itu, hasil-hasil yang diperoleh melalui implementasi algoritma ekspektasi-maksimasi akan disajikan. Interpretasi mendalam terhadap model yang dihasilkan akan diuraikan, dan untuk memberikan pemahaman yang lebih konkret, beberapa contoh kasus akan disajikan secara rinci.

BAB 5: Kesimpulan dan Saran

Pada bab ini, akan dibahas kesimpulan dari penulisan penelitian ini dan saran untuk pengembangan lebih lanjut.

