

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Perkembangan teori statistika telah mempengaruhi hampir semua aspek kehidupan. Hal ini disebabkan statistika merupakan salah satu disiplin ilmu yang berperan sebagai alat untuk mengumpulkan, menyusun, menyajikan, menganalisis data serta mengambil kesimpulan yang bersifat objektif mengenai populasi berdasarkan data sampel.

Jenis-jenis statistika dibagi menjadi dua yaitu statistika deskriptif dan statistika induktif. Statistika deskriptif adalah metode yang berkaitan dengan pengumpulan dan penyajian suatu hasil pengamatan (data) sehingga memberikan informasi yang berguna bagi pihak-pihak yang berkepentingan terhadap data dan informasi tersebut. Sedangkan statistika induktif adalah metode yang berkaitan dengan analisis sebagian data (data dari sampel) yang kemudian digunakan untuk melakukan peramalan atau penaksiran kesimpulan (generalisasi) mengenai data secara keseluruhan.[12]

Beberapa materi yang dibahas di statistika induktif adalah: probabilitas dan teori keputusan, metode sampling, teori pendugaan, pengujian hipotesa, regresi dan korelasi, dan statistika nonparametik.

Salah satu materi yang ada di statistika induktif adalah analisis regresi. Analisis regresi merupakan suatu metode yang digunakan untuk menganalisis hubungan antar variabel. Hubungan tersebut dapat dinyatakan dalam bentuk persamaan yang menghubungkan variabel respon Y dengan satu atau lebih variabel prediktor X_1, X_2, \dots, X_n . Pola hubungan ini dapat dinyatakan dengan sebuah persamaan yang dinamakan persamaan regresi dan bentuknya bisa linier maupun non-linier.[12]

Analisis regresi yang umumnya digunakan adalah analisis regresi klasik, dimana variabel responnya merupakan data kontinu yang mengikuti distribusi normal. Namun dalam perkembangannya model regresi klasik ini tidak mampu

mengatasi permasalahan–permasalahan dimana variabel respon berupa data diskrit dan tidak berdistribusi normal.

Solusi untuk mengatasi masalah tersebut adalah dengan menggunakan model linier tergeneralisir atau biasa disebut *Generalized Linear Model (GLM)*. Uji asumsi yang diterapkan pada GLM tidak mengharuskan asumsi kenormalan dari variabel respon dan juga tidak mengharuskan kehomogenan dari variansinya.[21]

Salah satu model regresi dengan variabel responnya bukan dari distribusi normal adalah regresi Poisson. Model regresi Poisson sering digunakan ketika variabel responnya merupakan data hitung (*count data*) atau dapat pula digunakan untuk memodelkan banyak kejadian dari sebuah peristiwa pengamatan. Contohnya angka kunjungan kedokter, jumlah panggilan telepon setiap jam, banyak siswa yang tidak masuk sekolah, dan sebagainya. Contoh–contoh tersebut dapat dimodelkan dengan menggunakan regresi Poisson. Dalam regresi Poisson variabel respon diasumsikan berdistribusi Poisson, dengan fungsi masa peluangnya adalah:

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

Regresi Poisson memiliki masalah khusus yang membuatnya berbeda dengan regresi linier klasik, yaitu: [21]

1. Regresi Poisson tidak simetris sedangkan regresi klasik mengasumsikan galatnya memiliki distribusi yang simetris.
2. Regresi Poisson memiliki nilai peramalan untuk variabel respon bersifat non-negatif.
3. Untuk regresi Poisson, nilai variansinya akan bertambah seiring dengan pertambahan nilai rata–ratanya sedangkan regresi klasik mengasumsikan variansi konstan.

Untuk melinearkan model regresi Poisson, digunakan fungsi penghubung, fungsi penghubung yang digunakan dalam regresi Poisson adalah fungsi penghubung log. Fungsi penghubung log sangat cocok diterapkan dalam regresi Poisson, karena fungsi tersebut menjamin bahwa seluruh nilai peramalan dari variabel responnya bersifat non-negatif.

Regresi Poisson digunakan untuk menganalisis hubungan antara variabel respon yang berupa data diskrit dengan satu atau lebih variabel prediktor, dimana nilai ekspektasi (mean) dan variansinya diasumsikan sama (*equidispersi*) yaitu $E(Y) = Var(Y) = \mu$, namun dalam analisis data diskrit dengan menggunakan model regresi Poisson terkadang terjadi pelanggaran asumsi tersebut, dimana nilai variansinya lebih besar dari nilai mean yang disebut *overdispersi* atau varian lebih kecil dari nilai mean yang disebut *underdispersi*.

Dalam beberapa kasus sering ditemukan bahwa varians data yang diamati lebih besar daripada meannya yang biasa disebut dengan *overdispersi*. Hal ini mengindikasikan bahwa model regresi Poisson tidak cocok untuk data tersebut. Ada beberapa hal yang menyebabkan terjadinya fenomena *overdispersi* dalam pengamatan sebuah variabel diantaranya adanya sumber keragaman yang tidak teramati (*unobserved heterogeneity*), adanya pengamatan yang hilang (*missing*) pada variabel prediktor, adanya pencilan pada data sehingga perlunya interaksi dalam model, variabel prediktor perlu ditransformasi atau kesalahan spesifikasi fungsi penghubung.[21]

Salah satu cara untuk mengatasi *overdispersi* adalah dengan menggunakan model regresi binomial negatif. Model regresi binomial negatif merupakan perluasan dari model regresi Poisson dimana variansi pada regresi binomial negatif lebih besar daripada meannya. Variabel respon pada regresi binomial negatif juga diasumsikan mengikuti distribusi binomial negatif sehingga tidak mengharuskan nilai meannya sama dengan nilai variansinya. Fungsi masa peluang dari distribusi binomial negatif adalah:[17]

$$P(x; k, p) = \binom{x-1}{k-1} p^k q^{x-k}, \quad x = k, k+1, k+2, \dots$$

untuk membentuk model regresi binomial negatif maka distribusi binomial negatif dapat dihasilkan dari distribusi *mixture* Poisson-gamma.

Sehingga fungsi masa peluang distribusi Binomial Negatif adalah:

$$f(y; \mu, k) = \frac{\Gamma\left(y_i + \frac{1}{k}\right)}{\Gamma\left(\frac{1}{k}\right) y_i!} \left(\frac{1}{1 + k\mu_i}\right)^{\frac{1}{k}} \left(\frac{k\mu_i}{1 + k\mu_i}\right)^{y_i}, \quad y_i = 0, 1, 2, \dots$$

dengan $\mu_i = \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi})$

Setelah rumusan model regresi binomial negatif didapatkan maka selanjutnya dilakukan penerapan model regresi Poisson dan model regresi binomial negatif pada pengaplikasian faktor-faktor yang mempengaruhi kecelakaan laulintas. Kecelakaan lalulintas merupakan salah satu penyebab kematian terbesar di Indonesia. Akibat dari kecelakaan laulintas adalah jumlah korban yang cukup besar yang akan memberikan dampak ekonomi (kerugian material) dan kerugian sosial yang tidak sedikit, berbagai usaha prefentif hingga perbaikan lalulintas dengan melibatkan berbagai pihak yang terkait hasilnya belum sesuai yang diharapkan.[19] Data kecelakaan lalulintas yang didapatkan dari POLDA JABAR dengan faktor-faktor yang mempengaruhi kecelakaan lalulintas adalah faktor manusia, faktor jalan, faktor alam dan faktor teknologi maka dari keempat faktor tersebut manakah faktor-faktor yang lebih signifikan dalam mempengaruhi kecelakaan lalulintas.

Berdasarkan uraian di atas, penulis tertarik untuk mengkaji permasalahan tersebut dalam tugas akhir (skripsi) yang selanjutnya diberi judul “Penerapan Regresi Binomial Negatif Dalam Mengatasi *Overdispersi* Pada Regresi Poisson Untuk Mengetahui Faktor-Faktor Yang Mempengaruhi Kecelakaan Lalulintas Di POLDA JABAR”.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, penulis dapat mengemukakan rumusan masalah sebagai berikut:

1. Bagaimana menerapkan data kecelakaan lalulintas menggunakan model regresi Poisson?
2. Bagaimana menggunakan model regresi binomial negatif apabila datanya mengalami *overdispersi*?
3. Bagaimana menentukan faktor-faktor yang mempengaruhi kecelakaan lalulintas di POLDA JABAR?

1.3 Batasan Masalah

Batasan masalah pada skripsi ini adalah:

- Variabel acak yang digunakan adalah berupa variabel acak diskrit.
- Variabel respon (Y) berupa data hitung (*count data*).
- Sampel yang digunakan adalah data kecelakaan lalulintas pada mobil penumpang di POLDA JABAR
- *Software* yang digunakan untuk studi kasus adalah MATLAB dan Maple 16
- Data kecelakaan lalulintas yang di ambil adalah data kecelakaan lalulintas pada mobil penumpang

1.4 Tujuan Penelitian

Adapun tujuan penelitian pada penulisan skripsi ini adalah:

1. Untuk menerapkan data kecelakaan lalulintas dengan menggunakan model regresi Poisson.
2. Menggunakan model regresi binomial negatif apabila datanya mengalami *overdispersi*.
3. Menentukan faktor-faktor yang mempengaruhi kecelakaan lalulintas di POLDA JABAR.

1.5 Manfaat Penelitian

Adapun manfaat penelitian pada penulisan skripsi ini adalah:

1. Manfaat Praktis

Adapun manfaat praktis penulisan skripsi ini adalah untuk memberikan informasi pada pihak terkait khususnya POLDA JABAR tentang faktor-faktor yang mempengaruhi kecelakaan lalulintas pada mobil penumpang.

2. Manfaat Teoritis

Adapun manfaat teoritis penulisan skripsi ini adalah untuk menambah dan memperluas pengetahuan tentang regresi, khususnya yang berhubungan dengan regresi Poisson, dimana untuk mengatasi *overdispersi* pada regresi Poisson dapat dilakukan dengan menggunakan regresi binomial negatif.

1.6 Metode Penelitian

Metode penelitian yang digunakan pada penulisan skripsi ini adalah sebagai berikut:

1. Studi Literatur

Dalam studi literatur peneliti mengkaji sumber-sumber, baik dari buku, jurnal maupun internet yang berhubungan dengan model regresi Poisson, regresi binomial negatif yang akan dimodelkan untuk mengetahui faktor-faktor yang mempengaruhi kecelakaan lalulintas di POLDA JABAR.

2. Bimbingan dengan Pihak Ahli

Pada tahap ini peneliti melakukan konsultasi pada pihak ahli pada dosen pembimbing I dan dosen pembimbing II, serta pihak-pihak yang membantu dalam pengerjaan skripsi mengenai model regresi Poisson, model regresi binomial negatif untuk mendapatkan informasi dan membantu dalam hal pemahaman.

3. Studi Lapangan

Pada studi lapangan peneliti mengambil data untuk mengaplikasikannya pada model regresi Poisson, model regresi binomial negatif pada kasus kecelakaan lalulintas dengan pengambilan data di POLDA JABAR.

1.7 Sistematika Penulisan

Sistematika penulisan pada skripsi ini adalah sebagai berikut:

BAB I PENDAHULUAN

Bab ini merupakan pendahuluan dari skripsi yang berisi latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, metode penelitian dan sistematika penulisan.

BAB II LANDASAN TEORI

Dalam bab ini akan dibahas mengenai landasan teori yang memuat teori pendukung pada bab III yaitu: variabel acak, variabel acak diskrit, variabel acak kontinu, distribusi keluarga eksponensial, distribusi Poisson, uji kecocokan distribusi Poisson, distribusi binomial negatif, analisis regresi, *Generalized Linier Model* (GLM), penaksir maksimum likelihood, eliminasi Gauss-jordan, uji *Goodness of Fit* dan uji skor.

BAB III PEMBAHASAN

Bab ini merupakan bab inti materi karena didalam bab ini akan dibahas pembentukan model regresi Poisson, penaksiran parameter model regresi Poisson, pengujian kesesuaian model regresi Poisson, *overdispersi*, model regresi binomial negatif, penaksir parameter model regresi binomial negatif serta pengujian kesesuaian model regresi binomial negatif .

BAB IV STUDI KASUS

Pada bab ini pengaplikasian rumus-rumus yang ada pada bab III dengan data faktor-faktor yang mempengaruhi kecelakaan lalulintas pada mobil penumpang di POLDA JABAR .

BAB V KESIMPULAN DAN SARAN

DAFTAR PUSTAKA

