

# BAB I

## PENDAHULUAN

Pada bab ini akan dibahas mengenai latar belakang tema yang diambil dalam tugas akhir ini, perumusan masalah yang akan dibahas, batasan masalah, tujuan penulisan, manfaat penulisan, dan sistematika penulisan.

### 1.1 Latar Belakang

Dalam kehidupan sehari-hari, sering ditemukan persoalan yang melibatkan dua atau lebih variabel yang diduga memiliki suatu hubungan tertentu. Bentuk hubungan ini dikenal dengan nama regresi dari satu variabel atas variabel lain. Dengan kata lain, analisis regresi merupakan sebuah metode statistika yang mempelajari hubungan fungsional atau kausal diantara variabel-variabelnya. Variabel-variabelnya dibedakan menjadi dua, yaitu variabel respon atau variabel terikat (*dependent variable*), dan variabel bebas atau variabel prediktor (*independent variable*). Variabel respon biasanya disimbolkan dengan Y dan variabel bebas biasanya disimbolkan dengan X. Hubungan ini biasanya dinyatakan dalam persamaan matematis yang dinamakan persamaan regresi. Persamaan regresi ini dapat berbentuk linear maupun non-linear.

Model regresi linear sederhana sudah mulai digunakan sejak awal abad 19, ditandai dengan kajian-kajian yang dilakukan oleh Francis Galton (1822-1911) tentang hubungan tinggi badan ayah dan anaknya (Manurung, 2005: 9). Dalam perkembangannya, model regresi linear dengan asumsi peubah respon  $Y_i \sim N(\mu_i, \sigma^2)$  tidaklah mampu menjawab semua masalah yang dihadapi dalam pemodelan statistik. *Generalized Linear Model* (GLM) merupakan pengembangan dari model linear "klasik" khususnya dalam mengatasi kendala peubah respon yang tidak normal. Pada umumnya, peubah respon dalam GLM diasumsikan memiliki distribusi yang termasuk dalam distribusi keluarga eksponensial. Distribusi yang termasuk kedalam distribusi keluarga eksponensial adalah distribusi normal, Poisson, binomial, eksponensial, dan gamma (Montgomery *et al*, 2006: 427).

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 ke dokter, jumlah panggilan telepon setiap jam, banyak siswa yang tidak masuk sekolah, dan sebagainya. Contoh-contoh tersebut dapat dimodelkan dengan menggunakan regresi Poisson.

Menurut Steve (2007), regresi Poisson memiliki masalah khusus yang membuatnya berbeda dengan regresi linear klasik, yaitu

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 yang konstan.

Dalam GLM, digunakan fungsi penghubung (*link function*), fungsi penghubung yang digunakan dalam regresi Poisson adalah logaritma. Fungsi penghubung log sangat cocok diterapkan dalam regresi Poisson, karena fungsi tersebut menjamin bahwa seluruh nilai peramalan dari variabel responnya bersifat non-negatif.

Karakteristik distribusi Poisson adalah nilai rata-rata sama dengan nilai variansinya. Dalam beberapa kasus, sering ditemukan bahwa variansi data yang terobservasi lebih besar daripada rata-ratanya, yang dikenal dengan overdispersi dan mengindikasikan bahwa model regresi Poisson tidak cocok untuk data tersebut. Dengan kata lain, kasus overdispersi menyiratkan bahwa taksiran yang diperoleh akurat tetapi tidak setepat yang dipikirkan.

Untuk mengatasi hal ini, ada dua alternatif yang dapat digunakan untuk memodelkan data tersebut. Yang pertama adalah distribusi binomial negatif yang merupakan perluasan dari distribusi Poisson. Regresi binomial negatif memiliki

variansi yang lebih besar daripada rata-ratanya. Yang kedua adalah model *Zero-Inflated Poisson* yang juga merupakan pengkhususan dari distribusi Poisson. Distribusi *Zero-Inflated Poisson* akan terbentuk, apabila pada suatu kumpulan data hitung terdapat nilai nol yang jumlahnya sangat besar (Steve, 2007).

Berdasarkan uraian di atas, penulis tertarik untuk menyikapi permasalahan tersebut dalam suatu karya tulis dalam rangka penyusunan tugas akhir. Untuk selanjutnya, tugas akhir ini diberi judul “Model *Zero-Inflated Poisson* untuk Mengatasi Overdispersi pada Regresi Poisson.”

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan di atas, maka permasalahan dalam pembahasan ini dapat dirumuskan sebagai berikut:

1. Bagaimanakah bentuk regresi linear yang variabel responnya mengikuti distribusi Poisson?
2. Bagaimana cara menguji adanya overdispersi pada model regresi Poisson?
3. Jika ternyata terdapat overdispersi, maka tindakan apakah yang dapat dilakukan untuk mengatasinya?
4. Bagaimana perancangan model *Zero-Inflated Poisson* dalam suatu contoh kasus?

### 1.3 Batasan Masalah

Masalah yang dipaparkan dalam penelitian ini akan dibatasi sebagai berikut:

1. Penaksiran parameter untuk regresi Poisson akan menggunakan *software S-PLUS*, dan penaksiran parameter untuk regresi *Zero-Inflated Poisson* akan menggunakan *software STATA*, tidak dilakukan secara manual.
2. Pengujian adanya overdispersi pada regresi Poisson dengan menggunakan statistik uji skor (C. B. Dean, 1992, dalam Arie S. Utami, 2003: 10).

### 1.4 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah:

1. Membentuk regresi linear yang variabel responnya mengikuti distribusi Poisson.
2. Menguji adanya kasus overdispersi dalam regresi Poisson.
3. Merumuskan tindakan yang tepat untuk mengatasi overdispersi.
4. Merancang model *Zero-Inflated Poisson* dalam suatu contoh kasus.

### 1.5 Manfaat Penelitian

1. Teoritis

Memperkaya dan memperluas pengetahuan tentang regresi, khususnya yang berkaitan dengan regresi Poisson. Disamping itu, tugas akhir ini sebagai evaluasi terhadap kemampuan dalam mengaplikasikan teori-teori tentang regresi yang telah disampaikan semasa perkuliahan.

## 2. Praktis

Sebagai bahan pertimbangan dan masukan bagi pihak yang berkepentingan, serta dapat dijadikan sebagai salah satu sumber informasi yang dapat mendukung tujuan dari pihak yang berkepentingan tersebut.

### 1.6 Sistematika Penulisan

Tugas akhir ini akan disusun dengan sistematika sebagai berikut:

#### BAB I PENDAHULUAN

Bab ini merupakan pendahuluan dari tugas akhir ini dan berisi latar belakang masalah yang akan dibahas, rumusan permasalahan, tujuan penelitian, batasan permasalahan, dan sistematika penulisan.

#### BAB II TINJAUAN TEORI

Dalam bab ini akan dibahas mengenai tinjauan pustaka yang memuat teori pendukung untuk permasalahan yang diangkat penulis.

#### BAB III MODEL *ZERO-INFLATED POISSON* UNTUK MENGATASI OVERDISPERSI PADA REGRESI POISSON

Bab ini merupakan bab inti karena di dalam bab ini akan dibahas mengenai regresi Poisson, overdispersi, dan model *Zero-Inflated Poisson*.

#### BAB IV CONTOH KASUS

Bab ini berisi contoh kasus untuk penerapan model *Zero-Inflated Poisson*, pengolahan datanya, dan pembahasan hasil pengolahan data.

## BAB V KESIMPULAN DAN SARAN

Bab V. Bab terakhir ini memuat kesimpulan dalam tugas akhir ini serta saran untuk pembaca.

