

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Ilmu statistika sering digunakan dalam berbagai bidang ilmu pekerjaan misalnya pada bidang ekonomi, sosial, politik, industri, pendidikan, dan kesehatan. Terdapat beberapa model regresi, namun pembedanya sendiri dilihat berdasarkan variabel terikat  $Y$  yang digunakan. Istilah regresi pertama kali dikenalkan oleh Francis Galton. Pada karya tulisnya yang terkenal, Galton menemukan bahwa adanya hubungan tinggi badan anak dengan tinggi badan orang tuanya. Orang tua yang memiliki badan tinggi cenderung akan memiliki anak yang tinggi pula, dan sebaliknya, orang tua yang memiliki badan pendek cenderung akan memiliki anak yang berbadan pendek. Rata-rata tinggi badan anak yang memiliki badan tinggi dan rata-rata tinggi badan anak yang memiliki badan pendek akan sama-sama menuju (beregresi) rata-rata tinggi badan populasi (Gujarati, 2013). Regresi yang dikemukakan Galton kemudian dikonfirmasi oleh temannya Karl Pearson.

Interpretasi istilah regresi *modern* agak sedikit berbeda. Menurut Gujarati (2013), analisis regresi berkaitan dengan studi mengenai ketergantungan satu variabel, yaitu variabel dependen, terhadap satu atau lebih variabel lainnya, yaitu variabel penjelas dengan tujuan untuk mengestimasi dan/atau memperkirakan nilai rerata atau rata-rata (populasi) variabel dependen dari nilai yang diketahui atau nilai tetap dari variabel penjelas (dalam sampling berulang). Pada analisis regresi linier klasik, umumnya variabel terikat yang digunakan hanya berdistribusi diskrit atau kontinu saja. Namun kenyataan di lapangan banyak ditemukan kasus dimana variabel terikat mempunyai distribusi gabungan diskrit dan kontinu. Variabel terikat yang mempunyai sifat gabungan diskrit dan kontinu ini dinamakan data tersensor (Greene, 2012). Dalam pengertian yang lebih luas, data tersensor merupakan data dimana informasi tentang variabel terikat yang diteliti tidak memberikan informasi yang lengkap (Gujarati, 2013). Analisis regresi klasik tidak dapat digunakan dalam kasus seperti ini, karena akan terjadi pelanggaran asumsi-asumsi regresi dan bias parameter. Salah satu jenis regresi lain yang dapat digunakan untuk mengestimasi data tersensor adalah regresi tobit (Tobin, 1958).

Regresi tobit pertama kali dikemukakan oleh James Tobin pada tahun 1958, pada analisis regresi tobit, estimasi yang digunakan bukan estimasi *Ordinary Least Square* (OLS). Metode OLS tidak dapat mengestimasi parameter regresi tobit karena jenis data yang digunakan adalah data tersensor. Jika dipaksakan mengestimasi menggunakan metode ini, maka akan terjadi pelanggaran asumsi regresi dan bias parameter. Estimasi yang dapat digunakan untuk menganalisis regresi tobit adalah *Maximum Likelihood Estimation* (MLE). Estimasi ini digunakan karena tahan terhadap pengujian jenis data berdistribusi campuran (tersensor). Menurut Greene (2008) menyebutkan bahwa variabel terikat yang bersifat campuran (mixture) memiliki struktur data dengan skala diskrit untuk yang bernilai nol, dan berskala kontinu untuk yang tidak bernilai nol. Data tersebut disebut juga data tersensor. Sebaran data tersensor adalah sebaran normal tersensor, yang mengikuti asumsi  $N(\mu, \sigma^2)$ . Penggunaan regresi tobit (regresi tersensor) pada data campuran akan mengurangi efek bias jika dibandingkan dengan data yang diolah menggunakan regresi linier klasik.

Kanker payudara terjadi karena adanya kerusakan pada gen yang mengatur pertumbuhan dan differensiasi sehingga sel itu tumbuh dan berkembang biak tanpa dapat dikendalikan. Penyebaran kanker payudara terjadi melalui pembuluh getah bening dan tumbuh di kelenjar getah bening, sehingga kelenjar getah bening aksila ataupun supraklavikula membesar. Kemudian melalui pembuluh darah kanker menyebar ke organ lain seperti paru-paru, hati dan otak. Kanker payudara ditemukan di seluruh dunia, tahun 2003, insiden kanker payudara di Belanda 91 per 100.000 penduduk, Amerika 71,7 per 100.000 penduduk, Swiss 70 per 100.000 wanita, Australia 83,2 per 100.000 penduduk, Kanada 84,7, Indonesia 26 per 100.000 wanita pada tahun 2002 dan Jepang 16 per 100.000 penduduk. dalam penelitian ini, metode yang akan digunakan. Kanker payudara lebih sering dijumpai pada umur 40-49 tahun yaitu sebesar 30,35%. Sehingga analisis seperti inilah yang sangat diperlukan untuk mengetahui seberapa lama pasien yang menderita kanker paru dapat bertahan hidup. Dan seberapa besar resiko kegagalan dalam bertahan hidup selama waktu yang diperkirakan

## 1.2 Rumusan Masalah

Berdasarkan uraian pada latar belakang maka disusun perumusan masalah sebagai berikut :

1. Bagaimana menentukan Fungsi likelihood estimator model regresi Tobit Kuantil Bayesian terbaik ?
2. Bagaimana perbandingan performa estimator parameter model regresi Tobit Kuantil Bayesian dengan estimator Powell?
3. Bagaimana model regresi Tobit Kuantil Bayesian terbaik bagi pasien pengidap kanker payudara?

## 1.3 Tujuan Penelitian

Tujuan dalam penelitian ini adalah :

1. Untuk mengetahui Fungsi likelihood estimator model regresi Tobit Kuantil Bayesian terbaik.
2. Untuk mengetahui perbandingan performa estimator parameter model regresi Tobit Kuantil Bayesian dengan estimator Powell.
3. Untuk mengetahui model regresi Tobit Kuantil Bayesian terbaik bagi bagi pasien pengidap kanker payudara.

## 1.4 Manfaat Penelitian

Adapun manfaat tugas akhir ini adalah sebagai berikut.

1. Bagi Mahasiswa
  - a. Mahasiswa dapat menerapkan ilmu pengetahuan dan metodologi yang telah diperoleh selama di bangku kuliah dalam penerapan di dunia nyata.
  - b. Menguji kemampuan mahasiswa mengenai penerapan ilmu yang telah diperoleh.
  - c. Memberi tambahan wawasan ilmu pengetahuan kepada mahasiswa mengenai penerapan ilmu statistika terutama dalam hal kesehatan
2. Bagi Jurusan Statistika
  - a. Membangun kerjasama pihak RSUP Dr. Kariadi Semarang dengan pihak Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Muhammadiyah Semarang.
  - b. Menjadi referensi untuk mahasiswa maupun dosen dalam melakukan penelitian lebih mendalam mengenai analisis statistik yang digunakan untuk pasien kanker payudara di RSUP Dr. Kariadi Semarang khususnya.

### 3. Bagi RSUP Dr.Kariadi Semarang

Sumber informasi kepada RSUP Dr. Kariadi Semarang mengenai rata-rata waktu hidup pasien kanker payudara peluang bertahan hidup, dan peluang kegagalan pasien untuk bertahan secara statistika.

#### 1.5 Batasan Masalah

Batasan dalam penelitian ini adalah :

1. Penelitian ini dibatasi hanya dengan melihat dari hasil dokumentasi pasien yang mengidap penyakit kanker payudara RSUP Dr. Kariadi Semarang yang dimulai pada tanggal 25 Maret 2017 sampai dengan 25 Mei 2017.
2. Data yang diambil yakni data pasien kanker payudara yang dibatasi dengan sensor tipe I yakni dimulai dari tanggal 1 Januari 2013 sampai dengan 31 Mei 2017 dan diperoleh sebanyak 50 data pasien.
3. Ukuran untuk mengevaluasi performa estimator adalah RMSE.
4. Estimator Powell yang akan dibandingkan dengan estimator parameter model
5. Kuantil yang akan dimodelkan adalah 0.05, 0.25, 0.50, 0.75 dan 0.95

