BAB I
PENDAHULUAN

A. Latar Belakang

Analisis regresi linear merupakan metode statistika yang digunakan untuk membentuk model hubungan antara variabel dependen (terikat; respon) dengan satu atau lebih variabel independen (bebas; prediktor). Analisa ini dipergunakan untuk menelaah hubungan antara dua variabel atau lebih, terutama untuk menelusuri pola hubungan yang modelnya belum diketahui dengan sempurna. Model regresi linear yang terdiri dari satu variabel independen dan satu variabel dependen disebut dengan regresi linear sederhana, sedangkan model regresi linear yang terdiri dari beberapa variabel independen dan satu variabel dependen merupakan model regresi linear berganda.

Analisis regresi linear memiliki beberapa kegunaan, yaitu (a) untuk mempelajari dan mengukur hubungan statistik yang terjadi antara dua variabel atau lebih, (b) untuk meramalkan atau memperkirakan nilai dari satu variabel dalam hubungannya dengan variabel lain yang diketahui melalui persamaan regresi, dan (c) untuk memeriksa dan memodelkan hubungan diantara variabel independen dan variabel dependen. Analisis regresi linear bertujuan untuk (a) mengestimasi parameter koefisien regresi, (b) untuk menguji hipotesis mengenai sifat alamiah ketergantungan, (c) untuk memperediksi atau meramalkan parameter koefisien regresi (Gujarati, 2004).

Model regresi merupakan suatu cara formal untuk mengekspresikan dua unsur penting pada suatu hubungan statistik, yaitu kecenderungan berubahnya

variabel dependen secara sistematis sejalan dengan berubahnya variabel independen dan berpencarnya titik-titik di sekitar kurva taksiran model tersebut. Bentuk model regresi linear berganda adalah $Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + e_i, i = 1,2,3, \dots, n$ dengan Y_i merupakan variabel dependen pada pengamatan ke-i, X_{ik} merupakan variabel independen pada pengamatan ke-i dan parameter ke-*k* dan $\beta_0, \beta_1, \dots, \beta_k$ merupakan parameter regresi yang tidak diketahui nilainya dan akan dicari nilai estimasinya.

Metode yang biasa digunakan untuk mengestimasi koefisien regresi adalah metode kuadrat terkecil (MKT) atau dikenal dengan *ordinary least squares* (OLS). Namun, metode ini mensyaratkan bahwa distribusi data harus memenuhi asumsi klasik dari regresi, yaitu harus linear dalam parameter, galat berdistribusi normal dengan rataan nol dan ragamnya konstan, antara galat tidak berkorelasi, dan antar variabel independen tidak saling berkorelasi. Jika data tidak memenuhi salah satu asumsi klasik regresi, maka penduga metode kuadrat terkecil tidak lagi efisien. Dalam kasus model regresi linear, dimungkinkan terdapat data *outlier*. *Outlier* merupakan observasi dengan nilai mutlak residunya jauh lebih besar daripada residu-residu lainnya, sehingga akan mempengaruhi model regresi yang terbentuk. Data *outlier* tersebut tidak boleh dibuang begitu saja karena akan mempengaruhi model prediksi serta akan menghasilkan estimasi parameter yang kurang tepat. Untuk menyelesaikan masalah tersebut diperlukan adanya metode yang bersifat *robust* (kekar), sehingga saat terjadi perubahan kecil dalam data tidak akan mempengaruhi nilai estimasinya. Regresi *robust* diperkenalkan oleh Andrews (1972) merupakan alat penting untuk menganalisis data yang

dipengaruhi oleh *outlier* untuk menghasilkan model yang *robust* atau *resistance* terhadap *outlier*. Suatu estimasi yang *resistance* adalah relatif tidak terpengaruh oleh perubahan besar pada bagian kecil data atau perubahan kecil pada bagian besar data (Widodo, Guritno, & Haryatmi, 2013).

Menurut Chen (2002) regresi *robust* terdiri dari 5 metode estimasi, yaitu (1) estimasi-M (*Maximum Likelihood type*), (2) estimasi-LMS (*Least Median Squares*), (3) estimasi-LTS (*Least Trimmed Squares*), (4) estimasi-MM (*Method of Moment*), dan (5) estimasi-S (*Scale*). Estimasi-M atau sering disebut juga dengan estimasi-M Huber merupakan metode regresi *robust* yang paling sering digunakan karena dipandang baik untuk mengestimasi parameter yang disebabkan oleh adanya *outlier*. Dan metode estimasi-M ini diperkenalkan oleh Huber (1973). Sedangkan estimasi-S merupakan estimasi *robust* yang mempunyai nilai *breakdown point* paling tinggi hingga 50%. *Breakdown point* merupakan fraksi terkecil dari data yang terkontaminasi *outlier* yang dapat menyebabkan estimator tidak berfungsi (Montgomery, Peck, & Vining, 2006).

Dari kelima metode regresi *robust* tersebut penulis tertarik untuk mempelajari dua diantaranya yaitu membandingkan regresi *robust* dengan estimasi-M Huber dan estimasi-S dalam mengatasi masalah *outlier*.

B. Batasan Masalah

Pembatasan masalah yang dikenakan pada penulisan skripsi ini yaitu pada penentuan estimator dengan menggunakan metode estimasi-M Huber dan estimasi-S untuk mengatasi kasus *outlier* pada model regresi linear berganda.

C. Rumusan Masalah

Berdasarkan uraian latar belakang rumusan masalah yang akan dibahas di dalam skripsi ini adalah:

1. Bagaimana mengestimasi parameter pada model regresi linear berganda dengan menggunakan regresi *robust* estimasi-M Huber dam estimasi-S.
2. Bagaimana hasil perbandingan regresi *robust* dengan menggunakan estimasi-M Huber dan estimasi-S.

D. Tujuan

Tujuan dalam penulisan skripsi ini adalah:

1. Menunjukkan langkah-langkah dalam mengestimasi parameter regresi *robust* dengan menggunakan estimasi-M Huber dan estimasi-S.
2. Menunjukkan hasil perbandingan regresi *robust* dengan menggunakan estimasi-M Huber dengan estimasi-S.

E. Manfaat

Manfaat dari penulisan ini adalah:

1. Bagi Jurusan Matematika FMIPA UNY

   Menambah kelengkapan koleksi pustaka dan menjadi dasar pertimbangan untuk penelitian-penilitian selanjutnya.
2. Bagi Mahasiswa

   Sebagai acuan untuk penulisan karya ilmiah selanjutnya khususnya mengenai regresi *robust*.