

# BAB I

## PENDAHULUAN

### A. Latar Belakang

Analisis regresi linier merupakan teknik dalam statistika yang digunakan untuk membentuk model hubungan antara variabel dependen dengan satu atau lebih variabel independen. Dalam analisis regresi dibedakan dua jenis variabel yaitu variabel dependen dan variabel independen. Regresi linier yang terdiri dari satu variabel dependen dan satu variabel independen disebut regresi linier sederhana, sedangkan regresi linier yang terdiri dari satu variabel dependen dan beberapa variabel independen disebut regresi linier berganda. Hubungan antar variabel-variabel tersebut dapat dinyatakan dalam model matematika. Bentuk umum model regresi linier

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon.$$

Keterangan:  $Y$  = variabel dependen

$\beta_0, \beta_1, \dots, \beta_k$  = koefisien regresi

$X_1, X_2, \dots, X_k$  = variabel independen

$\varepsilon$  = error

Salah satu tujuan dalam analisis regresi adalah untuk mengestimasi rata-rata populasi atau nilai rata-rata variabel dependen berdasarkan nilai variabel independen yang diketahui. Hasil dari analisis regresi berupa koefisien regresi untuk masing-masing variabel independen. Pada umumnya digunakan metode estimasi kuadrat terkecil untuk mengestimasi koefisien regresi. Metode kuadrat terkecil adalah suatu metode yang digunakan untuk mengestimasi koefisien garis

regresi dengan cara meminimumkan jumlah kuadrat residual. Penggunaan metode kuadrat terkecil memerlukan beberapa asumsi klasik yang harus dipenuhi. Beberapa asumsi itu antara lain:

- (1)  $\varepsilon$  merupakan variabel random dan mengikuti distribusi normal;
- (2) varians dari  $\varepsilon$  adalah konstan dan homoskedastisitas;
- (3) tidak ada autokorelasi; dan
- (4) tidak ada multikolinieritas di antara variabel independen.

Jika asumsi-asumsi klasik dalam metode kuadrat terkecil terpenuhi maka penduga parameter yang diperoleh bersifat *Best Linier Unbiased Estimation* (BLUE). Pada kenyataannya, asumsi ini tidak selalu terpenuhi sehingga penggunaan metode kuadrat terkecil kurang tepat. Salah satu penyebab tidak terpenuhinya asumsi klasik adalah adanya *outlier*. *Outlier* adalah satu atau beberapa data yang terlihat jauh dari pola kumpulan data keseluruhan.

Rousseeuw dan Leroy (1987) menyatakan terdapat dua cara untuk mengatasi masalah *outlier*:

1. Mengeluarkan *outlier* yang telah diidentifikasi, kemudian tetap menggunakan metode kuadrat terkecil.
2. Tetap menggunakan seluruh data, tetapi dengan memberikan bobot yang rendah untuk observasi yang terindikasi sebagai *outlier*, metode ini dikenal dengan nama metode regresi *robust*.

Adanya *outlier* dalam metode kuadrat terkecil mengakibatkan estimasi koefisien garis regresi yang diperoleh tidak tepat. Hal ini berarti nilai estimasi parameter-parameter dalam model regresi linier dapat dipengaruhi oleh satu titik

data ekstrim yang merupakan *outlier*. Pendeteksian *outlier* merupakan tahapan diagnosis yang perlu dilakukan terutama jika estimasi modelnya dengan metode kuadrat terkecil, yang dikenal cukup peka terhadap *outlier*. Metode pendeteksian *outlier* dapat dilakukan dengan beberapa metode, antara lain metode *boxplot*, *standardized residual*, dan *Cook's Distance*. Mengabaikan masalah *outlier* dalam data dapat mengakibatkan kesimpulan kurang akurat. Secara statistik, membuang *outlier* bukanlah tindakan yang bijaksana, karena suatu *outlier* dapat memberikan informasi cukup berarti. Oleh karena itu diperlukan suatu alternatif terhadap keberadaan *outlier*, yaitu dengan regresi *robust*.

Regresi *robust* diperkenalkan oleh Andrews (1972). “Regresi *robust* merupakan metode regresi yang digunakan ketika distribusi dari error tidak normal dan atau adanya beberapa *outlier* yang berpengaruh pada model ” (Olive, 2005: 3). Metode ini merupakan alat penting untuk menganalisis data yang dipengaruhi *outlier* sehingga dihasilkan model yang *robust* atau resisten terhadap *outlier*. Suatu estimasi yang resisten adalah estimasi yang relatif tidak terpengaruh oleh perubahan besar pada bagian kecil data atau perubahan kecil pada bagian besar data.

Menurut Chen (2002: 1), 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*). Kelima metode regresi tersebut mempunyai kelebihan dan kelemahan masing-masing. Dilihat dari nilai *breakdown point*-nya, estimasi-S merupakan estimasi *robust* yang mempunyai

nilai *breakdown point* paling tinggi hingga 50%. Menurut Huber (1981: 13), *breakdown point* adalah fraksi terkecil atau persentase dari *outlier* yang dapat menyebabkan nilai estimator menjadi besar. *Breakdown point* digunakan untuk menjelaskan ukuran kerobustan dari teknik *robust*. Kemungkinan tertinggi *breakdown point* untuk sebuah estimator adalah 50%. Estimasi-S memiliki beberapa fungsi pembobot yang dapat digunakan untuk mengestimasi parameter yang lebih resisten terhadap *outlier* diantaranya adalah *Welsch* dan *Tukey bisquare*. Fungsi pembobot ini digunakan untuk menghasilkan nilai skala pembobot yang diperoleh dengan cara melakukan iterasi hingga estimator yang diperoleh konvergen. Semakin kecil nilai skala yang dari suatu pembobot maka semakin *robust* (resisten) terhadap *outlier*.

Analisis regresi *robust* pernah dilakukan oleh beberapa peneliti. Peneliti terdahulu yang berkaitan dengan regresi *robust* antara lain: Yuliana Susanti, Hasih Pratiwi, Sri Sulistijowati dan Twenty Liana (2014) membandingkan estimasi-M, estimasi-S, dan estimasi-MM, diperoleh estimasi-S merupakan metode estimasi yang paling optimum; Anggono Harman (2014) membandingkan pembobot *Tukey Bisquare* dan pembobot *Fair* dalam regresi *robust* estimasi-M, dari kedua pembobot tersebut diperoleh pembobot *Tukey bisquare* menghasilkan model regresi yang lebih baik dibandingkan pembobot *Fair*; dan Dian Cahyawati, Hadi Tanuji, dan Riri Abdiati (2009) mengenai efektivitas metode regresi *robust* penduga *Welsch*, diperoleh hasil bahwa pendugaan parameter menggunakan metode regresi *robust* penduga *Welsch* menghasilkan model regresi yang lebih

baik dari metode kuadrat terkecil, untuk berbagai ukuran sampel dan banyak *outlier*.

Berdasarkan uraian tersebut, pada skripsi ini akan dibahas mengenai analisis regresi dari regresi *robust* estimasi-S menggunakan pembobot *Welsch* dan *Tukey bisquare* beserta modelnya dan perbandingan keefektifan dari estimasi-S menggunakan kedua pembobot tersebut. Data yang digunakan dalam contoh kasus penulisan skripsi ini adalah data Indeks Pembangunan Manusia (IPM) menurut provinsi di Indonesia tahun 2015. Variabel bebas yang digunakan adalah rata-rata lama sekolah, Upah Minimum Regional (UMR), dan Produk Domestik Regional Bruto (PDRB) dengan variabel terikat Indeks Pembangunan Manusia. Pemilihan variabel independen berdasarkan dimensi dasar dan konsep dari Indeks Pembangunan Manusia dan dipilih variabel data yang mengandung *outlier*. Kemudian untuk perbandingan keefektifan estimasi-S menggunakan pembobot *Welsch* dan pembobot *Tukey bisquare* ditinjau dari nilai *standard error* dan nilai *adj R-square* yang diperoleh dengan menggunakan *software R* versi 3.2.3 dan SPSS.

## **B. Batasan Masalah**

Pada skripsi ini, penulis memberikan batasan masalah sebagai berikut:

1. Data yang digunakan adalah data yang memuat *outlier*.
2. Model regresi yang digunakan adalah model regresi linier.
3. Metode yang digunakan adalah regresi *robust* estimasi-S dengan pembobot *Welsch* dan *Tukey bisquare*.

### C. Rumusan Masalah

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

1. Bagaimana prosedur analisis regresi *robust* estimasi-S dengan pembobot *Welsch* dan *Tukey bisquare* dalam mengatasi *outlier* pada analisis regresi?
2. Bagaimana model regresi *robust* estimasi-S dengan pembobot *Welsch* dan pembobot *Tukey bisquare* dalam mengatasi *outlier* pada data Indeks Pembangunan Manusia menurut provinsi di Indonesia tahun 2015?
3. Bagaimana perbandingan keefektifan regresi *robust* estimasi-S dengan menggunakan pembobot *Welsch* dan *Tukey bisquare* ditinjau dari nilai *standard error* dan nilai *adj R-square* pada data Indeks Pembangunan Manusia menurut provinsi di Indonesia tahun 2015?

### D. Tujuan

Tujuan dalam penulisan skripsi ini adalah:

1. Menjelaskan prosedur analisis regresi *robust* estimasi-S menggunakan pembobot *Welsch* dan *Tukey bisquare* dalam mengatasi masalah *outlier* pada analisis regresi.
2. Menentukan bentuk model regresi *robust* estimasi-S dengan pembobot *Welsch* dan *Tukey bisquare* dalam mengatasi *outlier* pada data Indeks Pembangunan Manusia menurut provinsi di Indonesia tahun 2015.
3. Menjelaskan keefektifan regresi *robust* estimasi-S dengan menggunakan pembobot *Welsch* dan *Tukey bisquare* ditinjau dari nilai *standard error* dan

*adj R-square* pada data Indeks Pembangunan Manusia menurut provinsi di Indonesia tahun 2015.

#### **E. Manfaat**

Manfaat dari penulisan skripsi ini adalah:

##### 1. Bagi Penulis

Menambah pengetahuan tentang metode regresi *robust*, khususnya regresi *robust* estimasi-S dengan menggunakan pembobot *Welsch* dan *Tukey bisquare* sebagai metode pembobot parameter alternatif untuk mengatasi masalah *outlier* dalam analisis regresi linear.

##### 2. Bagi Mahasiswa

Sebagai salah satu bahan tinjauan pustaka dalam mempelajari regresi *robust* dengan pembobot parameter alternatif untuk mengatasi masalah *outlier* yang berguna bagi setiap pihak yang membutuhkan, khususnya bagi mahasiswa Program Studi Matematika.

##### 3. Bagi Perpustakaan Universitas Negeri Yogyakarta

Menambah koleksi bahan pustaka yang bermanfaat bagi Universitas Negeri Yogyakarta pada umumnya dan mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam pada khususnya.