

# BAB I

## PENDAHULUAN

### A. Latar Belakang

Analisis regresi merupakan metode analisis yang dapat digunakan untuk menganalisis data dan mengambil kesimpulan yang bermakna tentang hubungan ketergantungan variabel satu terhadap variabel lainnya. Apabila banyaknya variabel bebas hanya ada satu dan berpangkat satu maka disebut regresi linier sederhana, sedangkan apabila terdapat lebih dari satu variabel bebas yang setiap variabelnya berpangkat satu maka disebut regresi linier berganda.

Salah satu metode estimasi dalam regresi linier yang sering digunakan adalah metode kuadrat terkecil (MKT) atau sering disebut *Ordinary Least Squares* (OLS). Penggunaan metode ini membutuhkan beberapa asumsi klasik yang harus dipenuhi untuk menghasilkan model yang baik, yaitu model estimasi linier tidak bias terbaik atau *Best Linier Unbiased Estimator* (BLUE), yaitu tidak terdapat multikolinieritas, tidak terdapat heteroskedastisitas, dan tidak terdapat autokorelasi. Selain itu, tidak adanya pencilan (*outlier*) merupakan syarat penggunaan MKT untuk menghasilkan model yang baik. Masalah yang sering terjadi pada data amatan adalah adanya pencilan (*outlier*). Keberadaan pencilan pada suatu data dapat mengganggu proses analisis data, sehingga pendeteksian pencilan merupakan hal yang sangat penting untuk dilakukan.

Ada banyak cara untuk mendeteksi pencilan, dan penyajiannya dapat berupa grafik atau nilai. Adanya pencilan dalam data dapat mengakibatkan estimasi koefisien regresi yang diperoleh kurang efisien karena memberikan nilai *error* yang besar apabila menggunakan MKT. Sedangkan tindakan membuang begitu saja suatu pencilan bukanlah tindakan yang tepat karena adakalanya pencilan memberikan informasi yang cukup berarti. Oleh karena itu, diperlukan suatu estimasi yang lebih efisien dalam menangani suatu pencilan yang dikenal dengan regresi *robust*. Regresi *robust* merupakan metode regresi yang digunakan ketika distribusi dari sisaan tidak normal dan/atau adanya beberapa pencilan yang berpengaruh pada model (Ryan, 1997:150). Metode ini merupakan alat penting untuk menganalisa data yang dipengaruhi oleh pencilan sehingga dihasilkan model yang *robust* yaitu *resistance* terhadap pencilan. Menurut Huber (2009:8), suatu estimasi yang *resistance* adalah relatif tidak terpengaruh oleh perubahan besar pada bagian kecil data atau perubahan kecil pada bagian besar data. Akan tetapi, adakalanya pencilan memberikan informasi yang tidak bisa diberikan oleh titik data lainnya, misalnya karena pencilan timbul dari kombinasi keadaan yang tidak biasa yang mungkin saja sangat penting dan perlu diselidiki lebih jauh.

Regresi *robust* adalah metode yang penting untuk menganalisis data yang mengandung pencilan. Untuk menaksir parameter regresi *robust* dapat digunakan beberapa metode antara lain: *S*-estimator, *M*-estimator, *MM*-estimator, *Least Median of Squares* (LMS)-estimator, dan *Least Trimmed Squares* (LTS)-estimator (Chen, 2002:1).

Metode estimasi-*S* pertama kali dikembangkan oleh Rousseeuw dan Yohai (1984). Metode ini memiliki *breakdown point* yang tinggi. *Breakdown point* yaitu bagian terkecil data yang menyimpang yang menyebabkan nilai estimator menjadi tidak berguna (Montgomery, Peck & Vining, 2006:385). Disebut estimasi-*S* karena mengestimasi berdasarkan skala. Skala yang digunakan adalah standar deviasi sisaan. Metode estimasi-*M* merupakan salah satu metode regresi *robust* yang penting dan luas digunakan. Selain itu, metode estimasi-*M* mempunyai efisiensi yang tinggi. Besarnya efisiensi dapat ditentukan melalui nilai ragam/varians yang minimum suatu penduga. Metode yang lainnya yaitu metode estimasi-*MM*. Metode estimasi-*MM* merupakan metode estimasi *robust* yang memadukan estimasi nilai *high breakdown point* dan *high efficiency*. Ketiga metode estimasi tersebut dapat menunjukkan model regresi yang optimal dalam memprediksi produksi kedelai di Indonesia. Setiap data akan memberikan optimasi yang berbeda.

. Selain padi dan jagung, kedelai merupakan komoditi pangan penting dalam kebutuhan pokok bagi masyarakat Indonesia. Kedelai bagi industri pengolahan pangan di Indonesia banyak digunakan sebagai bahan dasar pangan dalam pembuatan tempe, tahu, kecap dan susu. Komoditas kedelai ini terus meningkat dari tahun ke tahun karena mempunyai banyak fungsi, baik sebagai bahan pangan utama, pakan ternak, maupun sebagai bahan baku industri skala besar hingga kecil atau rumah tangga. Produksi kedelai dalam negeri hanya mampu memenuhi kebutuhan sekitar 30% dan setidaknya 70% harus impor (Satria, 2015).

Direktur Budidaya Aneka Kacang dan Umbi-umbian Ditjen Tanaman Pangan Kementerian Pertanian menyatakan bahwa setelah sempat mengalami swasembada kedelai pada 1992, para petani dihantam oleh regulasi impor yang dikeluarkan pada tahun 1998 ([www.distanhor.rokanhulukab.go.id](http://www.distanhor.rokanhulukab.go.id)). Kementerian Pertanian terus melakukan upaya untuk mewujudkan kembali swasembada tersebut, salah satunya dengan mendorong perluasan lahan tanam kedelai. Selain itu, kebijakan harga pangan merupakan salah satu instrumen penting dalam menciptakan ketahanan pangan nasional. Diperlukannya usaha untuk mengetahui ketersediaan kedelai pada tahun yang akan datang guna pentingnya memenuhi kebutuhan pangan.

Oleh karena itu diambil data produksi kedelai di Indonesia terbaru (pada tahun 2014) beserta faktor-faktor yang diduga mempengaruhinya, dengan mempertimbangkan keberadaan pencilan. Ada beberapa metode dalam bidang statistik yang dapat digunakan untuk memodelkan prediksi produksi kedelai serta menyelidiki faktor-faktor yang mempengaruhinya. Semua metode statistik secara eksplisit dan implisit bergantung pada sejumlah asumsi. Asumsi ini bertujuan untuk menentukan apa yang menjadi dugaan tentang analisis data atau masalah pemodelan statistik yang sedang dihadapi. Asumsi yang sering digunakan adalah data berdistribusi normal.

Akan tetapi meski data berdistribusi normal, sangat rawan terjadinya pencilan (*outlier*) pada data yang telah diperoleh. Pencilan dapat memiliki pengaruh besar pada metode klasik statistik yang optimal dengan asumsi normalitas atau linieritas.

Mengingat bahwa pencilan merupakan data yang kurang baik, mengeluarkannya dari pengolahan data merupakan hal yang tidak dibenarkan, karena pencilan sangat memungkinkan memiliki informasi tidak terduga yang cukup relevan. Sehingga digunakan metode alternatif untuk memperkecil nilai *error* karena pengaruh keberadaan pencilan tersebut tanpa harus menghilangkannya, yaitu dengan regresi *robust*.

Metode dalam regresi *robust* memiliki metode estimasi beragam dengan tes terkait (uji asumsi dan uji *outlier*). Metode ini dapat digunakan pada sebagian besar data, termasuk data yang tidak mengandung pencilan. Akan tetapi metode ini memberikan model regresi dengan efisiensi atau nilai *error* yang sama dengan MKT apabila data yang diolah tidak mengandung pencilan. Pada penelitian ini, akan dibahas regresi *robust* dengan estimasi-*M*, estimasi-*S*, dan estimasi-*MM*, serta ditunjukkan estimasi yang paling efektif dengan membandingkan nilai koefisien determinasi dari ketiga metode tersebut untuk memodelkan data pada produksi kedelai di Indonesia.

## **B. Rumusan Masalah**

Berdasarkan latar belakang di atas maka rumusan masalah yang akan dibahas adalah sebagai berikut:

1. Bagaimana alur menentukan model regresi *robust* dengan estimasi-*M*, estimasi-*S*, dan estimasi-*MM*?
2. Bagaimana model regresi *robust* dengan estimasi-*M*, estimasi-*S*, dan estimasi-*MM* untuk memprediksi produksi kedelai di Indonesia menggunakan?

3. Bagaimana perbandingan keefektifan model regresi *robust* antara metode estimasi-*M*, estimasi-*S*, dan estimasi-*MM* untuk memprediksi produksi kedelai di Indonesia?

### **C. Tujuan Penelitian**

Tujuan penulisan penelitian ini menurut rumusan masalah di atas adalah:

1. Menjelaskan alur regresi *robust* dengan estimasi-*M*, estimasi-*S*, dan estimasi-*MM*.
2. Menjelaskan model regresi *robust* untuk memprediksi produksi kedelai di Indonesia menggunakan estimasi-*M*, estimasi-*S*, dan estimasi-*MM*.
3. Menjelaskan perbandingan keefektifan model regresi *robust* antara estimasi-*M*, estimasi-*S*, dan estimasi-*MM* untuk memprediksi produksi kedelai di Indonesia.

### **D. Manfaat Penelitian**

Manfaat yang dapat diperoleh dari penelitian ini adalah:

1. Bagi Penulis

Penulisan tugas akhir ini dapat menambah pengetahuan mengenai model regresi *robust* untuk memprediksi produksi kedelai di Indonesia menggunakan estimasi-*M*, estimasi-*S*, dan estimasi-*MM*, serta mengetahui perbandingan dari ketiga metode estimasi untuk memprediksi produksi kedelai di Indonesia.

2. Bagi pembaca

Sebagai salah satu bahan dalam mempelajari model-model regresi *robust* serta mengetahui model optimumnya dan diharapkan penelitian

ini dapat dijadikan sebagai referensi untuk penelitian selanjutnya. Selain itu, Penelitian ini diharapkan dapat membantu pemerintah untuk melakukan upaya dalam meningkatkan produksi kedelai di Indonesia.

3. Bagi perpustakaan Universitas Negeri Yogyakarta

Penulisan tugas akhir ini juga bermanfaat dalam menambah koleksi bahan pustaka yang bermanfaat bagi Universitas Negeri Yogyakarta pada umumnya dan mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam pada khususnya.