

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Indeks Pembangunan Manusia (IPM) merupakan indikator penting untuk mengukur keberhasilan dalam upaya membangun kualitas hidup manusia (masyarakat/penduduk). IPM menjelaskan bagaimana penduduk dapat mengakses hasil pembangunan dalam memperoleh pendapatan, kesehatan, pendidikan, dan sebagainya.

IPM dibentuk oleh tiga dimensi dasar, yaitu umur panjang dan hidup sehat (*a long and healthy life*), pengetahuan (*knowledge*), dan standar hidup layak (*decent standard of living*). Umur panjang dan hidup sehat digambarkan oleh Angka Harapan Hidup saat lahir, pengetahuan diukur melalui indikator Rata-rata Lama Sekolah dan Harapan Lama Sekolah, serta standar hidup yang layak digambarkan oleh pengeluaran per kapita disesuaikan (*purchasing power parity*) (Badan Pusat Statistik, 2017).

IPM menjadi salah satu indikator apakah sebuah negara adalah negara maju, negara berkembang atau negara terbelakang dan juga mengukur pengaruh dari kebijaksanaan ekonomi terhadap kualitas hidup. Untuk mengetahui bentuk hubungan antara variabel IPM dengan variabel dimensi dasarnya maka dapat digunakan salah satu metode statistika yaitu analisis regresi.

Analisis regresi merupakan studi mengenai ketergantungan satu variabel yaitu variabel *dependent*, terhadap satu atau dua variabel lainnya, yaitu variabel *independent* (Gujarati & Porter, 2009). Regresi linier yang terdiri dari satu variabel *dependent* dan satu variabel *independent* disebut regresi linier sederhana, sedangkan regresi linier yang terdiri dari satu variabel *dependent* dan beberapa variabel *independent* disebut regresi linier berganda. Jika Y variabel *dependent* dan X_1, X_2, \dots, X_K ($k \geq 2$) variabel *independent*, maka model regresi linear berganda Y atas X_1, X_2, \dots, X_K secara umum dapat dinyatakan sebagai $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i$, dengan $\beta_0, \beta_1, \dots, \beta_k$ adalah parameter-parameter regresi dan ε_i adalah sisaan yang berdistribusi normal dengan mean nol dan variansi konstan (Sudjana, 2003).

Metode yang biasa digunakan untuk mengestimasi parameter regresi adalah Metode Kuadrat Terkecil (MKT) atau sering disebut *Ordinary Least Square* (OLS). Penggunaan metode ini membutuhkan beberapa asumsi yang harus dipenuhi agar menghasilkan model dengan parameter yang bersifat *Best Linier Unbiased Estimator* (BLUE) yaitu normalitas, tidak terdapat multikolinearitas, homokedastisitas, dan tidak terdapat autokorelasi.

Pada data hasil pengamatan seringkali ditemukan data pencilan, contohnya pada data Indeks Pembangunan Manusia Berdasarkan Provinsi di Indonesia tahun 2016. Data pencilan (*outlier*) adalah data pengamatan yang tidak mengikuti sebagian besar pola data (Rousseeuw & Zomeren, 1990). Data pencilan dapat ditanggulangi dengan cara menghapus atau menghilangkan data pengamatan tersebut. Akan tetapi, tindakan menghapus atau menghilangkan data pencilan bukanlah suatu tindakan yang efektif karena adakalanya data pencilan dapat memberikan informasi yang berarti. Keberadaan data pencilan akan menyebabkan residual atau sisaan yang besar dari model yang terbentuk (Paludi, Identifikasi dan Pengaruh Keberadaan Data Pencilan (Outlier), 2009). Mengestimasi data yang mengandung pencilan menggunakan Metode Kuadrat Terkecil akan memberikan nilai residual yang besar. Oleh karena itu, diperlukan suatu teknik analisis lain yang lebih efisien dalam menangani data hasil pengamatan yang mengandung data pencilan. Salah satu teknik analisis data yang dapat digunakan adalah regresi *robust*.

Regresi *robust* merupakan teknik analisis regresi alternatif dari Metode Kuadrat Terkecil yang digunakan ketika distribusi dari sisaan tidak normal atau adanya beberapa pencilan yang berpengaruh pada model (Ryan, 2003). Terdapat beberapa metode estimasi untuk menentukan penaksir parameter regresi *robust* yaitu metode estimasi-M, metode estimasi *Least Median Square* (LMS), metode estimasi *Least Trimmed Square* (LTS), metode estimasi-S, dan metode estimasi-MM (Chen, 2002).

Penelitian sebelumnya, regresi *robust* untuk data IPM telah dilakukan oleh Ma'rufatu Amaliyah pada tahun 2017 yang meneliti analisis regresi *robust* estimasi-S dalam mengatasi *outlier* dan Zuni Setiarini pada tahun 2016 meneliti analisis regresi *robust* estimasi-S menggunakan pembobot *Welsch* dan *Tukey*

Bisquare. Metode regresi *robust* estimasi-S memiliki *breakdown point* yang sama dengan estimasi-LMS dan estimasi-LTS yaitu sebesar 50%. Berdasarkan hal tersebut, penulis tertarik untuk meneliti regresi *robust* manakah yang terbaik antara estimasi-LMS dan estimasi-LTS pada studi kasus Indeks Pembangunan Manusia Berdasarkan Provinsi di Indonesia Tahun 2016 serta pembuatan program aplikasinya dengan menggunakan *Excel Visual Basic for Application* (VBA). Untuk mengetahui metode estimasi mana yang terbaik, maka kriteria yang digunakan adalah dengan membandingkan nilai koefisien determinasi atau *R-Square* nya, karena keseluruhan *goodness of fit* dari model diukur dari koefisien determinasi (Gujarati & Porter, 2009).

1.2. Rumusan Masalah

Berdasarkan uraian dari latar belakang di atas, maka permasalahan pada skripsi ini adalah sebagai berikut:

1. Bagaimana konstruksi program pemodelan regresi *robust* untuk estimasi-LMS dan estimasi-LTS dengan menggunakan *Excel Visual Basic for Application* (VBA) ?
2. Estimasi mana yang memberikan model regresi *robust* terbaik untuk data Indeks Pembangunan Manusia Berdasarkan Provinsi di Indonesia tahun 2016?
3. Bagaimana hasil regresi *robust* untuk estimasi-LMS dan estimasi-LTS apabila dilakukan simulasi pada pencilan data?

1.3. Batasan Masalah

Berdasarkan latar belakang masalah dan rumusan masalah yang telah diuraikan, maka penulis membatasi kajian pada tulisan ini yaitu sebagai berikut:

1. Data yang digunakan adalah data Indeks Pembangunan Manusia di Indonesia Berdasarkan Provinsi di Indonesia tahun 2016 yang diperoleh dari publikasi capaian pembangunan manusia di Indonesia dengan judul “Indeks Pembangunan Manusia 2016” yang diterbitkan oleh Badan Pusat Statistik.

2. Program pemodelan regresi *robust* dengan estimasi LMS dan LTS ini hanya untuk jumlah data sebanyak 34 sampel, satu variabel *dependent* dan empat variabel *independent*.

1.4. Tujuan Penulisan

Tujuan dari penulisan ini adalah sebagai berikut:

1. Menghasilkan konstruksi program pemodelan regresi *robust* untuk estimasi-LMS dan estimasi-LTS dengan menggunakan *Excel Visual Basic for Application* (VBA)
2. Mengetahui model regresi *robust* terbaik untuk data Indeks Pembangunan Manusia Berdasarkan Provinsi di Indonesia tahun 2016
3. Mengetahui model regresi *robust* untuk estimasi-LMS dan estimasi-LTS yang dilakukan simulasi pada pencilan data nya

1.5. Manfaat Penulisan

Penulisan skripsi ini diharapkan dapat memberikan manfaat baik secara teoritis maupun secara praktis.

Manfaat teoritis, dapat menambah wawasan dalam bidang statistika mengenai teknik analisis regresi dalam menangani data pencilan yaitu regresi *robust* dengan metode estimasi-LMS dan estimasi-LTS.

Manfaat praktis, dapat menjadi bahan pertimbangan atau referensi untuk penelitian selanjutnya khususnya tentang regresi *robust*. Selain itu program yang telah dibuat dapat digunakan sebagai alat untuk mengetahui model regresi *robust* estimasi-LMS dan estimasi-LTS pada data IPM Berdasarkan provinsi Indonesia untuk tahun-tahun berikutnya.