

BAB I

PENDAHULUAN

1.1 Latar Belakang

Statistika adalah salah satu cabang ilmu yang mempelajari prosedur-prosedur yang digunakan dalam pengumpulan, penyajian, analisis dan interpretasi data. Statistika inferensia adalah salah satu cabang statistika yang mempelajari semua metode yang berhubungan dengan analisis sebahagian data untuk kemudian melakukan peramalan atau penarikan kesimpulan mengenai keseluruhan data induknya. Dalam statistika inferensia, data induk yang merupakan himpunan keseluruhan data yang menjadi perhatian kita dinamakan *populasi*. Sebarang nilai yang menjadi ciri dari suatu populasi dinamakan *parameter*. Dalam statistika inferensia, nilai parameter inilah yang ingin diketahui nilainya.

Dalam kebanyakan kondisi, nilai parameter ini tidak dapat diketahui. Hal ini dapat terjadi karena sulitnya mendapatkan keseluruhan nilai populasi dikarenakan besarnya ukuran populasi, besarnya biaya dan waktu yang diperlukan untuk mendapatkan keseluruhan data populasi tersebut serta penyebab lainnya. Sehingga yang dapat dilakukan hanyalah menduga nilai parameter ini.

Pendugaan nilai parameter ini dilakukan dengan mengambil sebagian data dari populasi tersebut, yang biasa diistilahkan dengan contoh (*sample*). Sebarang nilai yang menjadi ciri dari sampel dinamakan *statistik*. Nilai statistik inilah yang akan digunakan untuk mengambil kesimpulan mengenai parameter populasi.

Masalah pendugaan parameter adalah bagian dari statistika inferensia yang merupakan suatu cara untuk memprediksi karakteristik dari suatu populasi berdasarkan sampel yang diambil. Secara umum pendugaan parameter digolongkan menjadi dua yaitu pendugaan titik dan pendugaan selang. Beberapa metode pendugaan titik yang digunakan untuk menduga parameter diantaranya adalah metode momen, metode kuadrat terkecil, metode MLE (*Maximum Likelihood Estimation*) dan metode Bayes. Metode momen menduga parameter dengan cara menyamakan momen ke- k sampel dengan momen ke- k populasi dan menyelesaikan sistem persamaan yang dihasilkan [1]. Sedangkan metode kuadrat terkecil prinsip kerjanya adalah meminimumkan jumlah kuadrat penyimpangan atau *error* nilai-nilai observasi terhadap rata-ratanya. Selanjutnya metode MLE merupakan suatu metode pendugaan parameter yang memaksimalkan fungsi *likelihood*. Kemudian metode Bayes merupakan metode pendugaan yang menggabungkan distribusi prior dan distribusi sampel. Distribusi prior adalah distribusi awal yang memberi informasi tentang parameter. Distribusi sampel yang digabung dengan distribusi prior akan menghasilkan suatu distribusi baru yaitu distribusi posterior yang menyatakan derajat keyakinan seseorang mengenai letak parameter setelah sampel diamati [8].

Analisis regresi adalah suatu teknik statistika yang digunakan untuk menarik kesimpulan mengenai hubungan antara satu variabel tak bebas (Y) dengan satu atau lebih variabel bebas (X) dalam suatu sistem. Hubungan antara variabel-variabel tersebut biasanya dinyatakan dalam suatu model yang biasanya diistilahkan sebagai model regresi yang secara umum dinyatakan sebagai $Y = f(X) + \varepsilon$, ε adalah

komponen galat. Model tersebut menghubungkan variabel bebas dan tak bebas melalui suatu parameter yang dinamakan sebagai parameter regresi, dinotasikan dengan β .

Dalam pemodelan regresi, metode pendugaan yang sering digunakan untuk mendapatkan nilai-nilai penduga parameter adalah metode kuadrat terkecil (*Ordinary Least Square*). Pada penggunaan metode kuadrat terkecil diperlukan beberapa asumsi klasik yang harus dipenuhi oleh komponen galat (ε) dalam model yang dihasilkan, diantaranya galat harus menyebar normal, kehomogenan ragam galat dan tidak terjadi autokorelasi. Jika asumsi terpenuhi maka penduga parameter yang diperoleh bersifat BLUE (*Best Linier Unbiased Estimator*). Namun jika ada asumsi yang tidak terpenuhi, maka metode kuadrat terkecil akan memberi kesimpulan yang bersifat kurang baik atau nilai penduga parameternya bersifat bias sehingga menyebabkan hasil yang diperoleh menjadi tidak valid. Pada kondisi seperti ini, diperlukan suatu metode pendugaan lain yang kekar terhadap ketidak terpenuhan asumsi tersebut. Metode *bootstrap* merupakan salah satu pilihan yang dapat digunakan untuk menghasilkan model yang lebih baik.

Metode *bootstrap* diperkenalkan oleh Bradley Efron pada tahun 1979. *Bootstrap* adalah suatu metode yang sangat umum untuk membuat ukuran ketidakpastian dan bias, khususnya pada penduga parameter dari peubah-peubah yang tersebar secara bebas dan identik (*Independently Identically Distributed Variable/variable i.i.d*). Metode *bootstrap* dapat bekerja tanpa membutuhkan asumsi mengenai distribusi karena pada pendekatan ini, sampel asli yang diperlakukan sebagai populasi. Sebagai pengganti, *bootstrap* menggunakan distribusi empiris untuk menduga sebaran penarikan contoh tersebut. Jadi jika penyelesaian analitik tidak

mungkin dilakukan karena asumsi mengenai sebaran tidak dipenuhi maka metode *bootstrap* masih memungkinkan seseorang untuk melakukan suatu inferensi statistik. Selain itu metode *bootstrap* didasarkan pada simulasi data yang sering kali digunakan untuk keperluan inferensi statistik.

Adapun ide dasar dari *bootstrap* memperlakukan data sampel asli yang berukuran n sebagai populasi. Kemudian dilakukan pengambilan sampel ulang (*resample*) yang juga berukuran n dari sampel asli secara acak dan dengan pengembalian. Dengan cara *resampling* seperti ini akan diperoleh sampel yang berukuran sama dengan sampel asli dan dengan pengembalian memungkinkan diperoleh sampel ulang yang sama, lebih dari sekali dan mungkin saja berbeda dengan sampel asli. Selanjutnya pengambilan kesimpulan mengenai parameter populasi didasarkan sampel ulang ini.

Dalam metode *bootstrap* biasanya *resampling* tersebut dilakukan sebanyak B kali ulangan *bootstrap*. Secara intuisi, semakin banyak ulangan *bootstrap* yang dilakukan maka dugaan yang dilakukan akan semakin baik, sehingga idealnya, untuk mendapatkan hasil yang paling baik, maka harus dilakukan replikasi *bootstrap* sebanyak mungkin. Namun, menambah satu ulangan *bootstrap* tentu memerlukan tambahan biaya analisis. Padahal, dengan menambahkan satu ulangan, boleh jadi perubahan kebaikan hasil pendugaan tidak terlalu banyak.

Suatu hal yang juga menarik untuk diketahui adalah berapa kali ulangan *bootstrap* sebaiknya dilakukan untuk menduga parameter regresi sehingga dapat dihasilkan penduga yang baik dan dengan biaya yang kecil. Metode yang dapat dilakukan untuk mengetahui banyaknya ulangan *bootstrap* ini adalah metode simulasi

Monte Carlo, yaitu suatu algoritma komputasi yang dilakukan untuk menirukan berbagai perilaku sistem matematika. Metode ini seringkali digunakan untuk menemukan solusi problem matematis yang sulit dipecahkan secara teori.

1.2 Rumusan Masalah

Dari penjelasan latar belakang di atas, permasalahan pada penelitian ini adalah berapa banyaknya ulangan *bootstrap* (B) untuk mendapatkan penduga parameter yang baik?

1.3 Pembatasan Masalah

Penelitian ini dibatasi pada penentuan banyaknya ulangan *bootstrap* pada ukuran contoh $(n) = 10, 15, 20, 25, 30, 35, 40, 45, 50$ untuk menduga parameter regresi linier sederhana dengan menggunakan metode simulasi Monte Carlo

1.4 Tujuan

Pada penelitian akan ditentukan banyaknya ulangan *bootstrap* untuk menduga β_0 dan β_1 pada analisis regresi linier sederhana dalam kondisi galat menyebar menurut sebaran Khi-kuadrat, sebaran Eksponensial, sebaran Gamma, sebaran-t, sebaran-F, sebaran Normal dengan ragam yang tidak sama dan untuk ukuran contoh $(n) = 10, 15, 20, 25, 30, 35, 40, 45, 50$.

1.5 Sistematika Penulisan

Skripsi ini dibagi ke dalam 5 bab dan masing–masing bab terdiri dari sub bab yang berisi penjelasan. Pada Bab I. Pendahuluan diuraikan tentang latar belakang masalah, perumusan masalah, pembatasan masalah, tujuan dan sistematika penulisan skripsi ini. Konsep dasar mengenai *bootstrap* serta beberapa teori pendukung untuk menyelesaikan permasalahan skripsi ini disajikan dalam Bab II sebagai landasan teori. Pada Bab III berisikan tentang data dan langkah-langkah penelitian yang merupakan langkah atau proses untuk memperoleh hasil. Analisis dan pembahasan akan dipaparkan dalam Bab IV yang merupakan hasil dari langkah-langkah pada Bab III. Penulisan skripsi ini diakhiri dengan kesimpulan yang dipaparkan dalam Bab V.