

BAB I PENDAHULUAN

1.1. Latar Belakang Masalah

Dalam sebuah penelitian di bidang statistika, peneliti akan berhubungan dengan data pengamatan, baik data kualitatif atau data kuantitatif yang akan diproses dalam bentuk angka berdasarkan sistem yang sedang ditelitinya. Sistem dalam keadaan yang sesungguhnya dapat dibuat dalam sebuah model pengamatan. Jika peneliti ingin mengetahui bagaimana hubungan antara faktor-faktor yang mempengaruhi sebuah sistem (peubah model) dan ingin meneliti hubungan antara peubah-peubah tersebut maka peneliti akan menggunakan sebuah model yang dapat menyatakan hubungan fungsional antara peubah dalam sistem. Model tersebut berupa fungsi yang biasanya dibentuk dengan menggunakan analisis regresi.

Analisis regresi dapat dibedakan berdasarkan jumlah peubah yang diamati menjadi analisis regresi sederhana dan analisis regresi berganda. Untuk tiap jenis analisis regresi tersebut dapat dibagi menjadi dua berdasarkan bentuk hubungannya yaitu linear dan non linear. Analisis regresi yang paling sederhana merujuk pada analisis hubungan linear antara dua peubah saja yakni hanya satu peubah bebas dan satu peubah terikat, analisis regresi inilah yang disebut sebagai analisis regresi linear sederhana.

Parameter-parameter dalam analisis regresi dapat dicari penduga parameter-nya dengan beberapa metode, metode yang paling umum digunakan adalah Metode Kuadrat Terkecil (MKT). Namun dalam penggunaannya MKT akan bekerja dengan baik apabila asumsi-asumsi pada analisis regresi telah terpenuhi. Asumsi tersebut antara lain: normalitas, homogenitas, tidak terjadi autokorelasi, heteroskedastisitas dan multikolinearitas. Untuk asumsi normalitas, asumsi ini tidak akan terpenuhi apabila pada data pengamatan terdapat pencilan (*outlier*). Paludi (2009) menyatakan bahwa ada dua hal yang menyebabkan model tidak cocok dengan data sampel atau data pengamatan, pertama modelnya belum tepat dan kedua, adanya pencilan (*outlier*) yang tidak mengikuti pola umum data pengamatan.

Pencilan (*outlier*) dapat diartikan sebagai suatu datum pada data pengamatan yang jauh (*ekstrim*) dari ukuran pemusatan data. Astari.(2014) menyatakan bahwa pencilan (*outlier*) tidak dapat dibuang, karena ada kemungkinan pencilan dapat memberikan informasi penting yang tidak ada pada datum lainnya pada data pengamatan yang dikumpulkan. Disamping itu, bila pencilan terdapat pada data pengamatan maka MKT tidak akan dapat bekerja dengan baik sehingga penduga parameter secara umum akan memberikan sifat bias yang membuat penduga parameter tidak lagi memiliki sifat *Best Linear Unbiased Estimator (BLUE)*. BLUE dalam analisis regresi menggambarkan sifat penduga parameter yang baik, penduga parameter dikatakan baik apabila penduga parameter memiliki bias yang kecil. Linear berarti penduga parameter tidak berbentuk eksponen ataupun berupa hasil perkalian atau pembagian dengan parameter yang lain dan dapat berupa penduga parameter yang tidak bias artinya nilai ekspektasi dari penduga parameter sama dengan nilai parameter populasinya.

Sesungguhnya pencilan lebih sering terjadi akibat kesalahan dalam melakukan pengamatan baik salah dalam input data atau kesalahan yang tidak disengaja lainnya. Tetapi ada kalanya data pencilan tersebut adalah murni dari pengamatan dalam kasus yang jarang atau langka. Dimana pencilan tersebut menyimpan informasi tertentu yang tidak dapat diganti oleh data yang lain. Dengan demikian kegiatan mengeluarkan atau membuang data pencilan tidak seharusnya dilakukan tanpa alasan yang tepat dan jika hal ini terjadi sebisa mungkin harus digunakan alternatif atau cara lain tanpa harus mengeluarkan atau membuang data pencilan tersebut.

Selain akibat dari pelanggaran asumsi karena adanya pencilan, bias pada penduga parameter statistika dapat juga terjadi karena hal-hal berikut: kesalahan dalam memilih subyek penelitian, informasi yang salah tentang subyek penelitian, kegagalan mengontrol variabel luar selain faktor penelitian yang berpengaruh dan kesalahan dalam pengumpulan data.

Cara alternatif yang dapat digunakan dalam masalah bias akibat pencilan ini adalah dengan menggunakan *bootstrap*. Dimana *bootstrap* merupakan metode resampling data yang dapat dilakukan hanya berdasarkan data yang dimiliki dan dapat digunakan dalam pendugaan parameter serta penyusunan selang kepercayaan tanpa perlu mengetahui distribusi populasi dari sampel yang dimiliki. Metode *bootstrap* diperkenalkan oleh Efron pada 1979. Nama *bootstrap* sendiri diambil

dari sebuah istilah “*to pull one self up by one’s bootstraps*” yang dapat diartikan berusaha dengan sumber daya yang minimal. Dalam permasalahan statistika sumber daya yang minimal dapat diartikan sebagai data yang sedikit, data yang menyimpang dari asumsi tertentu bahkan data yang tidak memiliki asumsi apapun tentang distribusinya. Ditegaskan oleh pernyataan Suparman (2012) bahwa tujuan utama penggunaan *bootstrap* adalah untuk memperoleh penduga (estimasi) yang sebaik baiknya berdasarkan data yang minimal dengan bantuan program komputer. Pendekatan pada *bootstrap* ini menggunakan metode pengambilan sampel berulang (*resample*) dari data pengamatan yang telah ada.

Dalam konteks regresi dan untuk menduga parameter dalam analisis regresi dengan melakukan *resampling* hanya berdasarkan sampel yang sudah ada, Shao dan Tu (1995) mengungkapkan bahwa metode *bootstrap* dapat dilakukan dengan beberapa cara yaitu:

1. *Bootstrap* berdasarkan residual (*Bootstrap based on residual*) atau dikenal dengan *resample fixed-x* atau model residual
2. *Bootstrap* data berpasangan (*Paired Bootstrap*) atau dikenal dengan *resample random-x* atau model korelasi

Bootstrap berdasarkan residual merupakan metode *resampling* yang proses *resamplingnya* dilakukan pada data residual yang diperoleh dari model analisis regresi sedangkan metode *bootstrap* data berpasangan merupakan metode *resampling* yang dilakukan berdasarkan hubungan pada pasangan variabel bebas dan variabel terikat dengan mempertahankan korelasi antara variabel bebas dan variabel terikat pada model analisis regresi.

Astari (2014) telah melakukan penelitian tentang mengatasi bias penduga parameter regresi linear sederhana dengan menggunakan metode *bootstrap* residual, dalam penelitian astari data yang digunakan adalah data simulasi yang dibangkitkan dengan program komputer. Kemudian menggunakan *bootstrap* residual dengan jumlah perulangan tertentu yang menghasilkan penduga parameter yang memiliki kriteria penduga parameter yang mendekati parameter populasinya (memenuhi sifat BLUE). Dengan menggunakan data hasil penelitian Astari (2014) sebagai pembandingan dalam penelitian ini, penulis mencoba menggunakan metode *bootstrap* yang kedua yaitu dengan menggunakan *bootstrap* data berpasangan (*paired bootstrap*). *Bootstrap* data berpasangan (*paired bootstrap*) memiliki prinsip kerja yang hampir sama dengan *bootstrap* residual. Tetapi, *bootstrap* data berpasangan

bekerja dengan cara *resampling* setiap nilai pasangan data dalam model regresi dan kemudian diharapkan menghasilkan penduga parameter yang juga mendekati parameter populasi seperti pada penduga parameter dengan menggunakan *bootstrap residual*.

1.2. Rumusan Masalah

Berdasarkan latar belakang masalah yang dikemukakan sebelumnya, permasalahan yang diangkat dalam penelitian ini antara lain :

1. Adanya suatu pencilan data yang mengakibatkan pelanggaran asumsi normalitas data sehingga penduga parameter regresi linear sederhana akan bersifat bias dimana data pencilan tidak dapat dibuang karena mengandung informasi penting yang tidak dapat digantikan oleh data yang lain.
2. Bagaimana mengatasi bias yang terjadi pada penduga parameter regresi linear sederhana yang disebabkan karena pelanggaran asumsi normalitas akibat adanya pencilan pada data tersebut ?
3. Apakah dengan menggunakan metode *bootstrap* data berpasangan dapat mengatasi masalah bias pada penduga parameter tanpa harus membuang data?

1.3. Batasan Masalah

Mengingat luasnya masalah yang akan diteliti dan untuk menghindari kesimpangsiuran, maka peneliti membatasi masalah yaitu:

1. Parameter yang diduga adalah parameter pada analisis regresi linear sederhana dengan metode kuadrat terkecil (MKT) .
2. Masalah bias yang diteliti adalah bias yang ditimbulkan karena adanya pencilan data yang merusak asumsi kenormalan data.
3. Data yang akan digunakan adalah data simulasi yang akan dibangkitkan dengan bantuan software MATLAB.
4. Metode bootstap yang digunakan untuk mengatasi masalah ini adalah metode *bootstrap* data berpasangan.

1.4. Tujuan Penelitian

Tujuan penelitian ini adalah :

1. Mengetahui apakah adanya pencilaan pada data pengamatan dapat mengakibatkan data tidak memiliki sebaran normal.
2. Meneliti bagaimana bias dapat terjadi pada penduga parameter regresi linear sederhana.
3. Meneliti apakah bias yang terjadi pada penduga parameter akibat pelanggaran asumsi normalitas karena adanya pencilaan data dapat berkurang (direduksi) tanpa harus menghilangkan data tersebut.
4. Membandingkan hasil dari penduga parameter dengan menggunakan metode *Paired Bootstrap* dengan hasil penduga parameter dengan metode *Bootstrap Residual* pada penelitian sebelumnya.

1.5. Manfaat Penelitian

Dengan diadakannya penelitian ini diharapkan dapat memberi manfaat sebagai berikut :

1. Bagi peneliti: Menjadi media belajar dalam meneliti apakah *bootstrap* data berpasangan efektif dalam mengatasi bias pada penduga parameter regresi linear sederhana akibat pencilaan data.
2. Bagi pembaca: Memberikan informasi tentang *bootstrap* data berpasangan dalam mengatasi bias pada penduga parameter regresi linear sederhana akibat pencilaan data dan memberikan bahan masukan bagi penelitian selanjutnya yang berkaitan dengan masalah tersebut.