

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Regresi linier adalah metode statistika yang digunakan untuk membentuk model hubungan antara variabel terikat dengan satu atau lebih variabel bebas. Dalam pasal ini akan dibicarakan masalah pendugaan nilai peubah tak bebas  $Y$  berdasarkan peubah bebas  $X$  yang telah diketahui nilainya. Apabila dua variabel  $X$  dan  $Y$  mempunyai hubungan, maka nilai variabel  $X$  yang sudah diketahui dapat dipergunakan untuk memperkirakan atau menaksir  $Y$ . Didalam bukunya (Susanti 1990) menggambarkan hubungan antara variabel  $X$  dan  $Y$  sebagai hubungan antara banyaknya pengunjung toko (Variabel  $X$ ) dengan banyaknya pembeli toko (Variabel  $Y$ ), maka hubungan keduanya dibaca sebagai regresi  $X$  atas  $Y$ . Artinya pembeli toko sebagai variabel bebas dengan  $Y$  sebagai variabel terikat. (Gujarati 1979) menyebutkan bahwa terdapat dua jenis regresi, yaitu regresi linier sederhana dan regresi linier berganda. Regresi linier sederhana digunakan untuk menggambarkan hubungan antara satu variabel bebas ( $X$ ) dengan satu variabel tak bebas ( $Y$ ). Sedangkan jika variabel bebas ( $X$ ) yang digunakan lebih dari satu, maka modelnya adalah model regresi linier berganda. Analisis regresi memiliki 3 kegunaan, yaitu untuk tujuan deskripsi dari fenomena data atau kasus yang sedang diteliti, untuk tujuan kontrol, serta untuk tujuan prediksi.

Model regresi linier dikatakan baik atau cocok, jika dipenuhi asumsi-asumsi ideal, yakni tidak adanya autokorelasi, heteroskedastisitas dan multikolinearitas dan normalitas. Autokorelasi terjadi ketika nilai korelasi dalam persamaan regresi bernilai nol, heteroskedastisitas terjadi ketika terjadi ketaksamaan varian dari residual untuk semua pengamatan pada model regresi, multikolinearitas terjadi ketika ada korelasi diantara variabel bebas sedangkan normalitas terjadi ketika variabel pengganggu atau residual dalam persamaan regresi memiliki distribusi normal. Apabila salah satu dari asumsi ini tidak dipenuhi maka persamaan regresi yang dihasilkan dengan metode kuadrat terkecil akan bersifat tidak valid . Tetapi asumsi yang akan dibahas dalam penelitian ini adalah asumsi multikolinearitas.

Beberapa penelitian sebelumnya telah dilakukan tentang data curah hujan. Diantaranya adalah "Analisis Regresi Komponen Utama Untuk Mengatasi Masalah Multikolinieritas Dalam Analisis Regresi Linier Berganda"(G. L. Marcus 2012). Pada penelitian ini data yang mengandung multikolinieritas diatasi dengan menggunakan metode Principal Component Analysis. Kemudian variabel-variabel yang mempengaruhi nya disederhanakan sehingga diperoleh sebuah persamaan regresi yang tepat untuk dianalisis.

(Kusrini Setiawan 2010) mendefinisikan multikolinieritas merupakan adanya hubungan korelasi linear yang tinggi diantara beberapa atau semua variabel bebas dari model regresi berganda. Multikolinieritas terjadi ketika menentukan model regresi populasi ada kemungkinan bahwa dalam sampel tertentu, beberapa atau semua variabel sangat kolinear (mempunyai hubungan linear sempurna atau hampir sempurna). Jika antara variabel berkorelasi tinggi, pengujian hipotesis parameter berdasarkan metode kuadrat terkecil (Ordinary Least Square) memberikan hasil yang tidak valid (galat yang dihasilkan akan menjadi besar, variansi dan kovariansi parameter tidak berhingga). Variabel-variabel bebas yang seharusnya berpengaruh signifikan terhadap variabel tak bebas akan dinyatakan tidak nyata secara statistik, koefisien regresi dugaan yang dihasilkan bertentangan dengan kondisi aktual, penduga koefisien regresi bersifat tidak stabil sehingga mengakibatkan sulitnya menduga nilai-nilai variabel tak bebas yang tentunya akan mengakibatkan tidak akuratnya pada peramalan. (Yildiz 2018) juga dalam penelitiannya yang berjudul *A Study On The Comparison of the Effectiveness of the Jackknife Method in the Biased Estimators* menyebutkan bahwa multikolinieritas memberikan hasil yang tidak valid dengan metode kuadrat terkecil dan ketidakstabilan dalam estimasi.

Terdapat dua cara yang dapat digunakan untuk mengatasi masalah multikolinieritas, diantaranya adalah Metode regresi Ridge dan Metode Principal Component Analysis. Metode regresi Ridge bertujuan untuk memperkecil variansi estimator koefisien regresi, sedangkan Metode Principal Component Analysis bertujuan untuk menyederhanakan variabel yang diamati dengan cara mereduksi dimensinya. Metode regresi sebelumnya telah diteliti oleh Tri Rahmadani dalam penelitiannya yang berjudul *Efektifitas Jacknife Ridge Regression Method Dalam Mengatasi Masalah Multikolinieritas*. Dalam penelitiannya, ia berhasil memperkecil variansi estimator koefisien regresi dalam masalah multikolinieritas. Namun metode yang akan digunakan dalam penelitian ini Metode Principal Component Analysis.

Penelitian lainnya juga dilakukan oleh Leonardo (2011) yaitu "Analisis Regresi Komponen Utama Untuk Mengatasi Masalah Multikolinieritas". Dalam penelitian ini juga data tentang gas mileage pada 30 mobil dan faktor-faktor yang mempengaruhinya yang mengandung multikolinieritas juga diatasi dengan menggunakan Metode Principal Component Analysis sehingga diperoleh persamaan regresi yang sudah tepat untuk analisis data. Namun pada penelitian ini akan digunakan jumlah variabel dan sampel yang berbeda. Berdasarkan uraian latar belakang diatas, penulis tertarik ingin melakukan penelitian mengenai "Mengatasi Masalah Multikolinieritas Dalam Mengestimasi Parameter Regresi Linear Berganda Dengan Menggunakan Principal Component Analysis"

## **1.2 Rumusan Masalah**

Dari latar belakang masalah diatas yang menjadi rumusan masalah adalah :

1. Bagaimana mengestimasi parameter regresi linear berganda dengan menggunakan Principal Component Analysis (PCA)
2. Apakah dengan menggunakan metode Principal Component Analysis (PCA) dapat memberikan hasil yang valid ditunjukkan dengan variansi yang lebih efisien

## **1.3 Batasan Masalah**

Untuk menjaga agar penelitian ini lebih terarah dan tidak meluas, maka peneliti membuat batasan masalah sebagai berikut.

1. Masalah yang ingin diatasi adalah masalah multikolinieritas dengan menganggap bahwa asumsi klasik lain terpenuhi.
2. Metode yang digunakan dalam mengatasi masalah multikolinieritas adalah Principal Component Analysis (PCA)
3. Data yang digunakan adalah data sekunder.

## **1.4 Tujuan penelitian**

Adapun tujuan dari penelitian ini adalah sebagai berikut

1. Mengestimasi parameter regresi linear berganda yang mengandung multikolinearitas dengan cara mengkombinasikan Principal Component Analysis (PCA) menggunakan Metode Kuadrat Terkecil.
2. Menghilangkan korelasi yang kuat diantara variabel bebas ditunjukkan dengan Standar *error* yang dihasilkan lebih efisien.

## 1.5 Manfaat Penelitian

Dengan diadakannya penelitian ini diharapkan dapat memberi manfaat sebagai berikut:

1. Bagi penulis, penelitian ini berguna untuk menambah wawasan mengenai multikolinearitas dengan metode Principal Component Analysis (PCA)
2. Bagi para pembaca, sebagai tambahan informasi dan referensi bacaan menggunakan metode Principal Component Analysis (PCA) bagi yang hendak melakukan penelitian serupa.
3. Bagi akademik Memberikan tambahan ilmu dan wawasan yang baru tentang cara mendeteksi dan mengatasi multikolinearitas dengan menggunakan metode Principal Component Analysis (PCA).