

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Statistika merupakan disiplin ilmu yang berkaitan dengan pengumpulan, peringkasan, dan analisis data untuk membuat pernyataan tentang suatu fenomena. Data yang dikumpulkan dapat berupa data yang dikuantifikasi yaitu data yang diubah menjadi skala numerik[1]. Banyak aturan yang harus dipahami dengan baik untuk proses peringkasan data yang diperoleh. Data yang telah dikumpulkan kemudian dapat dikaitkan dengan representasi teoretis atau model dari data untuk dilakukan tahap selanjutnya yaitu analisis data. Dengan mengaplikasikan model pada data, pernyataan tentang suatu fenomena dapat dibuat.

Analisis regresi dan analisis deret waktu adalah dua metode dasar yang sering digunakan untuk menganalisis data. Analisis regresi adalah teknik statistik digunakan untuk menyimpulkan hubungan antara variabel dependen dan variabel independen[2]. Analisis regresi akan memberikan informasi mengenai variabel yang paling penting, variabel yang dapat diabaikan, dan bagaimana variabel saling mempengaruhi. Analisis regresi dilakukan dengan menggunakan data *cross section* dari unit penelitian. Data *cross section* adalah data yang dikumpulkan dengan mengamati banyak subjek pada satu titik atau periode waktu[3]. Analisis deret waktu adalah metode yang digunakan ketika berhubungan dengan data deret waktu. Data deret waktu adalah data berada dalam rangkaian periode waktu atau interval tertentu[1].

Analisis regresi data longitudinal merupakan gabungan dari analisis regresi dan analisis deret waktu. Tidak seperti data regresi, dengan data longitudinal subjek diamati dari waktu ke waktu dan tidak seperti data deret waktu, data longitudinal mengamati banyak subjek. Pengamatan dari waktu ke waktu pada data longitudinal mengakibatkan adanya korelasi dalam unit penelitian. Oleh karena itu pada data longitudinal pengamatan dari unit penelitian yang sama cenderung lebih serupa jika dibandingkan dengan penelitian dari dua unit penelitian yang berbeda. Perbedaan atau heterogenitas antar unit penelitian seringkali tidak dapat diamati. Salah satu

cara untuk mendeteksi heterogenitas antar unit penelitian adalah dengan cara menyertakan efek acak ke dalam model. Efek acak adalah nilai acak yang terkait dengan tingkat faktor acak, yaitu variabel klasifikasi yang dapat dianggap sebagai sampel acak dari populasi yang sedang diteliti[4].

Model linier yang terdiri dari efek tetap dan efek acak disebut sebagai model linier campuran. Model linier campuran atau *Linear Mixed Model* (LMM) adalah pengembangan dari model linier klasik yang dapat digunakan untuk menganalisis data berkorelasi seperti data longitudinal. Istilah *Mixed* dalam LMM mengacu pada efek tetap dan efek acak yang secara bersamaan digunakan dan dianalisis dalam model. Perbedaan antara efek tetap dan acak adalah efek tetap menggambarkan pola rata-rata sistematis, sementara efek acak menggambarkan pola korelasi antara pengukuran berulang dalam unit penelitian atau heterogenitas antar unit penelitian atau keduanya.

Untuk mendeteksi heterogenitas, data longitudinal dapat dilihat sebagai data dengan struktur bertingkat. Tingkat pertama atau tingkat yang paling rendah akan mendeteksi korelasi dalam unit penelitian. Tingkat kedua akan mendeteksi heterogenitas antar unit penelitian. Keunggulan dari model linier campuran dari model linier klasik adalah pada model linier campuran heterogenitas dalam unit penelitian maupun heterogenitas antar unit penelitian dapat dimodelkan. Pada tingkat satu, karena korelasi ini ada di antara pengamatan yang diambil dari waktu ke waktu, korelasi ini disebut juga sebagai korelasi serial. Korelasi serial dimodelkan dalam bentuk dimodelkan dalam bentuk varians dan kovarians dari efek acak yang mewakili dan dapat diringkas dalam bentuk matriks kovarians (\mathbf{R}). Pemodelan matriks kovarians ini akan memberikan peningkatan akurasi pada hasil estimasi untuk parameter efek tetap. Penelitian ini akan fokus kepada pembuatan model linier campuran dua tingkat yang menggambarkan dua tingkat data longitudinal dari data Indeks Pendidikan di Jawa Barat dari tahun 2014 sampai 2018 yang dibagi menjadi dua kelompok berdasarkan wilayah yaitu kota dan kabupaten dengan memodelkan tiga struktur kovarians yang berbeda. Dengan menyesuaikan dengan karakteristik data Indeks Pendidikan, struktur kovarians yang akan dimodelkan adalah struktur kovarians *Diagonal*, *Compound Symmetry* (CS), dan *Unstructured* (UN). Untuk mengeksplorasi heterogenitas dalam unit

penelitian pada data longitudinal, dilakukan studi kasus dari data Indeks Pendidikan di Provinsi Jawa Barat.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dituliskan sebelumnya, maka rumusan masalah yang akan diteliti adalah sebagai berikut:

1. Data longitudinal dapat dilihat sebagai struktur data bertingkat. Penelitian yang dilakukan dengan analisis regresi longitudinal dilakukan untuk mengetahui perubahan pada subjek yang diteliti dari waktu ke waktu dan perbedaan antara subjek-subjek yang diteliti beserta faktor yang mempengaruhinya. Faktor-faktor yang mempengaruhi dapat berupa faktor yang diketahui dari data atau faktor acak yang berasal dari variasi keacakan data. Permasalahan tersebut kemudian dimodelkan dengan model linier campuran dua tingkat. Pokok permasalahan dari skripsi ini adalah apa yang dimaksud dengan model linier campuran dua tingkat dan bagaimana cara membangun model linier campuran dua tingkat dari data longitudinal.
2. Dalam model linier campuran dua tingkat, masing-masing faktor memiliki parameter yang kemudian akan diestimasi. Proses estimasi pada suatu penelitian adalah hal yang penting karena informasi hasil penelitian diperoleh dari hasil estimasi yang dilakukan. Oleh karena itu permasalahan selanjutnya dari skripsi ini adalah bagaimana langkah-langkah estimasi parameter yang akan disesuaikan dengan struktur matriks varians kovariansnya pada model linier campuran dua tingkat.
3. Analisis regresi longitudinal dapat digunakan di berbagai bidang, salah satunya adalah bidang pendidikan. Aplikasi analisis regresi longitudinal pada bidang pendidikan yang dilakukan adalah menganalisis data Indeks Pendidikan di Jawa Barat. Data Indeks Pendidikan dapat dilihat sebagai data longitudinal dengan struktur data bertingkat. Data ini terdiri dari beberapa waktu penelitian dan berbagai faktor yang mempengaruhi. Kasus ini digunakan sebagai permasalahan terakhir dari skripsi yaitu bagaimana aplikasi model linier campuran dua tingkat terhadap nilai Indeks Pendidikan di Jawa Barat dari tahun 2014 sampai 2018.

1.3 Batasan Masalah

Untuk mengarahkan pembahasan dalam penelitian ini pada sasaran yang dituju, batasan ruang lingkup permasalahannya adalah sebagai berikut:

1. Data yang digunakan merupakan data sekunder berupa data Indeks Pendidikan di Jawa Barat dari tahun 2014 sampai dengan 2018 yang diperoleh dari Badan Pusat Statistik (BPS).
2. Matriks varians kovarians yang dimodelkan adalah matriks kovarians *Diagonal*, *Compound Symmetry* (CS) dan tidak terstruktur (*Unstructured Covariance Matrix*).
3. Metode estimasi yang digunakan untuk model linier campuran dua tingkat yaitu metode Estimasi *Maximum Likelihood* (ML) dan *Restricted Maximum Likelihood* (REML).
4. Estimasi parameter kovarians dan estimasi parameter efek tetap dikerjakan menggunakan perangkat lunak Rstudio.

1.4 Tujuan dan Manfaat Penelitian

Berdasarkan rumusan masalah yang telah dituliskan sebelumnya, tujuan dan manfaat dari penelitian ini adalah:

1. Menjelaskan apa itu model linier campuran dua tingkat dan bagaimana menentukan bentuk model linier campuran dua tingkat yang sesuai untuk data longitudinal.
2. Menentukan langkah-langkah estimasi parameter dari model linier campuran dua tingkat untuk data longitudinal menggunakan metode *Maximum Likelihood Estimation* (MLE) dan *Restricted Maximum Likelihood* (REML).
3. Membentuk model prediksi untuk nilai Indeks Pendidikan di Jawa Barat

Adapun manfaat dari penelitian ini adalah:

1. Melengkapi pembendaharaan teori tentang model linier campuran dua tingkat, secara khusus dalam pemodelan struktur varians kovarians.
2. Hasil analisis dapat diterapkan pada berbagai bidang yang memiliki struktur data serupa.

1.5 Metode Penelitian

Metode penelitian yang akan dilakukan dalam penelitian ini terdiri dari langkah-langkah sebagai berikut:

1. Merumuskan masalah, dimulai dari heterogenitas dalam subjek yang ada pada data longitudinal.
2. Melakukan studi literatur yang berkaitan dengan penelitian yang dilakukan terhadap berbagai referensi seperti *e-book*, jurnal, dan referensi lain yang berkaitan dengan penelitian terdahulu.
3. Mencari data sekunder yang akan digunakan, yaitu data Indeks Pendidikan di Jawa Barat tahun 2014 sampai 2018.
4. Melakukan analisis dimulai dari menentukan bentuk model linier campuran dua tingkat yang sesuai dengan data dan menentukan struktur matriks varians kovarians
5. Melakukan estimasi parameter kovarians dengan menggunakan metode REML, estimasi parameter efek tetap dan parameter efek acak menggunakan metode ML dari model yang telah dibentuk.
6. Membuat program komputer menggunakan perangkat lunak R untuk melakukan perhitungan data.
7. Memberikan kesimpulan berdasarkan hasil analisis.

1.6 Sistematika Penulisan

Berdasarkan sistematika penulisannya, skripsi ini terdiri atas lima bab, daftar pustaka, dan lampiran, dimana setiap bab terdapat beberapa subbab.

BAB I : PENDAHULUAN

Bab I berisi latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, metodologi penelitian, dan sistematika penulisan.

BAB II : LANDASAN TEORI

Bab II berisi teori-teori yang melandasi pembahasan dalam penulisan skripsi. Secara garis besar, bab ini mencakup data longitudinal, model linier campuran, metode estimasi parameter menggunakan metode *Maximum Likelihood* dan *Restricted Maximum Likelihood*, dan uji kecocokan secara umum.

BAB III : MODEL LINIER CAMPURAN DUA TINGKAT DALAM ANALISIS DATA LONGITUDINAL

Bab III berisi pembahasan utama dari penelitian yang dikaji, meliputi pembahasan mengenai model linier campuran dua tingkat yang digunakan, struktur kovarians, dan estimasi parameter dari model linier campuran dua tingkat.

BAB IV : STUDI KASUS DAN ANALISIS

Bab IV berisi tentang studi kasus, perhitungan model linier campuran dua tingkat untuk data longitudinal serta analisis dan hasil studi kasus.

BAB V : PENUTUP

Bab V berisi kesimpulan dari seluruh pembahasan yang telah dikaji berdasarkan tujuan yang ingin dicapai. Selain itu diberikan saran untuk pengembangan lebih lanjut terhadap topik pembahasan dalam penelitian ini.