

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Data longitudinal adalah data hasil pengukuran berulang untuk satu atau beberapa variabel pada setiap anggota sejumlah subjek atau individu yang sama, yang diamati pada sejumlah titik waktu berbeda. Data longitudinal berbeda dengan data *time series* yang mana pada data *time series* hanya menyangkut satu objek dengan rentang seri pengukuran yang jauh lebih panjang, sedangkan data longitudinal menyangkut lebih banyak subjek dengan panel (gelombang waktu pengukuran data) yang relatif sedikit [1].

Dalam studi longitudinal, subjek sering memiliki data yang hilang karena terlewat atau tidak diisi. Data longitudinal yang tidak lengkap sering muncul karena adanya kesulitan dalam memastikan tanggapan pada waktu penelitian yang dijadwalkan, jawaban pada formulir yang tidak terselesaikan, maupun terjadinya penolakan pasien serta subjek studi yang gagal untuk menghadiri kunjungan yang dijadwalkan.

Saat memilih teknik statistik, penting untuk mempertimbangkan mekanisme data hilang. Data hilang terbagi menjadi *Missing Completely at Random* (MCAR), *Missing not at Random* (MNAR), dan *Missing at Random* (MAR). Secara kasar, data dikatakan MAR jika hilangnya tergantung pada data yang diamati tetapi tidak pada data yang tidak teramati, data dikatakan MNAR jika kekurangan tersebut terkait dengan respon yang tidak teramati dan data dikatakan MCAR jika probabilitas hilangnya respons tidak bergantung pada variabel dependen maupun independen, baik yang teramati maupun tidak (partisipan dengan data yang teramati lengkap dapat dilihat sebagai sampel acak dari keseluruhan) [2].

Pendekatan dominan untuk analisis data longitudinal adalah pendekatan GEE (*Generalized Estimating Equations*) yang sudah sering diterapkan dalam memperkirakan data longitudinal yang berkorelasi. Namun, ketika data longitudinal

yang akan dianalisis memiliki data hilang di dalamnya, terutama apabila mekanisme hilangnya data yaitu *missing at random* (MAR) atau hilang secara acak, metode yang digunakan yaitu *Weighted Generalized Estimating Equations* (WGEE). Metode WGEE ini merupakan hasil pengembangan dari metode GEE. Di mana pada metode WGEE data yang ada akan diboboti terlebih dahulu sebelum mengestimasi. Sehingga hasil estimasi yang dihasilkan akan lebih baik dan tidak bias.

Beberapa penelitian yang mengkaji tentang GEE dan WGEE diantaranya adalah jurnal yang ditulis oleh Ming Wang [3] membahas mengenai metode GEE kemudian menerapkannya pada data simulasi, jurnal yang ditulis oleh Cong Xu dkk [4] membahas mengenai pengaplikasian WGEE pada aplikasi R studio dan juga jurnal yang ditulis oleh Alejandro Salazar dkk yang membahas mengenai WGEE dan pengaplikasiannya menggunakan R Studio [5].

Dalam penelitian skripsi ini penulis akan membahas Analisis Data Longitudinal dengan Menggunakan WGEE Pada Kasus Data Hilang. Metode WGEE kemudian akan diaplikasikan pada data Indeks Pembangunan Kesehatan Masyarakat (IPKM) 33 Kabupaten / Kota di Provinsi Sumatera Utara.

## **1.2 Rumusan Masalah**

Rumusan masalah pada skripsi ini adalah bagaimana penggunaan metode *Weighted Generalized Estimating Equations* (WGEE) pada data longitudinal yang terdapat data hilang dengan struktur korelasi yang berbeda.

## **1.3 Batasan Masalah**

Dalam penelitian ini penulis hanya fokus pada:

1. Jenis data yang digunakan adalah data longitudinal dengan data hilang.
2. Jenis data hilang yang digunakan dalam skripsi ini adalah *Missing At Random* (MAR).
3. Data yang digunakan dalam penerapan metode WGEE adalah data Indeks Pembangunan Kesehatan Masyarakat (IPKM) 33 Kabupaten / Kota di Provinsi Sumatera Utara periode 5 tahunan yaitu tahun 2013 dan 2018.

## 1.4 Tujuan

Berdasarkan latar belakang dan rumusan masalah yang ada, maka tujuan dari penulisan penelitian ini adalah sebagai berikut:

1. Menentukan langkah – langkah regresi longitudinal dengan menggunakan metode *Weighted Generalized Estimating Equations* (WGEE) dengan kondisi adanya data hilang.
2. Menentukan model terbaik dari struktur korelasi yang berbeda.
3. Menerapkan metode WGEE untuk data hilang pada kasus data longitudinal IPKM 33 Kabupaten / Kota di Provinsi Sumatera Utara.

## 1.5 Metode Penelitian

Metode yang digunakan dalam penelitian ini adalah dengan mengkaji sumber pustaka berupa buku, jurnal, maupun skripsi yang berkaitan dengan *Weighted Generalized Estimating Equations* (WGEE) dan juga data hilang pada data longitudinal. Data yang dikumpulkan kemudian akan dianalisis dengan menggunakan bantuan aplikasi R Studio versi 4.1.2.

## 1.6 Sistematika Penelitian

Berdasarkan sistematika penulisanya, skripsi ini terdiri atas lima bab serta daftar pustaka dan lampiran. Di mana dalam setiap bab terdapat beberapa subbab yang dirincikan sebagai berikut:

### **BAB I : PENDAHULUAN**

Bab ini berisi tentang pendahuluan diantaranya adalah latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, ruang lingkup penelitian, dan sistematika penulisan dari masalah yang dikaji.

### **BAB II : LANDASAN TEORI**

Bab ini berisi penjelasan tentang teori-teori yang berkaitan dengan masalah yang akan dikaji. Secara garis besar, bab ini mencakup semua yang berkaitan dengan Data longitudinal dan Data Hilang

**BAB III : ANALISIS DATA LONGITUDINAL DENGAN MENGGUNAKAN WEIGHTED GENERALIZED ESTIMATING EQUATIONS (WGEE) PADA DATA HILANG**

Bab ini berisi pembahasan utama dari studi yang dikaji, meliputi pembahasan mengenai *Weighted Generalized Estimating Equations* dan bagaimana penerapannya pada data hilang di data longitudinal

**BAB IV : STUDI KASUS DAN ANALISIS DATA**

Bab ini merupakan intisari dari bab – bab sebelumnya yang terdiri penerapan WGEE pada data longitudinal dengan data hilang MAR.

**BAB V : PENUTUP**

Bab ini berisi penjelasan mengenai beberapa hal yang menjadi kesimpulan dan saran atas penelitian yang telah dilakukan serta beberapa saran pengembangan tulisan ini.

