

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Regresi logistik digunakan untuk mencari pengaruh variabel prediktor terhadap variabel respon kategori, berupa variabel dikotomis yang mengambil nilai 0 untuk tidak terjadi dan 1 untuk terjadi. Jika variabel prediktor (seperti faktor risiko) juga dikotomis, perkiraan jumlah sampel yang diperlukan dapat ditemukan dari tabel ukuran sampel yang dipublikasikan untuk perbandingan dua proporsi. Dalam melakukan sampling pada model regresi logistik penting diperhatikan besarnya ukuran sampel dan teknik pengambilan sampel yang digunakan. Penelitian Takeshi Amemiya[1], menjelaskan penggunaan data yang terbatas dengan hasil estimator maksimum likelihood menghasilkan kinerja yang lebih baik karena nilai biasanya lebih kecil daripada estimasi chi-square logit minimum. Sehingga estimasi maksimum likelihood lebih efektif ketika data yang tersedia terbatas.

Penelitian Daniel V. Gordon dan kawan-kawannya[2] menghasilkan metode untuk menghitung ukuran sampel dalam model regresi logistik yang paling baik. Metode yang beragam dapat menghasilkan ukuran sampel yang berbeda, pada penelitian tersebut hanya berfokus dalam jumlah sampel yang dibutuhkan. Sedangkan untuk memperbaiki tingkat estimasi parameter selain dengan ukuran sampel juga perlunya teknik pengambilan sampel yang tepat.

Ukuran sampel yang terlalu kecil, hasil estimasi parameter dengan metode estimasi maksimum likelihood menimbulkan estimasi yang bias. Formula untuk menentukan ukuran sampel yang sesuai untuk model logistik dimana probabilitas respon kecil dan ukuran sampel yang kecil itu sangat sensitif terhadap distribusi prediktor. Ukuran sampel dan strategi pengambilan sampel dapat mempengaruhi besar kecilnya bias sampel dari estimasi maksimum likelihood dan juga ketepatan estimasi selanjutnya[3].

Sampling adalah proses memilih Sebagian kecil dari populasi yang mewakili keseluruhan populasi. Teknik pengambilan sampel yang baik dapat memastikan bahwa sampel yang dipilih mencerminkan populasi yang lebih luas, karena sampel yang representatif memungkinkan estimasi parameter yang akurat dan dapat diandalkan serta memastikan bahwa kesimpulan yang diambil dari sampel tersebut dapat digeneralisasi ke populasi sehingga dapat memengaruhi nilai bias yang didapat. Karena diperlukannya pandangan yang komprehensif terhadap analisis teknik pengambilan sampel dengan karakteristik yang berbeda dari populasi logistik, penelitian ini mengambil judul “Pengaruh Metode Pengambilan Sampel Acak Pada Estimasi Maksimum Likelihood Dalam Model Regresi Logistik Sederhana ”.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang yang sudah disampaikan di atas, maka rumusan masalah pada penelitian ini adalah:

1. Apakah teknik pengambilan sampel berpengaruh terhadap estimasi maksimum likelihood pada model regresi logistik sederhana?
2. Bagaimana perbandingan efektifitas pengambilan sampel acak sederhana dengan pengambilan sampel acak bertingkat terhadap estimasi maksimum likelihood pada model regresi logistik sederhana?

## **1.3 Batasan Masalah**

Adapun Batasan masalah pada penelitian ini antara lain:

1. Metode pengambilan sampel yang digunakan yaitu pengambilan sampel acak sederhana, pengambilan sampel acak bertingkat dengan alokasi sama, dan pengambilan sampel acak bertingkat alokasi proporsional.
2. Sampel disimulasikan menggunakan metode Monte Carlo.
3. Hasil estimasi parameter dibandingkan dalam ukuran nilai bias, standar error, dan persentasi kelayakan model.
4. Variabel respon Y dikategorikan ke dalam dua kategori yaitu 1 dan 0.

#### 1.4 Tujuan dan Manfaat

Berdasarkan latar belakang dan rumusan masalah yang sudah penulis sampaikan di atas, tujuan penelitian ini diantaranya:

Tujuan

1. Mengetahui apakah teknik pengambilan sampel dapat memengaruhi estimasi maksimum likelihood pada model regresi logistik sederhana.
2. Mengetahui teknik pengambilan sampel mana yang lebih efektif digunakan dalam jumlah sampel yang kecil pada estimasi maksimum likelihood dari model regresi logistik sederhana.

Adapun manfaat dari penelitian ini adalah:

1. Mengembangkan wawasan ilmu pengetahuan yang berkaitan dengan metode pengambilan sampel.
2. Memberikan informasi berkaitan dengan teknik pengambilan sampel mana yang lebih efektif digunakan dalam estimasi maksimum likelihood pada model regresi logistik sederhana.

#### 1.5 Metode Penelitian

Metode penelitian yang digunakan dalam skripsi ini adalah :

##### 1. Studi Literatur

Dalam tahapan Studi Literatur, dilakukan pengumpulan teori-teori yang berupa identifikasi masalah dan pengumpulan data pustaka yang mendukung skripsi ini yang berkaitan dengan teknik sampling dan regresi logistik sederhana.

##### 2. Analisis dan Simulasi

Tahap ini adalah tahap pengkajian lebih mendalam tentang teknik sampling seperti teknik sampling acak sederhana, teknik sampling acak bertingkat, estimasi maksimum likelihood pada model regresi logistik sederhana, dan simulasi Monte Carlo. Kemudian dilanjutkan dengan simulasi menggunakan algoritma yang telah dibangun dan diimplementasikan menggunakan *software* R-studio.

## 1.6 Ruang Lingkup Penelitian

Penelitian skripsi ini merupakan pengembangan dari studi literatur yang dikaji ulang berdasarkan jurnal rujukan berjudul “*Effects of Random Sampling Methods on Maximum Likelihood Estimates of a Simple Logistic Regression Model*”. Pada jurnal tersebut, dijelaskan mengenai pengaruh teknik pengambilan sampel dimana teknik pengambilan sampel yang dimaksud adalah pengambilan sampel acak sederhana dan pengambilan sampel acak bertingkat, yang disimulasikan dengan metode Monte Carlo pada estimasi maksimum likelihood dari model regresi logistik sederhana[3]. Pada studi sebelumnya telah diteliti mengenai penaksiran parameter dari model logistik yang timbul dari berbagai macam strategi pengambilan sampel. Hasilnya menunjukkan ukuran sampel dan juga strategi dalam pengambilan sampelnya sendiri mempengaruhi nilai bias dari estimasi maksimum likelihood dan juga untuk ketepatan estimasi berikutnya.

Terdapat penelitian yang membahas tentang bagaimana desain dan analisis dua fase dapat digunakan untuk mengatasi masalah sampling yang tidak representatif dari populasinya dan mengoptimalkan penggunaan data yang tersedia. Hasil penelitian menunjukkan bahwa desain dan analisis dua fase dapat meningkatkan akurasi prediksi prognosis tumor Wilms dan memungkinkan analisis yang lebih efektif dari data yang terbatas[4].

Untuk mempelajari bagaimana strategi pengambilan sampel mempengaruhi sifat-sifat sampel kecil dari estimator logit, terdapat penelitian yang membahas tentang bagaimana penggunaan strategi pengambilan sampel yang berbeda dapat mempengaruhi bias dan presisi estimator logit dalam situasi ukuran sampel kecil. Hasil penelitian menunjukkan bahwa strategi pengambilan sampel yang tepat dapat mengurangi bias dan meningkatkan presisi estimator logit, sehingga memungkinkan analisis yang lebih akurat dalam situasi sampel berukuran kecil[5].

Analisis dilakukan pada data stunting Provinsi Jawa Barat tahun 2022 yang akan disimulasikan menggunakan metode Monte Carlo. Jawa Barat merupakan salah satu provinsi dengan populasi terbesar di Indonesia. Jawa Barat memiliki variasi geografis dan kondisi sosial ekonomi yang luas, dari daerah perkotaan hingga daerah pedesaan yang lebih terpencil. Tingginya jumlah penduduk membuat prevalensi stunting di Provinsi ini memiliki dampak yang signifikan terhadap angka

nasional. Mengidentifikasi dan memahami faktor-faktor yang berkontribusi terhadap stunting di daerah ini adalah langkah penting untuk menurunkan prevalensi stunting secara keseluruhan di Indonesia.

## **1.7 Sistematika Penulisan**

### **BAB I PENDAHULUAN**

Bab ini berisi mengenai latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, metode penelitian, dan sistematika penulisan.

### **BAB II LANDASAN TEORI**

Bab ini berisi mengenai penjelasan teori-teori penunjang penelitian seperti teknik pengambilan sampel, estimasi parameter menggunakan maksimum likelihood, dan model regresi logistik.

### **BAB III PENGARUH TEKNIK PENGAMBILAN SAMPEL TERHADAP ESTIMASI MAKSIMUM LIKELIHOOD PADA MODEL REGRESI LOGISTIK SEDERHANA**

Bab ini berisi pembahasan mengenai pengaruh teknik pengambilan sampel terhadap estimasi maksimum likelihood dalam mengestimasi model regresi logistik sederhana.

### **BAB IV STUDI KASUS DAN ANALISA**

Bab ini berisi penerapan studi kasus dan interpretasi dari hasil simulasi.

### **BAB V KESIMPULAN DAN SARAN**

Bab ini berisi kesimpulan dari pembahasan yang telah dianalisis serta saran untuk pengembangan penelitian selanjutnya