

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Statistika adalah ilmu yang mempelajari cara mengumpulkan, mengolah, menganalisis, dan menginterpretasikan data agar dapat disajikan dengan akurat dan efektif[1]. Metode statistik adalah prosedur yang digunakan dalam pengumpulan, penyajian, analisis, dan penafsiran. Metode tersebut dikelompokkan menjadi dua, yaitu statistika deskriptif dan statistika inferensia[2]. Pada statistika inferensi terdapat teknik sampling. Sampling adalah cara pengumpulan data di mana yang diselidiki adalah elemen sampel dari suatu populasi[3].

Pada penelitian ini, akan digunakan metode pengambilan sampel kasus kontrol dua fase dan uji kecocokan model regresi logistik. Pengambilan sampel kasus kontrol dilakukan dengan mencari kasus secara *purposive* atau *random sampling*. Studi kasus kontrol adalah rancangan studi epidemiologi yang mempelajari hubungan antara paparan dan penyakit dengan cara mengamati kelompok kasus dan kontrol lalu menelusuri status paparannya[4].

Pengambilan sampel kasus kontrol dua fase menggunakan model regresi logistik, karena model tersebut telah menjadi metode yang umum digunakan untuk memahami keterkaitan antara variabel independen dengan variabel dependen yang bersifat biner atau variabel yang menggambarkan terjadi atau tidaknya suatu peristiwa[5], memiliki fleksibilitas untuk menyesuaikan desain pengambilan sampel yang kompleks, dapat mengestimasi *odds ratio* dalam mengidentifikasi faktor risiko terkait dengan kasus penyakit, dan model regresi logistik dapat menangani ketidakseimbangan antara jumlah kasus dan kontrol, selain itu dapat memberikan estimasi yang konsisten bahkan dengan jumlah kasus yang relatif sedikit dibandingkan dengan jumlah kontrol[6]. Dengan demikian, model regresi logistik menjadi pilihan yang umum dan efektif untuk menganalisis data dari pengambilan sampel kasus kontrol, terutama dalam mengestimasi populasi dari data survei. Regresi logistik bermanfaat di berbagai bidang, salah satunya pada epidemiologi. Epidemiologi merupakan studi yang mempelajari distribusi,

determinan, frekuensi penyakit, dan faktor yang memengaruhi status kesehatan pada populasi manusia[7].

Salah satu komponen penting dalam melakukan analisis regresi logistik adalah pengambilan sampel yang representatif dari populasi yang diteliti. Dalam penelitian epidemiologi, pengambilan sampel dilakukan dengan metode kasus kontrol dua fase. Pengambilan sampel kasus kontrol dua fase telah menjadi metode populer dalam studi epidemiologi karena efisiensinya dalam penghematan biaya dan waktu[8]. Pada tahap pertama, kasus dan kontrol diambil dari populasi umum, dan beberapa informasi yang mengelompokkan, seperti ukuran paparan (misalnya penyakit atau non-penyakit) dapat diperoleh. Pada tahap kedua, sub-sampel dipilih dalam strata yang ditentukan oleh status penyakit, informasi stratifikasi lainnya, informasi yang lebih baik tentang paparan dan kovariat lainnya diperoleh untuk subsampel[7]. Estimasi parameter merupakan teknik statistik yang bertujuan untuk memperkirakan nilai parameter populasi berdasarkan data atau ukuran yang diperoleh dari sampel[9].

Dalam regresi logistik, parameter diestimasi menggunakan *Generalized Linear Models* (GLM) yang diterapkan dalam perangkat lunak statistik seperti R. GLM adalah kerangka kerja umum untuk regresi yang dapat diterapkan pada berbagai jenis data, termasuk data biner yang dianalisis melalui regresi logistik. Dasar utama yang digunakan oleh GLM untuk mengestimasi parameter adalah Metode *Maksimum Likelihood Estimation* (MLE). MLE adalah metode yang digunakan untuk mengestimasi parameter model dengan memaksimalkan *likelihood* (kemungkinan) dari pengamatan yang tersedia. Dengan GLM, MLE membantu memastikan bahwa model yang dibangun sesuai dengan data yang diamati sebaik mungkin, dengan menggunakan fungsi *likelihood* untuk mencari nilai parameter yang optimal.

Uji Kecocokan (*Goodness-of-Fit Test*) merupakan uji statistik untuk melihat apakah sebuah model cocok dengan sekumpulan data atau apakah data yang diamati sesuai dengan ekspektasi teoretis. Uji yang paling umum digunakan adalah uji chi-square. Semakin besar statistik chi-square maka semakin buruk kecocokannya, sedangkan semakin kecil statistik chi-square maka semakin baik kecocokannya.

Hal ini dikarenakan statistik chi-square yang kecil menunjukkan perbedaan yang kecil antara model dan data[10].

Penelitian ini bertujuan untuk menghasilkan estimasi parameter dan menguji kecocokan model regresi logistik pada pengambilan sampel kasus kontrol dua fase, dengan mengaplikasikan data kesehatan masyarakat. Salah satu masalah kesehatan masyarakat Indonesia adalah stunting. Stunting adalah gangguan pertumbuhan dan perkembangan anak akibat kekurangan gizi kronis dan infeksi berulang, yang ditandai dengan panjang atau tinggi badannya berada di bawah standar yang ditetapkan oleh menteri yang menyelenggarakan urusan pemerintahan di bidang kesehatan[11].

Penelitian ini berdasarkan jurnal tahun 2021[12]. Di mana jurnal tersebut membahas bagaimana cara memperluas kemungkinan empiris untuk data kasus kontrol standar menjadi data kasus kontrol dua fase, selain itu akan digunakan uji goodness of fit tipe Kolmogorov Smirnov untuk menguji validasi model regresi logistik. Tujuannya adalah untuk menyelidiki karakteristik estimasi statistik, seperti distribusi estimasi maksimum likelihood empiris dan distribusi rasio statistik likelihood empiris. Dalam jurnal tersebut digunakan data penyakit Kanker (*Tumor Wilms*) di beberapa rumah sakit China. Sedangkan untuk penelitian yang akan dilakukan pada penelitian ini akan digunakan data Prevalensi stunting di 4 Provinsi Jawa Barat, Jawa Tengah, Jawa Timur, dan DI Yogyakarta. Hasilnya menunjukkan bahwa penggunaan metode berbasis rasio statistik likelihood memiliki keunggulan, dan uji kecocokan yang diusulkan berhasil dalam mengevaluasi model regresi logistik. Menurut penelitian tahun 2020[13] menjelaskan bahwa mengenai analisis faktor stunting menggunakan regresi logistik biner dengan tujuan untuk memperoleh model regresi logistik biner dari faktor yang memengaruhi terjadinya stunting dan mengetahui faktor apa yang paling berpengaruh terhadap resiko terjadinya stunting pada penelitian. Dalam penelitian tahun 2021[14], membahas mengenai analisis faktor risiko stunting pada balita usia 6 bulan-23 bulan di suatu puskesmas di kabupaten Pekkae Makassar dan bertujuan untuk menganalisis satu persatu faktor yang ada pada stunting diantaranya riwayat pemberian asi eksklusif, MP ASI, imunisasi, tinggi badan ibu, pendapatan ekonomi keluarga dan lain-lain. Pada penelitian tahun 2015[15], bertujuan untuk mengetahui faktor-faktor yang

berhubungan dengan kejadian stunting anak usia sekolah dasar pada siswa/siswi Madrasah Ibtidaiyah (MI) Muhammadiyah Haurgelis Indramayu tahun 2015. Membahas mengenai penelitian case control dengan menggunakan kriteria inklusi dan eksklusi serta *frame sampling*, menggunakan data primer dan data sekunder dengan perbandingan siswa yang stunting (kasus) dan yang tidak (kontrol) adalah 1:1 dengan jumlah populasi yang digunakan adalah 174 orang dan sampel 70 orang.

Oleh karena itu, berdasarkan uraian diatas, penulis mengambil judul “Uji Kecocokan untuk Model Regresi Logistik pada Pengambilan Sampel Kasus Kontrol Dua Fase”.

1.2 Rumusan Masalah

Rumusan masalah yang akan dikaji dalam skripsi ini adalah:

1. Bagaimana hasil estimasi parameter model regresi logistik pada pengambilan sampel kasus kontrol dua fase?
2. Bagaimana hasil uji kecocokan model regresi logistik pada pengambilan sampel kasus kontrol dua fase?

1.3 Batasan Masalah

Adapun batasan masalah pada penelitian ini, yaitu Data yang akan digunakan adalah data sekunder mengenai prevalensi stunting 4 Provinsi di Pulau Jawa, yaitu Jawa Barat, Jawa Tengah, Jawa Timur, dan DI Yogyakarta sebagai variabel dependen, serta variabel Independen bersumber dari Statistik Kesejahteraan Rakyat 4 Provinsi.

1.4 Tujuan dan Manfaat Penelitian

Berdasarkan latar belakang dan rumusan masalah yang diuraikan sebelumnya, penelitian ini memiliki tujuan sebagai berikut:

1. Memperoleh hasil estimasi parameter model regresi logistik pada pengambilan sampel kasus kontrol dua fase.
2. Memperoleh dan menjelaskan hasil uji kecocokan untuk model regresi logistik pada pengambilan sampel kasus kontrol dua fase.

Adapun manfaat dalam penelitian ini adalah:

1. Memberikan informasi mengenai hasil estimasi parameter model regresi logistik pada pengambilan sampel kasus kontrol dua fase,
2. Memberikan pengetahuan dan wawasan mengenai hasil uji kecocokan untuk model regresi logistik pada pengambilan sampel kasus kontrol dua fase dengan menerapkan data kesehatan masyarakat (stunting).

1.5 Metode Penelitian

Metode yang dilakukan dalam menyelesaikan skripsi ini bersifat studi literatur dengan cara mengumpulkan teori yang berhubungan dengan pengambilan sampel kasus kontrol dua fase dan model regresi logistik yang diperoleh dari berbagai sumber berupa buku, jurnal, dan skripsi. Sebelum melakukan uji kecocokan model, akan dicari terlebih dahulu hasil estimasi parameter model regresi logistik untuk pengambilan sampel kasus kontrol dua fase. Penelitian ini akan menggunakan data sekunder mengenai prevalensi stunting.

1.6 Sistematika Penulisan

BAB I

PENDAHULUAN

Berisi mengenai latar belakang masalah, rumusan masalah, batasan masalah, tujuan dan manfaat penelitian, metode penelitian dan sistematika penulisan.

BAB II

LANDASAN TEORI

Berisi penjelasan mengenai teori-teori yang menjadi landasan dalam literatur yang diteliti, yaitu statistika, teknik sampling, model regresi logistik, estimasi parameter, GLM, MLE, Uji Kecocokan, Bootstrap, Stunting, dan Program R.

BAB III

UJI KECOCOKAN UNTUK MODEL REGRESI LOGISTIK PADA PENGAMBILAN SAMPEL KASUS KONTROL DUA FASE

Berisi tentang pengambilan sampel kasus kontrol, pengambilan sampel kasus kontrol satu fase, pengambilan

sampel kasus kontrol dua fase, estimasi parameter dan uji kecocokan model regresi logistik, langkah penelitian, dan flowchart penelitian.

BAB IV STUDI KASUS DAN ANALISA

Bab ini berisi penerapan dari estimasi parameter regresi logistik pada pengambilan sampel kasus kontrol dua fase, hasil bootsrap, dan hasil uji kecocokan.

BAB V KESIMPULAN DAN SARAN

Berisi tentang kesimpulan dan saran yang berkaitan dengan penulisan

