

# BAB I

## PENDAHULUAN

Bab ini membahas tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, metode penelitian, dan sistematika penulisan.

### 1.1 Latar Belakang Masalah

Statistika sangat berhubungan dengan data karena statistika merupakan sebuah ilmu pengetahuan yang mengumpulkan data, mengelompokkan data, menganalisis data, dan membuat kesimpulan berdasarkan data sehingga dapat diinterpretasikan menjadi sebuah informasi yang berguna. Salah satu data yang dapat diperoleh adalah data yang berkaitan dengan waktu hidup suatu individu. Suatu analisis statistika yang dapat digunakan untuk menganalisis data waktu hidup adalah analisis ketahanan hidup (analisis *survival*). Analisis *survival* memiliki tujuan untuk menaksir peluang kematian, kesembuhan, kekambuhan atau kejadian lain sampai pada periode waktu tertentu.

Terdapat tiga pendekatan dalam analisis *survival*, yaitu metode parametrik, semiparametrik dan nonparametrik. Ketika model secara eksplisit menggunakan parameter untuk menggambarkan distribusi data *survival*, maka disebut metode parametrik. Contoh metode parametrik diantaranya model *Weibull*, eksponensial, *log-normal*, *log-logistik*. Ketika sebuah model membuat asumsi tentang *hazard rate* tetapi tidak menggunakan parameter tambahan untuk menggambarkan distribusi waktu *survival*, maka disebut metode semiparametrik. Contoh metode semiparametrik adalah *Cox regression*. Ketika model tidak memiliki parameter tertentu untuk menggambarkan distribusi data *survival*, maka disebut metode nonparametrik. Contoh metode nonparametrik diantaranya metode *life-table*, estimasi kaplan meier, dan model *survival* waktu diskrit.

Pada metode parametrik dan semiparametrik diasumsikan bahwa data *survival* adalah data waktu kontinu. Model *survival* waktu kontinu merupakan metode yang paling sering digunakan dalam analisis *survival*. Model tersebut biasanya digunakan ketika waktu dari suatu kejadiannya diketahui secara tepat. Tetapi jika peneliti menganggap data waktu diskrit sebagai data waktu kontinu

maka dapat menyebabkan hasil dan interpretasi yang salah. Model *survival* waktu diskrit yang diperkenalkan oleh Allison (1982) [1] dapat digunakan jika data *survival* yang digunakan berupa data waktu diskrit. Data waktu diskrit terbagi menjadi dua kasus, kasus pertama ketika periode transisinya diskrit secara struktural dan kasus kedua ketika waktu *survival* dikelompokkan sebagai interval waktu diskrit. Model *survival* waktu diskrit dapat diestimasi dengan menggunakan fungsi *logit* dan fungsi *complementary log-log*.

Berikut ini merupakan beberapa penelitian mengenai model *survival* waktu diskrit yaitu Singer dan Willet [2] yang membahas tentang penggunaan analisis *survival* waktu diskrit untuk mempelajari durasi dan waktu dari kejadian, Xie dkk [3] yang membahas tentang penggunaan model *survival* waktu diskrit untuk menganalisis pola remisi pola remisi dari penyalahgunaan zat di antara orang dengan penyakit mental berat, Odeniya dkk [4] yang membahas tentang model *survival* waktu diskrit untuk menganalisis usia saat kehamilan pertama di antara wanita Nigeria, dan Raykov dkk [5] yang membahas tentang aplikasi model *survival* waktu diskrit untuk studi hipotesis depresi vaskular (*vascular depression*).

Perbedaan model *survival* waktu diskrit dengan model *survival* waktu kontinu terletak pada struktur datanya. Data *survival* umumnya dikelompokkan dalam data *person-oriented* (dimana setiap subjek hanya mempunyai satu baris data), sedangkan model *survival* waktu diskrit harus diaplikasikan pada data *person-period* (dimana setiap individu mempunyai beberapa baris data yang bergantung pada periode waktu diskrit dari setiap individu). Oleh karena itu, hal pertama yang harus dilakukan adalah mengubah data *person-oriented* menjadi data *person-period*. Hal tersebut menyebabkan ukuran dari data menjadi besar dan situasi ini dapat memperpanjang waktu analisis sedangkan para peneliti seringkali mengandalkan data yang kompleks dan besar tetapi harus menganalisis dengan cepat dan efisien dalam periode yang singkat.

Karakteristik tersebut memotivasi peneliti untuk menggunakan metode pengambilan sampel pada model *survival* waktu diskrit, khususnya untuk data *person-oriented* yang besar. McIntyre (1952) mengusulkan sebuah metode yang dinamakan pengambilan sampel peringkat [6]. Metode pengambilan sampel peringkat merupakan suatu metode pengambilan sampel dimana setiap unit yang

diambil dari populasi dapat diberi peringkat dengan cara tertentu tanpa melakukan pengukuran sebenarnya sehingga dapat menghemat waktu dan biaya penelitian [7]. Metode pengambilan sampel tersebut akan digunakan dalam model *survival* waktu diskrit agar proses analisis menjadi lebih efisien dan menghemat waktu.

Penelitian terdahulu yang mengkaji tentang analisis *survival* waktu diskrit dengan pengambilan sampel peringkat yaitu terdapat pada jurnal yang ditulis oleh Nihal Ata Tutkun, Nursel Koyuncu, dan Ugur Karabey pada tahun 2019 yang berjudul “*Discrete-time Survival Analysis Under Ranked Set Sampling: An Application to Turkish Motor Insurance Data*” [8]. Jurnal tersebut memperkenalkan penggunaan metode pengambilan sampel peringkat pada model *survival* waktu diskrit untuk pertama kalinya dan penerapannya pada bidang asuransi kendaraan bermotor.

Selain pada bidang sosial, model *survival* waktu diskrit juga dapat diterapkan dalam bidang kesehatan, seperti pada kasus COVID-19. *Coronavirus disease 2019* (COVID-19) disebabkan oleh virus *Severe Acute Respiratory Syndrome Coronavirus 2* (SARS-CoV-2). Virus tersebut telah menyebar secara global dan menyebabkan banyak kematian. Sehingga peneliti tertarik untuk membahas lebih lanjut mengenai analisis *survival* waktu diskrit untuk memodelkan kasus terjadinya kematian pada pasien COVID-19 berdasarkan faktor-faktor yang mempengaruhinya dan juga melakukan metode pengambilan sampel peringkat pada pasien COVID-19. Oleh karena itu, skripsi ini diberi judul “Analisis *Survival* Waktu Diskrit dengan Pengambilan Sampel Peringkat pada Pasien COVID-19”.

## 1.2 Rumusan Masalah

Rumusan masalah yang akan dikaji dalam skripsi ini adalah :

1. Bagaimana estimasi model *survival* waktu diskrit dengan menggunakan fungsi *logit* dan fungsi *complementary log-log* serta langkah-langkah estimasi parameternya?
2. Bagaimana menerapkan model *survival* waktu diskrit pada data pasien COVID-19?
3. Bagaimana menentukan model terbaik pada data pasien COVID-19 dengan menggunakan metode *Akaike Information Criterion* (AIC)?

### 1.3 Batasan Masalah

Batasan masalah dalam skripsi ini adalah :

1. Data yang digunakan adalah data sekunder yang diperoleh dari rekam medis RSUD Kota Bandung. Data tersebut merupakan data rekam medis pasien COVID-19 dari bulan Januari-Juni 2021.
2. Penyensoran yang terjadi adalah sensor kanan.
3. Estimasi model *survival* waktu diskrit dengan menggunakan fungsi *logit* dan fungsi *complementary log-log*.
4. Estimasi parameter model *logit* dan model *complementary log-log* menggunakan *Maximum Likelihood Estimation* (MLE) dan iterasi *Fisher Scoring*.
5. Uji signifikansi parameter dengan menggunakan uji rasio *likelihood* dan uji *wald*.
6. Pemilihan model terbaik dengan menggunakan metode *Akaike Information Criterion* (AIC).
7. *Software* yang digunakan yaitu R *software*.

### 1.4 Tujuan Penelitian

Tujuan dari penulisan skripsi ini adalah :

1. Menentukan estimasi model *survival* waktu diskrit dengan menggunakan fungsi *logit* dan fungsi *complementary log-log* serta langkah-langkah estimasi parameternya.
2. Menerapkan model *survival* waktu diskrit pada data pasien COVID-19.
3. Menentukan model terbaik pada data pasien COVID-19 dengan menggunakan metode *Akaike Information Criterion* (AIC).

### 1.5 Metode Penelitian

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

1. Melakukan studi literatur terhadap berbagai referensi seperti buku, jurnal dan penelitian sebelumnya yang berhubungan dengan penelitian yang dilakukan.

Kemudian memahami secara mendalam mengenai model *survival* waktu diskrit dan metode pengambilan sampel peringkat.

2. Menentukan data, dalam skripsi ini menggunakan data sekunder yaitu data pasien COVID-19 dan sampel penelitian ditentukan dengan menggunakan metode pengambilan sampel peringkat.
3. Melakukan analisis, yaitu menentukan estimasi model *survival* waktu diskrit dengan menggunakan fungsi *logit* dan fungsi *complementary log-log* kemudian menentukan nilai estimasi parameter model *logit* dan model *complementary log-log* dengan menggunakan *Maximum Likelihood Estimation* (MLE) dan metode iterasi *Fisher Scoring*.
4. Melakukan uji signifikansi parameter yaitu uji rasio *likelihood* dan uji *wald* untuk mengetahui peranan variabel bebas yang terdapat dalam model *survival* waktu diskrit.
5. Menentukan model terbaik dengan menggunakan *Akaike Information Criterion* (AIC) terhadap data pasien COVID-19.
6. Menyusun algoritma langkah-langkah yang dilakukan dalam skripsi dan membuat program berdasarkan algoritma tersebut dengan menggunakan *software R*.
7. Memberikan interpretasi dan kesimpulan berdasarkan hasil analisis.

## 1.6 Sistematika Penulisan

Berdasarkan sistematika penulisannya, skripsi ini terdiri dari lima bab, serta daftar pustaka di mana setiap bab terdapat beberapa sub bab.

### **BAB I      PENDAHULUAN**

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

### **BAB II     LANDASAN TEORI**

Bab ini berisi teori-teori penunjang yang berkaitan dengan masalah yang akan dikaji dalam skripsi ini. Secara garis besar, bab ini mencakup semua yang berkaitan dengan sampel, peubah acak,

analisis *survival*, penyensoran, model waktu diskrit, *Generalized Linear Model* (GLMs), regresi logistik, regresi *complementary log-log*, *Maximum Likelihood Estimation* (MLE), metode *Fisher Scoring*, uji signifikansi parameter, dan *Akaike Information Criterion* (AIC).

### **BAB III ANALISIS SURVIVAL WAKTU DISKRIT DENGAN PENGAMBILAN SAMPEL PERINGKAT**

Bab ini berisi pembahasan inti dari penelitian yang dilakukan, meliputi pembahasan mengenai estimasi model *survival* waktu diskrit dengan menggunakan fungsi *logit* dan fungsi *complementary log-log*, estimasi parameter model *logit* dan model *complementary log-log* dengan metode MLE dan iterasi *Fisher Scoring*, metode pengambilan sampel peringkat pada model *survival* waktu diskrit, uji signifikansi parameter, pemilihan model terbaik, dan algoritma analisis *survival* waktu diskrit dengan menggunakan Program R.

### **BAB IV STUDI KASUS DAN ANALISIS**

Bab ini berisi pembahasan mengenai penerapan model *survival* waktu diskrit dengan pengambilan sampel peringkat pada pasien COVID-19, hasil analisis data dan interpretasinya.

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

Bab ini berisi kesimpulan dari pembahasan yang telah dikaji. Selain itu, diberikan saran yang dapat dilakukan untuk pengembangan penelitian agar lebih baik lagi.

## **DAFTAR PUSTAKA**

## **LAMPIRAN**