

# BAB I

## PENDAHULUAN

Pada bab ini berisi tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, metode penelitian, sistematika penulisan, dan ruang lingkup penelitian.

### 1.1 Latar Belakang Masalah

Pendugaan parameter dalam masalah statistik di berbagai bidang sangat penting, salah satunya yaitu estimasi parameter untuk variabel respon biner. Variabel respon biner adalah variabel respon atau dependen memiliki dua kemungkinan kejadian yang biasa dinotasikan dengan nilai 1 dan 0. Masalah ini sangat penting dan bisa ditemukan dalam ilmu sosial, biologi, dan ekonomi.

Kasus yang sering ditemui dalam pendugaan parameter adalah variabel respon  $Y_i$  berdistribusi tidak normal, dalam hal ini  $Y_i$  berdistribusi dalam keluarga eksponensial yaitu  $Y_i$  berdistribusi binomial berupa data biner. Misalnya, dalam bidang biologi dilakukan percobaan untuk mengetahui efektifitas suatu zat dengan berbagai tingkat konsentrasi terhadap insektisida, variabel respon memiliki kemungkinan antara insektisida yang mati dan hidup, insektisida yang mati diasumsikan sebagai “sukses” sedangkan insektisida yang tetap hidup diasumsikan sebagai “gagal”. Istilah “sukses” dan “gagal” dalam distribusi binomial dinotasikan dengan nilai 1 dan 0 untuk setiap kejadian.

Salah satu metode yang dapat digunakan untuk menangani masalah tersebut dengan menggunakan *Generalized Linier Models (GLM)*. Terdapat tiga komponen dalam GLM yaitu komponen random ( $\mu$ ), komponen sistematik ( $\eta$ ), dan fungsi *link* ( $g(\cdot)$ ). Fungsi *link* yang penting dan umum digunakan dalam GLM yaitu logit *link* dan probit *link*[10].

Logit *link* merupakan hasil transformasi dari model regresi logistik yang dibangun dari fungsi distribusi kumulatif model Logistik. Probit *link* merupakan transformasi dari model regresi probit yang dibangun dari fungsi distribusi kumulatif normal. Kedua model ini digunakan untuk menggambarkan hubungan

antara variabel respon yang mempunyai skala pengukuran dikotomus atau biner biasa disimbolkan  $Y$  dengan variabel prediktor yang mempunyai skala pengukuran bersifat dikotomus, polikotomus, atau kontinu. Akan tetapi, penelitian lain menunjukkan bahwa kedua fungsi *link* tersebut, terkadang bisa menghasilkan ketidaksesuaian dan ketidakcocokan model yang baik[4].

Untuk mengatasi masalah keterbatasan fungsi *link* tersebut, maka diusulkan fungsi *link* berdasarkan distribusi Weibull. Distribusi Weibull diperkenalkan oleh fisikawan Swedia Waloddi Weibull pada tahun 1939[14]. Penggunaan distribusi Weibull dikarenakan kesederhanaan dari distribusinya dan memiliki banyak “bentuk”. Untuk membangun fungsi *link* tersebut menggunakan model distribusi Weibull dua parameter, dimana parameter yang digunakan yaitu parameter lokasi dan parameter bentuk. Dengan menggunakan fungsi distribusi kumulatif dari model distribusi Weibull yang digunakan, maka dibangun fungsi *link* dengan komponen sistematis menggunakan model regresi sederhana  $\eta = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n$  dan komponen random sebagai nilai tengah respon menggunakan nilai tengah dari respon yang berdistribusi binomial ( $\pi$ ).

Dalam menyelesaikan model yang dibangun dari suatu distribusi perlu dilakukan estimasi parameter dari model tersebut, karena pada umumnya estimasi parameter merupakan hal yang tidak dapat dipisahkan dalam pengkajian distribusi dan nilai suatu parameter tidaklah diketahui. Estimasi parameter sering dipakai sebagai prosedur untuk mencari parameter dari sebuah model yang paling cocok pada suatu data pengamatan yang ada. Oleh karena itu, untuk menentukan nilai estimasi parameter dari model regresi yang diusulkan ditentukan dengan menggunakan metode *Maximum Likelihood Estimation* (MLE).

Uji perbandingan dilakukan untuk mengetahui efektifitas dari model yang diusulkan. Model yang dibandingkan yaitu logit *link*, probit *link*, dan model *link* yang dibangun dari distribusi Weibull dua parameter. Pengujian ini dilakukan dengan pemilihan model menggunakan pemilihan model terbaik yang ditentukan dengan menggunakan metode *Bayesian Information Criterion* (BIC), dan *Akaike Information Criterion* (AIC). Model dikatakan yang terbaik yaitu memiliki nilai terkecil dari masing-masing metode.

Beberapa penelitian yang telah dilakukan yaitu dalam jurnal yang ditulis oleh Georgina M.T[17] dalam penelitiannya membahas tentang aplikasi dari model regresi logit dan model regresi probit pada data kategori, estimasi parameter menggunakan prinsip dari MLE untuk memperoleh taksiran dari setiap parameter yang memaksimumkan fungsi likelihood. Prentice[13] di dalam penelitiannya melakukan uji kesesuaian fungsi *link* umum dua parameter yang berisi kasus khusus untuk model probit dan logit. Caron dan Polpo[3], secara singkat mengusulkan sebuah fungsi *link* asimetris yang dibangun dari distribusi Weibull tiga parameter khusus untuk kasus data biner. Selanjutnya, Caron, Renault, dkk[4] mengembangkan hasil penelitian sebelumnya dengan mengusulkan sebuah fungsi *link* asimetris, bernama Weibull *link* untuk data biner dan multinomial, dan membandingkan model yang diusulkan dengan model simetris yaitu logit *link*, probit *link*, dan lain-lain.

Penggunaan model regresi distribusi Weibull, probit *link*, dan logit *link* akan dianalisis untuk jenis data kategori. Berdasarkan penjelasan di atas, dalam skripsi ini penulis mengangkat judul “Perbandingan Model Regresi Distribusi Weibull dengan Beberapa Fungsi *Link* untuk Data Kategori”.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan sebelumnya, maka rumusan masalah dalam penelitian ini, yaitu:

1. Bagaimana model regresi Weibull, logit *link*, dan Probit *link* untuk data kategori?
2. Bagaimana estimasi parameter model regresi Weibull, model logit *link*, dan model probit *link* untuk data kategori menggunakan metode *Maximum Likelihood Estimation* (MLE)?
3. Bagaimana perbandingan dengan pemilihan model terbaik dari model regresi Weibull, logit *link*, dan probit *link* untuk data kategori menggunakan metode AIC dan BIC?

### 1.3 Batasan Masalah

Untuk mengarahkan pembahasan dalam penelitian ini agar tidak menyimpang dari sasaran yang dituju, maka perlu membuat batasan ruang lingkup permasalahan. Batasan masalah penelitian ini adalah sebagai berikut:

1. Distribusi Weibull yang digunakan yaitu distribusi Weibull dengan dua parameter.
2. Model logit *link* dan probit *link* digunakan untuk membandingkan model yang diusulkan.
3. Model regresi hanya bergantung pada satu variabel bebas yang berupa data kontinu dan satu variabel terikat yang merupakan data kategori.
4. Data yang digunakan berdistribusi binomial dan merupakan data sekunder.
5. Pemilihan model terbaik menggunakan metode AIC dan BIC.
6. *Software* yang digunakan yaitu R *Software*.

### 1.4 Tujuan dan Manfaat Penelitian

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

1. Membangun model regresi Weibull, logit *link*, dan Probit *link* untuk data kategori.
2. Menentukan estimasi parameter model regresi Weibull, model logit *link*, dan model probit *link* untuk data kategori menggunakan metode *Maximum Likelihood Estimation* (MLE).
3. Menentukan hasil perbandingan dengan pemilihan model terbaik dari model regresi Weibull, logit *link*, dan probit *link* untuk data kategori menggunakan metode AIC dan BIC.

Manfaat dari penelitian ini adalah membantu memudahkan untuk analisis data kategori dengan variabel respon atau variabel dependennya berupa data biner yang berdistribusi binomial. Selain itu, karena dalam penelitian ini bersifat aplikatif maka model yang diusulkan dapat diterapkan untuk analisis kasus yang lainnya.

## 1.5 Metode Penelitian

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

1. Merumuskan masalah, yaitu dimulai dari suatu masalah data kategori.
2. Melakukan studi kepustakaan terhadap berbagai referensi yang berkaitan dengan penelitian yang dilakukan bersumber dari buku, jurnal, dan penelitian sebelumnya. Kemudian, memahami secara mendalam tentang model regresi distribusi Weibull untuk data kategori.
3. Menentukan data, dalam penelitian ini menggunakan data sekunder.
4. Melakukan analisis, yaitu melakukan estimasi parameter dengan menggunakan metode MLE. Selanjutnya, melakukan perbandingan model dengan pemilihan model terbaik menggunakan *Akaike Information Criterion* (AIC), dan *Bayesian Information Criterion* (BIC) terhadap data yang telah diperoleh.
5. Perhitungan data dan analisis dibantu dengan menggunakan *R software*.
6. 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 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 yang melandasi pembahasan dalam penulisan skripsi ini. Secara garis besar, bab ini mencakup semua yang berkaitan dengan fungsi *link*, distribusi Weibull, probit *link*, logit *link*, metode *Maximum Likelihood Estimation* (MLE), metode *Newton Raphson*, pemilihan model terbaik yaitu *Akaike*

*Information Criterion (AIC) dan Bayesian Information Criterion (BIC).*

**BAB III : PERBANDINGAN MODEL REGRESI DISTRIBUSI WEIBULL DENGAN BEBERAPA FUNGSI *LINK* UNTUK DATA KATEGORI**

Bab ini berisi pembahasan utama dari skripsi yang dikaji, meliputi pembahasan mengenai fungsi *link* yang dikembangkan dari distribusi Weibull, model probit *link*, dan logit *link*, serta estimasi parameter untuk ketiga fungsi *link* tersebut.

**BAB IV : STUDI KASUS DAN SIMULASI PERHITUNGAN**

Bab ini berisi data dan juga pembahasannya berupa perhitungan dengan bantuan *software* yaitu R *software* untuk mencari nilai estimasi parameter dari setiap model dan membandingkan dengan memilih model terbaik.

**BAB V : PENUTUP**

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

UNIVERSITAS ISLAM NEGERI  
SUNAN GUNUNG DJATI  
BANDUNG