

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Pendugaan parameter dalam masalah statistik pada variabel respon biner di berbagai bidang merupakan hal yang sangat penting termasuk dalam ilmu sosial, biologi, dan ekonomi. Variabel respon biner atau variabel respon yang bersifat biner bisa disebut juga variabel dikotomis. Variabel dikotomis adalah variabel yang hanya mempunyai dua kemungkinan nilai, misalnya sukses dan gagal. Sedangkan variabel prediktor sering disebut juga dengan kovariat atau variabel penjelas.

Dalam pendugaan parameter dibutuhkan model statistik, salah satunya yaitu *Generalized Linear Model* (GLM). *Generalized Linear Models* (GLM) merupakan perluasan dari model regresi linear ‘klasik’ dengan asumsi prediktor (variabel bebas) memiliki efek linear akan tetapi tidak mengasumsikan suatu distribusi tertentu dari respon (variabel terikat). Terdapat tiga komponen dalam *Generalized Linear Models* (GLM) yaitu komponen random (μ), komponen sistematis (η), dan fungsi *link* $g(\cdot)$. Komponen yang sering digunakan dalam *Generalized Linear Model* (GLM) adalah fungsi *link*. Fungsi *link* yang penting dan umum digunakan dalam *Generalized Linear Models* (GLM) merupakan fungsi *link* simetris, yaitu Logit *link* dan Probit *link*.

Regresi Logit digunakan untuk menggambarkan hubungan variabel dependen (variabel respon) dengan variabel independen (variabel prediktor) yang bersifat kategori, kontinu, atau bersifat keduanya. Sedangkan regresi probit digunakan untuk menggambarkan antara variabel dependen dan variabel independen.

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 dikotomis atau biner biasa disimbolkan Y dengan variabel prediktor yang mempunyai skala pengukuran bersifat dikotomis, polikotomis, atau kontinu.

Banyak penelitian yang menunjukkan bahwa fungsi *link* simetris menghasilkan ketidaksesuaian dan ketidakcocokan model yang baik, serta berlaku hanya untuk kasus yang terbatas. Hal ini juga dibuktikan bahwa ketika probabilitas respons biner mendekati 0 pada

tingkat yang berbeda dari tingkat (sebagai fungsi kovariat) pendekatan 1, fungsi *link* simetris tidak dapat sesuai.

Untuk mengatasi masalah keterbatasan fungsi *link* tersebut, maka diusulkan fungsi *link* asimetris berdasarkan distribusi Eksponensial. Penggunaan distribusi Eksponensial dikarenakan kesederhaan dari distribusinya. Pada tahun 1999 Guptu dan Kundu [1] memperkenalkan hasil penelitiannya berupa perluasan pada distribusi Eksponensial. Perluasan dari distribusi Eksponensial tersebut adalah distribusi Eksponensial yang dipangkatkan ini diambil dari salah satu fungsi kepadatan komulatif yang digunakan pada pertengahan abad 19 (Gompertz-Verhulst) untuk membandingkan tabel kematian dan menghasilkan pertumbuhan penduduk.

Distribusi Eksponensial yang dipangkatkan dapat digunakan untuk menganalisis keandalan atau reliabilitas dari data masa hidup suatu benda atau individu yang penaksir parameternya ditentukan dengan menggunakan metode maksimum *likelihood*. Selain itu dilakukan juga pencarian nilai parameter dari distribusi Eksponensial yang dipangkatkan dengan menggunakan metode *Newton Raphson*.

Berbagai penelitian yang telah dilakukan yaitu dalam jurnal yang ditulis oleh Georgina M.T [2] dalam penelitiannya membahas tentang aplikasi dari model regresi Logit dan model regresi Probit pada data kategori, estimasi parameter menggunakan prinsip dari MLE (*Maximum Likelihood Estimation*) untuk memperoleh taksiran dari setiap parameter yang memaksimumkan fungsi *likelihood*. Prentice [3] dalam penelitiannya melakukan uji kesesuaian fungsi *link* umum dua parameter yang berisi kasus khusus untuk model Probit dan Logit. Caron dan Polpo [4] secara singkat mengusulkan sebuah fungsi *link* asimetris yang dibangun dari distribusi Weibull tiga parameter khusus untuk kasus data biner. Selanjutnya, Caron, Renault, dkk [5] 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. Penelitian yang dilakukan oleh Caron, Renault, dkk [5], dijadikan sebagai landasan penulis, hanya saja penulis mencoba mengusulkan distribusi yang berbeda namun tetap asimetris, yaitu distribusi Eksponensial.

Dalam statistika, ada istilah uji perbandingan, uji perbandingan ini dilakukan untuk mengetahui efektifitas dari model yang di usulkan. Model yang akan di bandingkan yaitu Logit *link*, Probit *link*, dan model *link* yang di bangun dari distribusi Eksponensial yang dipangkatkan. 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 di katakana yang terbaik yaitu memiliki nilai terkecil dari masing-masing metode.

Berdasarkan paparan latar belakang diatas, pada skripsi ini akan dibahas mengenai model Eksponensial yang dipangkatkan yang dibangun dengan fungsi *link* untuk data respon kategori. Penulis memberikan judul **“Perbandingan Model Regresi Distribusi Eksponensial Yang Dipangkatkan Dengan Model Probit *Link* dan Logit *Link* Untuk Data Kategori”**.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah disampaikan sebelumnya, rumusan masalah dalam penulisan tugas akhir ini, yaitu:

1. Bagaimana model regresi Eksponensial yang dipangkatkan, Logit *link*, dan Probit *link* untuk data kategori?
2. Bagaimana estimasi parameter model regresi Eksponensial yang dipangkatkan, model Logit *link*, dan model Probit *link* untuk data kategori menggunakan metode *Maximum Likelihood Estimation* (MLE)?
3. Bagaimana perbandingan dan pemilihan model terbaik untuk model regresi Eksponensial yang dipangkatkan, Logit *link*, dan Probit *link* untuk data kategori menggunakan metode *Bayesian Information Criterion* (BIC), dan *Akaike Information Criterion* (AIC)?

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 Eksponensial yang digunakan yaitu distribusi Eksponensial yang dipangkatkan dengan tiga parameter.
2. Model regresi hanya bergantung pada satu variabel bebas yang berupa data kontinu dan satu variabel terikat yang merupakan data kategori.
3. Data yang digunakan berdistribusi binomial dan merupakan data sekunder.
4. Pemilihan model terbaik menggunakan metode AIC dan BIC.
5. *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 skripsi ini, sebagai berikut:

1. Membangun model regresi Eksponensial yang dipangkatkan, Logit *link*, dan Probit *link* untuk data kategori
2. Menentukan estimasi parameter model regresi Eksponensial yang dipangkatkan, model Logit *link*, dan model Probit *link* untuk data kategori menggunakan metode *Maximum Likelihood Estimation* (MLE).
3. Menentukan hasil perbandingan dan pemilihan model terbaik untuk model regresi Eksponensial yang dipangkatkan, Logit *link*, dan Probit *link* untuk data kategori menggunakan metode *Bayesian Information Criterion* (BIC), dan *Akaike Information Criterion* (AIC)

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:

- a. Merumuskan masalah, yaitu dimulai dari suatu masalah data kategori.
- b. 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 Eksponensial yang dipangkatkan untuk data kategori.
- c. Menentukan data, dalam penelitian ini menggunakan data sekunder.
- d. 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.
- e. Perhitungan data dan analisis dibantu dengan menggunakan R *software*.
- f. Memberikan kesimpulan berdasarkan hasil analisis.

1.6. Sistematika Penulisan

Berdasarkan sistematika penulisannya, skripsi ini terdiri atas lima bab, serta daftar pustaka, dimana setiap bab terdapat beberapa sub bab.

BAB I : PENDAHULUAN

Bab ini berisi latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, ruang lingkup penelitian, metodologi penelitian, dan sistematika penulisan.

BAB II : LANDASAN TEORI

Berisi teori-teori yang melandasi pembahasan dalam penulisan skripsi. Secara garis besar, bab ini mencakup semua yang berkaitan dengan model *link*, distribusi Eksponensial, metode yang digunakan untuk menentukan estimasi dari parameter yang digunakan, dan pemilihan model yang terbaik.

BAB III : PERBANDINGAN MODEL REGRESI DISTRIBUSI EKSPONENSIAL YANG DIPANGKATKAN DENGAN MODEL PROBIT *LINK* DAN LOGIT *LINK* UNTUK DATA KATEGORI

Berisi pembahasan utama dari tugas akhir yang dikaji, meliputi pembahasan mengenai model Eksponensial *link* yang digunakan, dan estimasi parameter pada model eksponensial *link*.

BAB IV : STUDI KASUS DAN SIMULASI

Berisi tentang studi kasus, simulasi uji coba model menggunakan fungsi *link*, serta analisis dan hasil simulasi

BAB V : PENUTUP

Berisi kesimpulan dari seluruh pembahasan yang telah dikaji berdasarkan tujuan yang ingin dicapai. Selain itu, diberikan saran untuk pengembangan lebih lanjut terhadap topik pembahasan selanjutnya.

DAFTAR PUSTAKA

LAMPIRAN