

BAB I

PENDAHULUAN

Model regresi merupakan bagian yang tidak terpisahkan dari beberapa analisis data dengan menggambarkan hubungan antara variabel tak bebas (respon) dengan satu atau lebih variabel bebas (prediktor). Model regresi yang sering digunakan adalah model regresi linier sederhana yang mempunyai bentuk persamaan

$$Y_i = \alpha + \beta X_i + \varepsilon_i \quad i = 1, 2, \dots, n \quad (1)$$

dengan Y_i merupakan variabel respon ke- i , X_i merupakan variabel prediktor ke- i , α dan β , merupakan parameter yang belum diketahui yang akan diestimasi, dan ε_i merupakan error yang diasumsikan mempunyai rata-rata 0 (nol) dan varian σ^2 (Hines & Montgomery, 1990).

Model regresi linier tersebut mempunyai asumsi bahwa variabel respon Y_i berdistribusi Normal dan bersifat kuantitatif. Akan tetapi sering juga ditemui suatu kasus dimana variabel respon Y_i berdistribusi tidak Normal dalam hal ini Y_i berdistribusi dalam keluarga eksponensial yaitu Y_i dapat berdistribusi Binomial, Poisson, Normal, atau Gamma, data dapat bersifat kualitatif atau dapat bersifat kuantitatif dan model tidak linier sehingga persamaan (1) tidak dapat digunakan. Untuk mengatasi kasus tersebut di atas maka digunakan generalisasi model linier (Dobson, 1990)

Generalisasi model linier membuat suatu fungsi link yang menghubungkan antara variabel prediktor dengan nilai rata-rata dari variabel respon Y_i , dan

membuat model menjadi linier. Setelah diperoleh hasil fungsi link-nya, kemudian hasil tersebut disubstitusikan ke dalam persamaan model awal (model tidak linier).

Generalisasi model linier mempunyai tiga komponen yaitu : variabel respon Y_i , $i = 1, 2, \dots, N$ yang diasumsikan mempunyai distribusi yang sama yaitu distribusi dari keluarga eksponensial, β merupakan himpunan parameter dan X adalah himpunan variabel penjelas, serta sebuah fungsi link yang menghubungkan antara variabel penjelas dengan nilai ekspektasi (rata-rata) dari variabel respon Y , dengan bentuk persamaan $g(\mu) = \mathbf{x}'\beta$, dengan $\mu = E(Y)$.

Dalam generalisasi model linier terdapat beberapa fungsi link di antaranya fungsi identitas, fungsi log, fungsi logit, fungsi probit, fungsi log-log komplementer, dan sebagainya (Agresti, 1996).

Dalam penulisan tugas akhir ini, akan dibahas mengenai fungsi link pada generalisasi model linier untuk data biner, dimana variabel respon Y_i termasuk ke dalam keluarga eksponensial dan berdistribusi binomial. Fungsi link yang digunakan adalah fungsi link probit, fungsi link logit, dan fungsi link log-log komplementer. Untuk itu, tujuan yang ingin dicapai dari penulisan tugas akhir ini adalah mengetahui fungsi link pada generalisasi model linier untuk data biner yaitu fungsi link probit untuk model probit, fungsi link logit untuk model logistik, dan fungsi link log-log komplementer untuk model nilai ekstrim, dimana model-model tersebut digunakan dalam menaksir probabilitas kejadian sukses dalam suatu percobaan, kemudian membandingkan nilai Deviansi dari ketiga model tersebut.

Dalam penulisan tugas akhir ini dibagi menjadi empat bab yaitu :

- Bab I mengenai pendahuluan yang menjelaskan tentang latar belakang, permasalahan, pembatasan masalah, tujuan penulisan, dan sistematika penulisan.
- Bab II berisi tentang teori dasar / penunjang yang menguraikan apa saja yang diperlukan sebagai pendukung untuk memperjelas pemahaman pada bagian inti permasalahan yaitu mengenai probabilita, variabel random kontinu, fungsi distribusi, nilai harapan, percobaan Bernoulli, distribusi Binomial, metode maksimum likelihood (MLE), deviansi, distribusi keluarga eksponensial, dan model regresi linier berganda,
- Bab III menguraikan tentang Generalisasi model linier yang meliputi distribusi probabilitas, fungsi link, fungsi link untuk data berdistribusi binomial, penaksiran parameter, pengujian kecocokan model, pengujian koefisien regresi secara individu, serta contoh penerapan generalisasi model linier untuk data berdistribusi binomial.
- Bab IV berisi kesimpulan.