

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Analisis regresi merupakan metode statistika yang digunakan untuk meramalkan sebuah variabel respon Y dari satu atau lebih variabel bebas X , selain itu juga digunakan untuk menaksir pengaruh-pengaruh variabel bebas X terhadap variabel respon Y (Johnson dan Wichern, 1992: 285).

Model regresi yang biasanya digunakan untuk menentukan hubungan variabel bebas X dengan variabel respon Y adalah model regresi linear, baik regresi linear sederhana maupun berganda. Variabel dalam regresi linear biasanya bertipe data kontinu (interval atau rasio), walaupun demikian data nominal atau ordinal dapat digunakan sebagai variabel. Dalam regresi linear, variabel yang bertipe data nominal atau ordinal dapat digunakan sebagai variabel bebas, tetapi tidak dapat digunakan sebagai variabel respon.

Untuk memahami mengapa variabel yang bertipe data nominal atau ordinal tidak dapat digunakan sebagai variabel respon dalam regresi linear, perhatikan situasi dimana variabel respon Y dalam suatu masalah regresi hanya mengambil dua nilai kemungkinan, yaitu 0 dan 1. Variabel respon ini bisa jadi diperoleh dari suatu pengamatan terhadap respon yang bertipe data kualitatif (nominal atau ordinal). Misalnya, dalam suatu studi ingin diketahui penyebab kecelakaan lalu lintas kendaraan roda empat di suatu kota, dimana jika penyebabnya tidak

digunakannya sabuk pengaman (*seat-belt*) saat mengemudi, kategori variabel responnya diberi label 1 dan jika penyebabnya karena faktor lain, misalnya mabuk, mengantuk, sedang menelepon ketika mengemudi, kategori variabel responnya diberi label 0.

Andaikan bahwa model regresi dari permasalahan di atas memiliki bentuk

$$y_i = \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i \quad (1.1)$$

Dimana $\mathbf{x}_i = (x_{i0}, x_{i1}, \dots, x_{ik})'$, $\boldsymbol{\beta}' = (\beta_0, \beta_1, \dots, \beta_k)$, dan variabel respon y_i mengambil nilai 0 atau 1. Diasumsikan bahwa variabel respon y_i adalah suatu variabel acak Bernoulli dengan distribusi peluang sebagai berikut:

y_i	Peluang
1	$P(y_i = 1) = \pi_i$
0	$P(y_i = 0) = 1 - \pi_i$

Karena $E(\varepsilon) = 0$, dan nilai ekspektasi dari variabel respon adalah

$$E(y_i) = 1(\pi_i) + 0(1 - \pi_i) = \pi_i$$

Yang menunjukkan bahwa

$$E(y_i) = \boldsymbol{\beta}'\mathbf{x}_i = \pi_i \quad (1.2)$$

Hal ini berarti bahwa ekspektasi dari variabel respon yang diberikan menurut fungsi variabel respon $E(y_i) = \boldsymbol{\beta}'\mathbf{x}_i$ adalah peluang dari variabel respon yang mengambil nilai 1.

Sen dan Srivastava (1990: 11) mengemukakan bahwa ada tiga kondisi yang harus dipenuhi oleh regresi linear yang menggunakan metode kuadrat terkecil

untuk menaksir parameter-parameter agar dihasilkan taksiran parameter yang baik, yaitu:

1. Rata-rata dari galat adalah nol,

$$E(\varepsilon_i) = 0, \text{ untuk semua } i$$

2. Variansi dari galat adalah suatu konstanta,

$$\text{Var}(\varepsilon_i) = E(\varepsilon_i - E(\varepsilon_i))^2 = E(\varepsilon_i^2) = \sigma^2, \text{ untuk semua } i$$

3. Galat-galatnya tidak berkorelasi atau saling bebas,

$$E(\varepsilon_i \varepsilon_j) = 0, \text{ untuk semua } i \neq j$$

Ketiga kondisi di atas disebut kondisi *Gauss-Markov*, dengan galat harus berdistribusi normal.

Namun, model regresi linear dalam persamaan (1.1) di atas melanggar dua kondisi pertama Gauss-Markov dan batasan nilai duga (*fitted value*) y_i . Pelanggaran terhadap kondisi Gauss-Markov dan batasan nilai duga y_i tersebut adalah:

- a. Jika variabel responnya biner, maka galat ε_i dapat mengambil dua nilai,

$$\varepsilon_i = 1 - \beta' \mathbf{x}_i \quad \text{jika } y_i = 1$$

$$\varepsilon_i = -\beta' \mathbf{x}_i \quad \text{jika } y_i = 0$$

Akibatnya, galat dalam model ini tidak mungkin berdistribusi normal.

- b. Variansi dari galat bukanlah suatu konstanta, karena

$$\sigma_{y_i}^2 = E\{y_i - E(y_i)\}^2 = (1 - \pi_i)^2 \pi_i + (0 - \pi_i)^2 (1 - \pi_i) = \pi_i (1 - \pi_i)$$

Persamaan ini dapat ditulis sebagai

$$\sigma_{y_i}^2 = E(y_i)[1 - E(y_i)]$$

karena $E(y_i) = \beta' \mathbf{x}_i = \pi_i$. Hal ini mengindikasikan bahwa variansi dari pengamatan (yang mana sama seperti variansi galat sebab $\varepsilon_i = y_i - \beta' \mathbf{x}_i = y_i - \pi_i$, dan π_i adalah suatu konstanta) adalah suatu fungsi rata-rata.

c. Pembatasan pada fungsi variabel respon,

$$0 \leq E(y_i) = \pi_i \leq 1$$

Perhatikan bahwa peluang π_i harus terletak di antara 0 dan 1. Sementara $E(y_i)$ pada persamaan (1.2) bisa mengambil nilai di luar interval 0 dan 1 untuk setiap nilai x yang kecil atau besar.

Dengan demikian, berdasarkan a, b, dan c, model regresi linear kurang tepat digunakan jika variabel responnya bertipe data nominal atau ordinal.

Untuk mengatasi masalah ini, diperkenalkanlah regresi logistik. Regresi logistik merupakan bagian dari metode statistika yang dikenal sebagai model linear yang digeneralisasi (*generalized linear model*). Menurut Kleinbaum (1994: 5), regresi logistik adalah suatu pendekatan pemodelan matematika yang dapat digunakan untuk menggambarkan hubungan variabel bebas X dan variabel respon Y , dimana Y adalah variabel respon biner yaitu variabel yang memiliki dua nilai kemungkinan atau kategori.

Regresi logistik dapat dibedakan berdasarkan variabel bebas dan variabel responnya. Berdasarkan variabel bebasnya, regresi logistik terbagi menjadi dua, yaitu regresi logistik sederhana (*simple logistic regression*) dan regresi logistik

berganda (*multiple logistic regression*). Sedangkan berdasarkan variabel responnya, regresi logistik terbagi menjadi dua, yaitu: regresi logistik biner (*binary logistic regression*) dan regresi logistik multinomial (*multinomial logistic regression*). Regresi logistik biner digunakan ketika hanya ada dua kategori (*dichotomous*) variabel respon Y . Sedangkan regresi logistik multinomial digunakan ketika pada variabel respon Y terdapat lebih dari dua kategori (*polytomous* atau *polychotomous*).

Data respon *polytomous* banyak ditemui dalam kehidupan sehari-hari. Misalnya, dalam suatu studi mengenai pilihan sistem transportasi untuk bekerja, variabel responnya bisa berupa mobil pribadi, bus, busway, kereta api, sepeda motor, sepeda, atau jalan kaki. Dalam hal ini, variabel responnya memiliki tujuh kategori. Pemilihan transportasi ini akan bergantung pada variabel bebasnya. Variabel bebasnya bisa berupa jenis kelamin, ongkos, lama perjalanan sampai ke tempat bekerja, kenyamanan, dan sebagainya. Untuk menaksir variabel bebas tersebut terhadap variabel responnya yang bertipe data kategorik dapat digunakan model regresi logistik multinomial.

Kategori variabel respon sistem transportasi untuk bekerja dalam contoh di atas tidak terurut. Variabel respon dengan kategori tidak terurut dinamakan variabel respon nominal dan variabel respon dengan kategori terurut dinamakan variabel respon ordinal. Ada situasi dimana kategori responnya terurut. Misalnya, dalam suatu studi mengenai kepuasan pelanggan terhadap pelayanan toko A, variabel responnya bisa berupa sangat puas, puas, tidak puas, sangat tidak puas. Dalam suatu percobaan klinik (*clinical trial*), respon untuk suatu perlakuan bisa

diklasifikasi sebagai semakin membaik, tidak ada perubahan, semakin memburuk. Untuk kedua situasi ini digunakan suatu metode yang berbeda yang disebut model odds proporsional (*proportional odds model*) (Chatterjee dan Hadi, 2006: 329). Dalam tugas akhir ini, yang akan menjadi bahasan penulis adalah model regresi logistik multinomial dengan kategori variabel respon tidak terurut (nominal).

Untuk lebih memahami regresi logistik multinomial ini, penulis mengambil contoh kasus mengenai pengaruh gender dan ras terhadap pemilihan suatu partai politik. Untuk selanjutnya, tugas akhir ini diberi judul **“REGRESI LOGISTIK MULTINOMIAL”**.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang masalah di atas, rumusan masalah tugas akhir ini adalah:

1. Bagaimanakah cara menaksir parameter pada regresi logistik multinomial?
2. Bagaimana cara menguji keberartian parameter pada regresi logistik multinomial?
3. Bagaimana cara menguji kecocokan model pada regresi logistik multinomial?
4. Bagaimana aplikasi regresi logistik multinomial pada suatu contoh kasus mengenai pengaruh variabel bebas gender dan ras terhadap pemilihan suatu partai politik?

1.3 Batasan Masalah

Adapun batasan masalah pada tugas akhir ini adalah:

1. Model yang digunakan dalam regresi logistik multinomial adalah model logit yang digeneralisasi (*Generalized Logit Model*) dan disebut juga sebagai model logit multinomial.
2. Variabel respon yang digunakan merupakan variabel respon nominal atau variabel respon kategori yang tidak memiliki urutan data.
3. Penggunaan perangkat lunak SPSS versi 16.0 dalam pengolahan data.

1.4 Tujuan Penulisan

Berdasarkan uraian rumusan masalah di atas, tujuan penulisan tugas akhir ini adalah:

1. Mengetahui taksiran parameter pada regresi logistik multinomial.
2. Mengetahui uji keberartian parameter pada regresi logistik multinomial.
3. Mengetahui uji kecocokkan model pada regresi logistik multinomial.
4. Mengetahui aplikasi regresi logistik multinomial pada contoh kasus mengenai pengaruh variabel bebas gender dan ras terhadap pemilihan suatu partai politik.

1.5 Sistematika Penulisan

BAB I: PENDAHULUAN

Bab ini berisi tentang latar belakang masalah, rumusan masalah, tujuan penulisan, dan sistematika penulisan.

BAB II: LANDASAN TEORI

Bab ini berisi tentang teori-teori dasar yang digunakan untuk mendukung pernyataan pada bab-bab selanjutnya.

BAB III: REGRESI LOGISTIK MULTINOMIAL

Bab ini membahas tentang regresi logistik multinomial.

BAB IV: CONTOH KASUS

Bab ini berisi contoh kasus dari regresi logistik multinomial beserta pengolahan data dan hasil analisisnya.

BAB V: PENUTUP

Bab ini berisi kesimpulan dari bab-bab sebelumnya, juga disertakan saran dari penulis.

