BAB II

TEORI PENUNJANG

Untuk dapat melakukan analisis statistik dengan benar, diperlukan pengetahuan yang benar tentang teori – teori statistik. Berikut disampaikan beberapa teori yang menunjang analisis statistik dalam pembahasan.

2.1 Probabilitas Bersyarat

Probabilitas terjadinya suatu kejadian B bila diketahui kejadian A telah terjadi disebut dengan probabilitas bersyarat, dan dinyatakan dengan P(B|A). Keadaan ini dirumuskan sebagai berikut:

$$P(B|A) = \frac{P(A \cap B)}{P(A)}, \text{ bila } P(A)>0 \qquad (2.1.1)$$

Contoh:

Misalkan diketahui sejumlah orang dewasa yang telah tamat SMA sebagai ruang sampel di suatu kota kecil. Mereka dikelompokkan menurut jenis kelamin dan status pekerjaan.

Tabel 2.1.1

	bekerja	tidak bekerja	jumlah
lakli-laki	460	40	500
Wanita	140	260	400
Jumlah	600	300	900

This document is Undip Institutional Repository Collection. The author(s) or copyright owner(s) agree that UNDIP-IR may, without changing the content, translate the submission to any medium or format for the purpose of preservation. The author(s) or copyright owner(s) also agree that UNDIP-IR may keep more than one copy of this submission for purpose of security, back-up and preservation:
(http://eprints.undip.ac.id)

Daerah tersebut akan dijadikan daerah pariwisata dan seseorang dipilih secara acak untuk mempropagandakannya ke seluruh negeri. Akan diteliti kejadian berikut:

M : laki – laki yang terpilih

E : orang yang terpilih dalam status bekerja

Berapa peluang terpilih seorang laki – laki dalam status bekerja?

Jawab:

$$P(M|E) = \frac{P(E \cap M)}{P(E)}$$

$P(E \cap M)$ dan $P(E)$ diperoleh dari ruang sampel, sebagai berikut:

$P(E) = 600/900 = 2/3$ dan $P(E \cap M) = 460/900 = 23/45$

Jadi $P(M|E) = \frac{23/45}{2/3} = \frac{23}{30}$ (Walpole dan Myers,1989).

2.2 Macam – macam Skala Pengukuran Data

Dalam analisis statistika ditemui data kuantitatif yang dihasilkan melalui perhitungan dan pengukuran, maupun data kualitatif (data kategori) yang dikelompokkan menurut kriteria tertentu. Analisis terhadap data hasil pengamatan bergantung pada tingkat atau skala pengukuran data. Skala pengukuran data dapat dikelompokkan sebagai berikut:

1. Skala nominal

Merupakan skala pengukuran data yang menggunakan angka – angka untuk mengelompokkan obyek dan tidak ada urutan tertentu.

This document is Undip Institutional Repository Collection. The author(s) or copyright owner(s) agree that UNDIP-IR may, without changing the content, translate the submission to any medium or format for the purpose of preservation. The author(s) or copyright owner(s) also agree that UNDIP-IR may keep more than one copy of this submission for purpose of security, back-up and preservation:
(http://eprints.undip.ac.id)

Contoh:

a. pemberian kode pada hasil produksi; 0 jika rusak dan 1 jika baik,

b. pemberian kode 0 jika sakit dan 1 jika sehat.

2. Skala ordinal

Merupakan skala pengukuran data yang menggunakan angka – angka untuk mengelompokkan obyek, serta ada urutan tertentu berdasarkan kriteria tertentu.

Contoh:

a. penggolongan PNS; golongan IA, IIA, IIIA, dan seterusnya,

b. jenjang struktural pemerintahan desa; lurah sekdes, dan sebagainya.

3. Skala interval

Merupakan skala pengukuran data yang menggunakan angka – angka untuk mengelompokkan obyek, ada urutan berdasarkan kriteria tertentu serta jarak antara dua angka skala diketahui ukurannya.

Contoh:

Pengukuran suhu dengan skala Celcius dan Fahrenheith.

Celcius	0	10	20	30
Fahrenheith	32	50	68	86

4. Skala rasio

Merupakan skala pengukuran data yang menggunakan angka – angka untuk menunjukkan pengelompokkan data, ada urutan berdasarkan kriteria tertentu, jarak antara dua angka skala diketahui ukurannya dan mempunyai titik nol sejati.

Contoh:

Pengukuran tinggi badan, bobot badan, dan pendapatan (Agus Rusgiono dkk, 2000).

This document is Undip Institutional Repository Collection. The author(s) or copyright owner(s) agree that UNDIP-IR may, without changing the content, translate the submission to any medium or format for the purpose of preservation. The author(s) or copyright owner(s) also agree that UNDIP-IR may keep more than one copy of this submission for purpose of security, back-up and preservation:
(http://eprints.undip.ac.id)

2.3 Distribusi Binomial dan Multinomial

Dilakukan sebuah percobaan yang mempunyai hasil yang dapat diklasifikasikan sebagai sukses atau gagal. Misalkan Y = 1 jika hasilnya sukses dan Y = 0 jika hasilnya gagal. Variabel random Y dikatakan variabel random Bernoulli jika distribusi probabilitas adalah

$$f(y) = \pi^y (1-\pi)^{1-y}, \qquad 0 \le \pi \le 1, \tag{2.3.1}$$

dengan π adalah probabilitas bahwa percobaan sukses (Walpole dan Myers,1989).

Jika dilakukan n percobaan independen dengan hasil sukses mempunyai probabilitas π dan gagal mempunyai probabilitas $1 - \pi$ dan Y mewakili jumlah sukses yang terjadi dalam n percobaan, maka Y dikatakan variabel random binomial dengan parameter (n,π). Distribusi probabilitas dari variabel random binomial yang mempunyai parameter (n, π) adalah

$$b(y;\pi,n) = \binom{n}{y} \pi^n (1-\pi)^{n-y}, \qquad y = 0, 1, 2, \ldots, n. \tag{2.3.2}$$

(Walpole dan Myers,1989).

Contoh 1: (Walpole dan Myers,1989)

Probabilitas seorang penderita kanker darah dapat sembuh adalah 0,4. Jika diketahui ada 15 orang yang telah mengidap penyakit tersebut, berapakah probabilitas

a. paling sedikit 10 akan sembuh,

b. antara 3 sampai 8 akan sembuh,

c. tepat 5 yang sembuh.

This document is Undip Institutional Repository Collection. The author(s) or copyright owner(s) agree that UNDIP-IR may, without changing the content, translate the submission to any medium or format for the purpose of preservation. The author(s) or copyright owner(s) also agree that UNDIP-IR may keep more than one copy of this submission for purpose of security, back-up and preservation:
(http://eprints.undip.ac.id)

Penyelesaian:

Misalkan Y banyaknya penderita yang sembuh, maka

a. $P(Y \geq 10) = 1 - P(Y < 10)$

$$= 1 - \sum_{y=0}^{9} b(y;15,(0,4))$$

$$= 1 - 0{,}9662$$

$$= 0{,}0338,$$

b. $P(3 \leq Y \leq 8) = \sum_{y=3}^{8} b(y;15,(0,4))$

$$= \sum_{y=0}^{8} b(y;15,(0,4)) - \sum_{y=0}^{3} b(y;15,(0,4))$$

$$= 0{,}9050 - 0{,}0271$$

$$= 0{,}8779,$$

c. $P(Y=5) = b(5;15,(0,4))$

$$= \sum_{y=0}^{5} b(y;15,(0,4)) - \sum_{y=0}^{4} b(y;15,(0,4))$$

$$= 0{,}4032 - 0{,}2173$$

$$= 0{,}1859.$$

Percobaan multinomial merupakan perluasan dari percobaan binomial. Percobaan binomial menjadi percobaan multinomial jika setiap usaha dapat memberikan kemungkinan hasil lebih dari dua. (Walpole dan Myers,1989).

Contoh:

a. pembagian hasil pabrik menjadi ringan, berat, dan masih dapat diterima,

This document is Undip Institutional Repository Collection. The author(s) or copyright owner(s) agree that UNDIP-IR may, without changing the content, translate the submission to any medium or format for the purpose of preservation. The author(s) or copyright owner(s) also agree that UNDIP-IR may keep more than one copy of this submission for purpose of security, back-up and preservation:
(http://eprints.undip.ac.id)

b. pencatatan kecelakaan disuatu simpang jalan menurut hari dalam seminggu.

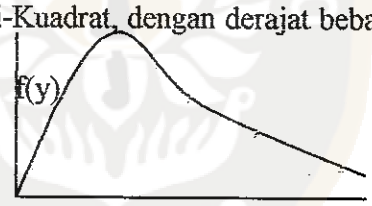
Jika suatu usaha tertentu dapat menghasilkan k macam hasil $E_1, E_2, E_3, \ldots, E_r$ dengan probabilitas $\pi_1, \pi_2, \ldots, \pi_r$, maka distribusi probabilitas peubah acak $Y_1, Y_2, Y_3, \ldots, Y_r$ yang menyatakan banyak terjadinya $E_1, E_2, E_3, \ldots, E_r$ dalam n usaha bebas adalah:

$$f(y_1, y_2, y_3, \ldots, y_r; \pi_1, \pi_2, \pi_3, \ldots, \pi_r, n) = \binom{n}{y_1, y_2, y_3, \ldots, y_r} \pi_1^{y_1} \pi_2^{y_2} \pi_3^{y_3} \ldots \pi_r^{y_r},$$

dengan $\sum_{j=1}^{r} y_j = n$ dan $\sum_{j=1}^{r} \pi_j = 1$ (Walpole dan Myers,1989).

2.4 Distribusi Chi-Kuadrat

Peubah acak kontinyu Y berdistribusi Chi-Kuadrat, dengan derajat bebas v bila fungsi padat peluang dan grafiknya berbentuk



$$f(y) = \begin{cases} \frac{1}{2^{v/2}\Gamma(v/2)} y^{(v/2)-1} e^{-y/2}, & y>0 \\ 0, & \text{untuk yang lainnya,} \end{cases} \qquad (2.4.1)$$

dengan v bilangan bulat positif. Rata- rata dan variannya adalah $\mu = v$ dan $\sigma^2 = 2v$ (Walpole dan Myers,1989).

Distribusi Chi-Kuadrat merupakan hal khusus dari distribusi gamma. Distribusi Chi-Kuadrat diperoleh dengan mengambil $\alpha = v/2$ dan $\beta = 2$, untuk v bilangan bulat positif.

Pada distribusi gamma, peubah acak kontinyu Y berdistribusi gamma dengan parameter α dan β bila fungsi padat peluangnya berbentuk

This document is Undip Institutional Repository Collection. The author(s) or copyright owner(s) agree that UNDIP-IR may, without changing the content, translate the submission to any medium or format for the purpose of preservation. The author(s) or copyright owner(s) also agree that UNDIP-IR may keep more than one copy of this submission for purpose of security, back-up and preservation:
(http://eprints.undip.ac.id)

$$f(y) = \begin{cases} \frac{1}{\beta^{\alpha}\Gamma(\alpha)} y^{\alpha-1} e^{-y/\beta}, & y>0 \\ 0, & \text{untuk yang lainnya,} \end{cases} \tag{2.4.2}$$

dengan $\alpha > 0$ dan $\beta > 0$. Rata – rata dan variannya adalah $\mu = \alpha\beta$ dan $\sigma^2 = \alpha\beta^2$ (Walpole dan Myers,1989).

2.5 Distribusi Normal

Distribusi peluang kontinyu yang paling penting dalam seluruh bidang statistika adalah distribusi normal. Suatu peubah acak kontinyu Y dikatakan berdistribusi normal apabila mempunyai fungsi padat peluang sebagai berikut:

$$f(y;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2((y-\mu)/\sigma)^2}$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right), \quad -\infty < y < \infty. \tag{2.5.1}$$

Teorema 2.5.1

Jika peubah acak $Y \sim N(\mu,\sigma^2)$, $\sigma^2 > 0$, maka peubah acak $W = \frac{(y-\mu)}{\sigma} \sim N(0,1)$.

Bukti:

Karena $\sigma^2 > 0$, maka $\sigma > 0$. Sehingga fungsi distribusi G(w) dari W adalah

$$G(w) = \Pr\left(\frac{y-\mu}{\sigma} \le w\right)$$

$$= \Pr\left(y \le \sigma w + \mu\right)$$

This document is Undip Institutional Repository Collection. The author(s) or copyright owner(s) agree that UNDIP-IR may, without changing the content, translate the submission to any medium or format for the purpose of preservation. The author(s) or copyright owner(s) also agree that UNDIP-IR may keep more than one copy of this submission for purpose of security, back-up and preservation:
(http://eprints.undip.ac.id)

$$= \int_{-\infty}^{\sigma w+\mu} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(y-\mu)}{2\sigma^2}\right) dy. \qquad (2.5.2)$$

Jika peubah integrasinya diubah dengan menulis $x = \frac{(y-\mu)}{\sigma}$,maka diperoleh $dy = \sigma$ dx. Sehingga

$$G(w) = \int_{-\infty}^{w} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \qquad (2.5.3)$$

Fungsi padat peluang $g(w) = G^{-1}(w)$.Sehingga

$$g(w) = \frac{1}{\sqrt{2\pi}} e^{-w^2/2}, \qquad -\infty < w < \infty \qquad (2.5.4)$$

Jadi $W \sim N(0,1)$.

Teorema 2.5.2

Jika peubah acak $Y \sim N(\mu,\sigma^2)$, $\sigma^2 > 0$, maka peubah acak $V = \left(\frac{y-\mu}{\sigma}\right)^2 \sim \chi^2$.

Bukti:

Karena $Y \sim N(\mu,\sigma^2)$, maka menurut teorema 2.5.1, $W = \frac{y-\mu}{\sigma} \sim N(0,1)$.

Jika $V = \left(\frac{y-\mu}{\sigma}\right)^2$, maka $V = W^2$. Sehingga fungsi distribusi G(v) dari V adalah

$$\begin{aligned} G(v) &= \Pr(V \le v) \\ &= \Pr(W^2 \le v) \\ &= \Pr(|W| \le \sqrt{v}) \\ &= \Pr(-\sqrt{v} \le W \le \sqrt{v}), \text{ untuk } v \ge 0 \end{aligned}$$

This document is Undip Institutional Repository Collection. The author(s) or copyright owner(s) agree that UNDIP-IR may, without changing the content, translate the submission to any medium or format for the purpose of preservation. The author(s) or copyright owner(s) also agree that UNDIP-IR may keep more than one copy of this submission for purpose of security, back-up and preservation:
(http://eprints.undip.ac.id)

$$= \int_{-\sqrt{v}}^{\sqrt{v}} \frac{1}{\sqrt{2\pi}} e^{-w^2/2} dw, \text{ untuk } v \geq 0$$

$$= 2\int_{0}^{\sqrt{v}} \frac{1}{\sqrt{2\pi}} e^{-w^2/2} dw. \qquad (2.5.5)$$

Jika peubah integrasinya diubah dengan menulis $w = \sqrt{y}$, maka diperoleh $dw = ½\, y^{-1/2}\, dy$. Sehingga diperoleh

$$G(v) = \frac{2}{2}\int_{0}^{v} \frac{1}{\sqrt{2\pi}\sqrt{y}} e^{-y^2/2} dy,\ v \geq 0. \qquad (2.5.6)$$

Fungsi padat peluang $g(v) = G'(v)$.

$$g(v) = \begin{cases} \frac{1}{\sqrt{2}\sqrt{\pi}} v^{1/2-1} e^{-v/2}, 0 < v < \infty \\ 0, \quad \text{untuk yang lainnya.} \end{cases} \qquad (2.5.7)$$

Karena $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$. Sehingga

$$g(v) = \begin{cases} \frac{1}{\Gamma(1/2)\sqrt{2}} v^{1/2-1} e^{-v/2}, 0 < v < \infty \\ 0, \quad \text{untuk yang lainnya.} \end{cases} \qquad (2.5.8)$$

2.6 Metode Maksimum Likelihood

Metode maksimum likelihood digunakan untuk menaksir parameter yang tidak diketahui. Prinsip metode ini adalah mencari nilai β dengan memaksimumkan fungsi likelihood.

This document is Undip Institutional Repository Collection. The author(s) or copyright owner(s) agree that UNDIP-IR may, without changing the content, translate the submission to any medium or format for the purpose of preservation. The author(s) or copyright owner(s) also agree that UNDIP-IR may keep more than one copy of this submission for purpose of security, back-up and preservation:
(http://eprints.undip.ac.id)

Misalkan Y variabel random dengan distribusi probabilitas $f(\pi_j(x_i);y_{ji})$. Misalkan $y_1, y_2, y_3, \ldots, y_n$ menjadi nilai yang diobservasi didalam suatu sampel random yang besarnya n, maka fungsi likelihoodnya adalah:

$$l(\beta) = \prod_{i=1}^{n} f(\pi_j(x_i); y_{ji}), \qquad j = 0, 1, 2, \ldots, r \qquad (2.6.1)$$

Untuk memaksimumkan $l(\beta)$ dibentuk logaritma dari fungsi ini, $L(\beta) = \ln l(\beta)$. Menurut hitung defferensial, persamaan maksimum likelihood didapat dari turunan parsial pertama $L(\beta)$ terhadap β, kemudian disamadengankan dengan nol.

Dalam regresi logistik, persamaan log likelihoodnya bukan merupakan fungsi yang linier dalam β, sehingga harga taksiran β dicari dengan menggunakan iterasi Newton-Raphson.

2.7 Metode Newton-Raphson

Metode Newton-Raphson merupakan sebuah prosedur untuk mendapatkan suatu nilai pada fungsi yang dimaksimumkan. Dalam penggunaannya harus dibuat sebuah perkiraan awal nilai itu. Selanjutnya fungsi didekati melalui polynomial berderajat dua. Berikutnya fungsi didekati pada bagian kedua melalui polynomial berderajat tiga.

Secara lebih khusus, adalah sebagai berikut. Andaikan ingin didapatkan nilai β dari β yang memaksimumkan $L(\beta)$. Misalkan $q = \frac{\partial L(\beta)}{\partial \beta_j}$, misalkan $I = \frac{\partial^2 L(\beta)}{\partial \beta_j \partial \beta_k}$, dan misalkan q_m dan I_m merupakan iterasi ke-m dalam barisan pendekatan untuk $\hat{\beta}$. Pada

This document is Undip Institutional Repository Collection. The author(s) or copyright owner(s) agree that UNDIP-IR may, without changing the content, translate the submission to any medium or format for the purpose of preservation. The author(s) or copyright owner(s) also agree that UNDIP-IR may keep more than one copy of this submission for purpose of security, back-up and preservation:
(http://eprints.undip.ac.id)

langkah ke-m dari iterasi (m = 0, 1, 2, ...), $L(\beta)$ didekati dengan ekspansi deret Taylor,

$$Q_m(\beta) \equiv L(\beta_m) + q_m(\beta-\beta_m) + (1/2)(\beta-\beta_m)^1 I_m(\beta-\beta_m). \quad (2.7.1)$$

Pemecahannya $\frac{\partial Q_m}{\partial \beta} = q_m + I_m(\beta-\beta_m) = 0$, untuk β hasil pendekatan selanjutnya,

$$\beta_{m+1} = \beta_m + (I_m)^{-1} q_m, \quad (2.7.2)$$

dengan asumsi I_m nonsingular.

2.8 Model Regresi Linier

Model regresi linier merupakan salah satu cara dalam analisis data statistik. Model regresi linier terdiri dari model regresi linier sederhana dan model regresi linier berganda. Model regresi linier berganda adalah sebagai berikut:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_r x_{ri} + \varepsilon_i \quad (2.8.1)$$

Error diasumsikan berdistribusi normal dengan rata – rata nol dan varian konstan.

Taksiran parameternya diperoleh dengan menggunakan metode kuadrat terkecil dan metode maksimum likelihood.

2.9 Variabel Dummy

Variabel yang digunakan dalam analisis regresi biasanya variabel kuantitatif, yaitu variabel yang didefinisikan dengan skala ukuran. Kadang – kadang perlu digunakan variabel kategori atau variabel kualitatif. Secara umum dalam variabel kualitatif tidak ada skala ukuran mutlak.

This document is Undip Institutional Repository Collection. The author(s) or copyright owner(s) agree that UNDIP-IR may, without changing the content, translate the submission to any medium or format for the purpose of preservation. The author(s) or copyright owner(s) also agree that UNDIP-IR may keep more than one copy of this submission for purpose of security, back-up and preservation:
(http://eprints.undip.ac.id)

Jika dalam suatu masalah terdapat variabel prediktor yang berskala nominal, seperti jenis kelamin, ras, dan lain – lain, maka digunakan variabel dummy untuk mewakilinya dalam model. Pada umumnya jika sebuah variabel prediktor berskala nominal mempunyai r kategori, maka dibutuhkan r-1 variabel dummy. Jika variabel prediktor ke-j berskala nominal dengan r kategori, maka akan digunakan r-1 variabel dummyuntuk mewakili X_k dalam model. Jika variabel dummy tersebut dinyatakan dengan X_{ki} dan koefisien – koefisiennya dinyatakan dengan β_{ji}, j = 0, 1, 2, ..., r-1, maka bentuk fungsi logit untuk model dengan k variabel bebas dengan variabel ke-k berskala nominal adalah:

$$g_j(x_i) = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_{10} x_{11} + \ldots + \beta_{(r-1)i} x_{(r-1)i} + \beta_p x_p.$$

$$= \beta_0 + \beta_1 x_{1i} + \ldots + \sum_{j=0}^{r-1} \beta_{ji} x_{ki} + \beta_p x_p.$$

2.10 Model Regresi Logistik Biner

Regresi linier dan regresi logistik mempunyai kegunaan yang sama, yaitu untuk menyelidiki hubungan antara sebuah variabel respon dengan satu atau lebih variabel penjelas. Baik regresi linier maupun regresi logistik menggunakan metode estimasi parameter untuk memperoleh kecocokan model yang diharapkan. Regresi linier menggunakan metode kuadrat terkecil untuk memperoleh estimasi parameter yang diharapkan. Sedangkan regresi logistik menggunakan metode maksimum likelihood untuk memperoleh estimasi parameter yang diharapkan.

This document is Undip Institutional Repository Collection. The author(s) or copyright owner(s) agree that UNDIP-IR may, without changing the content, translate the submission to any medium or format for the purpose of preservation. The author(s) or copyright owner(s) also agree that UNDIP-IR may keep more than one copy of this submission for purpose of security, back-up and preservation:
(http://eprints.undip.ac.id)

Perbedaan pokok antara regresi linier dan regresi logistik terletak pada tipe variabel responnya. Regresi linier menggunakan variabel respon kontinyu, sedangkan regresi logistik menggunakan variabel respon berkategori.

Model regresi logistik terdiri dari model regresi logistik biner, nominal, dan ordinal. Model regresi logistik biner merupakan model regresi dengan variabel respon berskala nominal dengan dua kategori, dan variabel penjelasnya bisa berupa variabel kategori bisa pula variabel kontinyu. Model regresi logistik biner dapat dinyatakan sebagai berikut:

$$y = \pi(\mathbf{x}) + \varepsilon\,, \tag{2.10.1}$$

dengan $\pi(\mathbf{x}) = \dfrac{e^{\beta_0+\beta_1x_1+\beta_2x_2+\ldots+\beta_rx_r}}{1+e^{\beta_0+\beta_1x_1+\beta_2x_2+\ldots+\beta_rx_r}}$.

Kemudian dilakukan transformasi logit terhadap $\pi(\mathbf{x})$ sebagai berikut:

$$\begin{aligned}
g(\mathbf{x}) &= \ln\left[\frac{\pi(\mathbf{x})}{1-\pi(\mathbf{x})}\right] \\
&= \ln\left[\frac{\dfrac{e^{\beta_0+\beta_1x_1+\beta_2x_2+\beta_rx_r}}{1+e^{\beta_0+\beta_1x_1+\beta_2x_2+\beta_rx_r}}}{1-\dfrac{e^{\beta_0+\beta_1x_1+\beta_2x_2+\beta_rx_r}}{1+e^{\beta_0+\beta_1x_1+\beta_2x_2+\beta_rx_r}}}\right] \\
&= \ln\left[e^{\beta_0+\beta_1x_1+\beta_2x_2+\ldots+\beta_rx_r}\right] - \ln\left[1+e^{\beta_0+\beta_1x_1+\beta_2x_2+\ldots+\beta_rx_r}\right] - \\
&\quad \ln[1] + \ln\left[1+e^{\beta_0+\beta_1x_1+\beta_2x_2+\ldots+\beta_rx_r}\right] \\
&= \ln\left[e^{\beta_0+\beta_1x_1+\beta_2x_2+\ldots+\beta_rx_r}\right] \\
&= \beta_0 + \beta_1x_1 + \beta_2x_2 + \ldots + \beta_rx_r.
\end{aligned}$$

Sehingga model regresi logistik biner adalah:

This document is Undip Institutional Repository Collection. The author(s) or copyright owner(s) agree that UNDIP-IR may, without changing the content, translate the submission to any medium or format for the purpose of preservation. The author(s) or copyright owner(s) also agree that UNDIP-IR may keep more than one copy of this submission for purpose of security, back-up and preservation:
(http://eprints.undip.ac.id)

$$y = \pi(\mathbf{x}) + \varepsilon, \qquad (2.10.2)$$

dengan $\pi(\mathbf{x}) = \dfrac{e^{g(\mathbf{x})}}{1+e^{g(\mathbf{x})}}$

Y merupakan variabel respon dengan dua kategori, misalkan disimbolkan dengan y = 1 atau y = 0. Jika y = 1, maka $\varepsilon = 1 - \pi(\mathbf{x})$ dengan probabilitas $\pi(\mathbf{x})$. Sedangkan jika y = 0, maka $\varepsilon = -\pi(\mathbf{x})$ dengan probabilitas $1 - \pi(\mathbf{x})$. Jadi ε mempunyai suatu distribusi dengan mean 0 dan varian $\pi(\mathbf{x})(1 - \pi(\mathbf{x}))$. Dan probabilitas bersyarat variabel respon dapat ditulis dengan $P(y=1|\mathbf{x}) = \pi(\mathbf{x})$ dan $P(y=0|\mathbf{x}) = 1 - \pi(\mathbf{x})$. Dilakukan n percobaan yang saling bebas, dengan Y_i adalah variabel respon dari observasi ke-i yang berdistribusi Bernoulli yang memiliki probabilitas sukses $\pi(\mathbf{x}_i)$ dan probabilitas gagal $1 - \pi(\mathbf{x}_i)$. Y_i mempunyai fungsi densitas sebagai berikut:

$$f(\pi(\mathbf{x}_i);y_i) = \pi(\mathbf{x}_i)^{y_i}(1-\pi(\mathbf{x}_i))^{1-y_i} \quad y_i = 0, 1 \text{ dan } i = 1, 2, 3, \ldots, n. \qquad (2.10.3)$$

Karena observasinya saling bebas, maka fungsi likelihood didapat sebagai hasil perkalian dari masing – masing fungsi densitasnya, yaitu sebagai berikut:

$$l(\beta) = \prod_{i=1}^{n} \pi(\mathbf{x}_i)^{y_i}(1-\pi(\mathbf{x}_i))^{1-y_i}, \qquad (2.10.4)$$

dengan β adalah parameter yang tidak diketahui dan x_i adalah variabel penjelas dari observasi ke-i.

Prinsip dari metode maksimum likelihood adalah mencari nilai β dengan memaksimumkan fungsi likelihood. Oleh karena itu diperlukan turunan parsial pertama dari fungsi likelihood terhadap masing – masing parameter β_{jk}. Untuk lebih

This document is Undip Institutional Repository Collection. The author(s) or copyright owner(s) agree that UNDIP-IR may, without changing the content, translate the submission to any medium or format for the purpose of preservation. The author(s) or copyright owner(s) also agree that UNDIP-IR may keep more than one copy of this submission for purpose of security, back-up and preservation:
(http://eprints.undip.ac.id)

memudahkan dalam memaksimumkan fungsi likelihood, terlebih dahulu dibentuk log dari fungsi likelihood, yaitu sebagai berikut:

$$L(\beta) = \ln l(\beta)$$

$$= \sum_{i=1}^{n} \{y_i \ln[\pi(x_i)] + (1-y_i)\ln[1-\pi(x_i)]\}$$

$$= \sum_{i=1}^{n} \left\{ y_i \ln\left[\frac{e^{g(x_i)}}{1+e^{g(x_i)}}\right] + (1-y_i)\ln\left[1-\frac{e^{g(x_i)}}{1+e^{g(x_i)}}\right]\right\}$$

$$= \sum_{i=1}^{n} \{y_i \left(\ln\left[e^{g(x_i)}\right] - \ln\left[1+e^{g(x_i)}\right]\right) + (1-y_i)\left(\ln[1] - \ln\left[1+e^{g(x_i)}\right]\right)\}$$

$$= \sum_{i=1}^{n} \{y_i\, e^{g(x_i)} - \ln\left[1+e^{g(x_i)}\right]\} \qquad (2.10.5)$$

dengan $g(x_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_r x_{ri}$.

Sehingga persamaan (2.10.5) menjadi:

$$L(\beta) = \sum_{i=1}^{n} \{y_i \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_r x_{ri})$$

$$\ln[1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_r x_r)]\} \qquad (2.10.6)$$

Dari turunan pertama fungsi log likelihood terhadap β_0, β_1, β_2, ..., β_r, maka didapatkan (p+1) persamaan likelihood, dengan bentuk umum sebagai berikut:

$$\frac{\partial L(\beta)}{\partial \beta_j} = \sum_{i=1}^{n} x_{ji}(y_i - \pi(x_i)) = 0 \qquad (2.10.7)$$

dengan $j = 0, 1, 2, \ldots, r$ dan $x_{0i} = 1$, untuk setiap i.

Persamaan log likelihood pada persamaan (2.10.7) bukan merupakan fungsi yang linier dalam β_0, β_1, β_2, ..., β_r. Sehingga harga taksiran β dicari dengan

This document is Undip Institutional Repository Collection. The author(s) or copyright owner(s) agree that UNDIP-IR may, without changing the content, translate the submission to any medium or format for the purpose of preservation. The author(s) or copyright owner(s) also agree that UNDIP-IR may keep more than one copy of this submission for purpose of security, back-up and preservation:
(http://eprints.undip.ac.id)

menggunakan metode numerik. Metode yang digunakan untuk memecahkan masalah ini adalah metode Newton Raphson. Untuk itu diperlukan turunan parsial kedua dari fungsi log likelihood, dengan bentuk umum sebagai berikut:

$$\frac{\partial L(\beta)}{\partial \beta_j^{\,2}} = -\sum_{i=1}^{n} x_{ji}^{\,2}\pi(\mathbf{x}_i)(1-\pi(\mathbf{x}_i)) \qquad (2.10.8)$$

dan

$$\frac{\partial L(\beta)}{\partial \beta_j \partial \beta_u} = -\sum_{i=1}^{n} x_{ji}x_{ui}\pi(\mathbf{x}_i)(1-\pi(\mathbf{x}_i)), \quad u = 0, 1, 2, \ldots, p. \qquad (2.10.9)$$

Prosedur Newton Raphson untuk mencari taksiran b_j, j= 0, 1, 2, ..., r, yaitu sebagai berikut:

1. Pilih taksiran awal b_{jm}, misalkan $b_{j1} = 0$, m = 1, 2, 3, ... ,
2. Pada setiap iterasi ke-(m+1) hitung taksiran baru $b_{j(m+1)} = b_{jm} + [I(\beta)]^{-1}x'[y-\pi_i]$, dengan $[I(\beta)]^{-1}$ adalah invers dari matriks informasi yang merupakan matriks berukuran (p+1) x (p+1) yang memiliki elemen – elemen negatif dari nilai – nilai dalam persamaan (2.10.8) dan (2.10.9).
3. Iterasi berlanjut hingga diperoleh $b_{j(m+1)} \approx b_{jm}$ (Neneng Sitinurkhasanah, 2000).

This document is Undip Institutional Repository Collection. The author(s) or copyright owner(s) agree that UNDIP-IR may, without changing the content, translate the submission to any medium or format for the purpose of preservation. The author(s) or copyright owner(s) also agree that UNDIP-IR may keep more than one copy of this submission for purpose of security, back-up and preservation:
(http://eprints.undip.ac.id)