

PEMODELAN FAKTOR-FAKTOR YANG MEMPENGARUHI KEJADIAN DBD (DEMAM BERDARAH *DENGUE*) MENGGUNAKAN REGRESI LOGISTIK BINER UNTUK WILAYAH REGIONAL 2 INDONESIA (SUMATERA)

DINA MONICA, DODI DEVIANTO, FERRA YANUAR

*Program Studi Matematika,
Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Andalas,
Kampus UNAND Limau Manis Padang, Indonesia,
email : dinamonica77@yahoo.com*

Abstrak. Penelitian ini bertujuan untuk menjelaskan beberapa faktor yang mempengaruhi kejadian Demam Berdarah *Dengue* (DBD) pada kabupaten atau kota di wilayah regional 2 Indonesia (Sumatera) tahun 2012. Faktor-faktor tersebut menggunakan metode Regresi Logistik Biner yang merupakan salah satu teknik estimasi parameter dengan pendekatan *likelihood*. Pada penelitian ini diperoleh tiga variabel prediktor yang berpengaruh signifikan terhadap kejadian demam berdarah *dengue*. Variabel tersebut adalah rumah atau bangunan bebas jentik nyamuk Aedes, rumah tangga ber-PHBS dan sumur terlindung. Dengan nilai *Odds ratio* untuk rumah atau bangunan bebas jentik nyamuk Aedes, rumah tangga ber-PHBS, dan sumur terlindung masing-masing sebesar 0,968, 0,974, dan 0,980. Nilai *hit ratio* keakuratan model peluang logit sebesar 71,233%. Dengan demikian dapat disimpulkan bahwa model peluang logit yang terbentuk sudah layak digunakan untuk mengetahui faktor-faktor yang mempengaruhi kejadian DBD.

Kata Kunci: Model Regresi Logistik Biner, metode Maximum Likelihood, Demam Berdarah *Dengue*

1. Pendahuluan

Penyakit demam berdarah *dengue* (*Dengue Hemorrhagic Fever* / DHF) adalah suatu jenis penyakit yang disebabkan oleh virus *Dengue* (*Viral Dease*) dari genus *Flavivirus*, family *Flaviviridae* dengan perantara nyamuk *Aedes Aegypti* yang terinfeksi virus *Dengue*. Penyakit DBD merupakan salah satu penyakit yang perjalanan penyakitnya cepat dan dapat menyebabkan kematian dalam waktu singkat.

Meningkatnya penyakit demam berdarah ini dipengaruhi oleh faktor cuaca dan musim. Selain itu faktor yang paling menonjol yaitu faktor lingkungan. Kondisi lingkungan yang kotor dan tidak terjaga dengan baik akan membuat nyamuk berkembang biak. Faktor-faktor tersebut seperti tempat penampungan air sebagai tempat perindukan nyamuk *Aedes Aegypti*, ketinggian tempat suatu daerah, curah hujan pada musim hujan dan sanitasi lingkungan sehingga terjadinya perkembangan nyamuk *Aedes Aegypti*.

Kejadian DBD pada kabupaten atau kota yang ada di Sumatera perlu dimodelkan agar diketahui faktor-faktor apa saja yang mempengaruhi kejadian DBD di

daerah tersebut. Dalam model hubungan sebab akibat, maka kejadian DBD didefinisikan sebagai variabel respon dan faktor-faktor yang mempengaruhinya didefinisikan sebagai variabel prediktor. Hubungan antara variabel respon dengan faktor-faktor yang mempengaruhinya biasanya dapat dimodelkan dengan menggunakan analisis regresi. Karena kejadian DBD adalah variabel random diskrit yang bersifat biner (dikotomi), maka pada kasus ini salah satu cara untuk memodelkannya adalah dengan menggunakan regresi logistik biner. Oleh karena itu, pada penelitian ini akan diteliti faktor-faktor apa saja yang mempengaruhi kejadian DBD sehingga dapat ditentukan suatu pendugaan peluang terhadap kejadian DBD pada kabupaten atau kota yang ada di Sumatera dengan menggunakan analisis regresi logistik biner.

2. Model Regresi Logistik Biner

Model regresi logistik biner digunakan untuk memodelkan hubungan antara beberapa variabel prediktor dengan suatu variabel respon biner yaitu variabel biner yang hanya mempunyai dua nilai kemungkinan yang dinyatakan dengan 0 (gagal) dan 1 (sukses). Model regresi logistiknya [2] :

$$p_i = \frac{\exp(\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki})}{1 + \exp(\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki})}. \quad (2.1)$$

Jika $\eta_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}$ maka

$$p_i = \frac{1}{1 + e^{-\eta_i}}, i = 1, 2, \dots, n. \quad (2.2)$$

Logit dari p_i adalah

$$\text{Logit}(p_i) = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}; -\infty \leq \text{Logit}(p_i) \leq \infty. \quad (2.3)$$

2.1. Pendugaan Parameter pada Regresi Logistik Biner

Pendugaan parameter dalam regresi logistik dilakukan dengan cara metode maksimum *likelihood*. Metode tersebut menduga koefisien β dengan cara memaksimumkan fungsi *likelihood* dan mensyaratkan bahwa data harus mengikuti suatu distribusi tertentu. Pada regresi logistik biner, setiap pengamatan mengikuti distribusi Bernoulli sehingga dapat ditentukan fungsi *likelihood*-nya [2]. Fungsi kepadatan peluang distribusi Bernoulli sebagai berikut :

$$f(x_i) = p_i^{y_i} (1 - p_i)^{i - y_i}; y_i = 0, 1. \quad (2.4)$$

Setiap pasangan pengamatan diasumsikan bebas sehingga fungsi *likelihood*-nya merupakan gabungan dari fungsi distribusi masing-masing pasangan yaitu sebagai berikut :

$$l(\beta) = \prod_{i=1}^n f(x_i) = \prod_{i=1}^n p_i^{y_i} (1 - p_i)^{i - y_i}. \quad (2.5)$$

Fungsi *likelihood* tersebut lebih mudah dimaksimumkan dalam bentuk $\log L(\beta)$ dan dinyatakan dengan $L(\beta)$

$$\begin{aligned} L(\beta) &= l(\beta), \\ &= \ln \left[\prod_{i=1}^n p_i^{y_i} (1 - p_i)^{i - y_i} \right], \end{aligned}$$

sehingga diperoleh

$$L(\beta) = \sum_{i=1}^n [y_i \eta_i - \ln(1 + e^{\eta_i})]. \quad (2.6)$$

Untuk mencari taksiran dari parameter-parameter yaitu $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ fungsi \log *likelihood* pada persamaan \log *likelihood* diturunkan secara parsial terhadap masing-masing parameter yang bersesuaian kemudian disamakan dengan nol. Untuk mencari taksiran dari $\beta_0, \beta_1, \dots, \beta_k$ digunakan metode numerik yaitu metode Newton-Raphson. Metode Newton-Raphson digunakan untuk menemukan solusi dari fungsi \log *likelihood* sehingga diperoleh nilai konvergen untuk dijadikan sebagai taksiran untuk masing-masing parameter. Sehingga didapat taksiran untuk model regresi logistik biner yaitu :

$$\hat{\beta}_{i+1} = \hat{\beta}_i + \{\mathbf{X}^T \mathbf{V} \mathbf{X}\}^{-1} \{\mathbf{X}^T (\mathbf{Y} - \mathbf{P}_i)\}. \quad (2.7)$$

2.2. Interpretasi Koefisien Regresi Logistik Biner

Misalkan diketahui model regresi logistik dengan k variabel prediktor sebagai berikut.

$$\ln\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki}, \quad i = 1, 2, \dots, n.$$

Persamaan ini juga dapat ditulis menjadi

$$\frac{p_i}{1 - p_i} = \exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki}). \quad (2.8)$$

Ruas kiri dari Persamaan (2.8) merupakan perbandingan antara peluang berhasil p_i dengan peluang gagal $1 - p_i$ yang disebut *odds*. Sedangkan perbandingan nilai *odds* antara dua individu disebut *odds ratio* [6], yang dinotasikan :

$$\begin{aligned} \theta &= \left[\frac{(p_i(1)/(1 - p_i(1)))}{(p_i(0)/(1 - p_i(0)))} \right], \\ &= \frac{(\exp(\beta_0 + \beta_1(1) + \beta_2 X_{2i} + \dots + \beta_k X_{ki}))}{(\exp(\beta_0 + \beta_1(0) + \beta_2 X_{2i} + \dots + \beta_k X_{ki}))}, \\ &= \exp(\beta_1). \end{aligned}$$

Artinya, *odds* (resiko) terjadinya $Y = 1$ pada kategori $X = 1$ adalah sebesar $\exp(\beta_1)$ kali *odds* (resiko) terjadinya $Y = 1$ pada kategori $X=0$. Untuk variabel prediktor kontinu, interpretasinya adalah setiap kenaikan nilai X sebesar satu satuan (unit) akan mengakibatkan perubahan nilai *odds* (resiko) terjadinya $Y = 1$ sebesar $\exp(\beta_1)$ kali.

2.3. Uji Signifikansi Model

Dalam uji signifikansi model dilakukan uji G yaitu uji signifikansi secara bersama dan uji Wald yaitu uji signifikansi secara terpisah.

Uji Signifikansi Bersama

Dalam regresi linier biasa untuk menguji hipotesis tentang pengaruh koefisien regresi β secara bersama digunakan uji F, sementara dalam regresi logistik digunakan uji perbandingan *likelihood* (ratio test) atau disebut juga statistik uji G. Hipotesis pada uji ini yaitu :

$$\begin{aligned} H_0 : \beta_0 = \beta_1 = \dots = \beta_k = 0, \\ H_1 : \beta_j \neq 0 ; j = 1, 2, \dots, k. \end{aligned}$$

Rumus untuk statistik uji G adalah [6]:

$$\begin{aligned} G &= -2 \ln L_0 - (-2 \ln L_k) \\ &= -2 \ln \frac{L_0}{L_k} \end{aligned}$$

dengan

L_k : likelihood model yang terdiri dari k peubah
 L_0 : likelihood model yang hanya terdiri atas β_0 .

Statistik uji G mengikuti distribusi khi-kuadrat (χ^2) dengan derajat bebas k dengan aturan pengambilan keputusan tolak H_0 , yaitu:

$$\begin{aligned} \text{Jika } G > \chi^2_{\alpha, db} \quad \text{maka } H_0 \text{ ditolak,} \\ \text{Jika } G \leq \chi^2_{\alpha, db} \quad \text{maka } H_0 \text{ tidak ditolak.} \end{aligned}$$

Uji Signifikansi Terpisah

Dalam regresi linier biasa untuk menguji hipotesis tentang pengaruh koefisien regresi β secara terpisah digunakan uji t, sementara dalam regresi logistik digunakan statistik uji Wald. Hipotesis pada uji ini yaitu :

$$\begin{aligned} H_0 : \beta_j = 0, \\ H_1 : \beta_j \neq 0 ; j = 0, 1, 2, \dots, k. \end{aligned}$$

Rumus umum untuk uji Wald adalah [6]:

$$W_j = \left(\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \right)^2,$$

dengan

$\hat{\beta}_j$: nilai pendugaan koefisien regresi ke- j
 $SE(\hat{\beta}_j)$: nilai galat baku (*standard error*) pendugaan koefisien regresi ke- j

Statistik uji ini berdistribusi khi-kuadrat (χ^2) dengan derajat bebas 1 dengan aturan pengambilan keputusan tolak H_0 , yaitu:

$$\begin{aligned} \text{Jika } W_j > \chi^2_{\alpha,1} & \text{ maka } H_0 \text{ ditolak,} \\ \text{Jika } W_j \leq \chi^2_{\alpha,1} & \text{ maka } H_0 \text{ tidak ditolak.} \end{aligned}$$

2.4. Uji Keباikan Model

Ukuran kebaikan model pada regresi linier ditentukan dengan menggunakan R^2 , dimana :

$$R^2 = \frac{JK_{total} - JK_{sisaan}}{JK_{total}} = \frac{JK_{regresi}}{JK_{total}}$$

Jumlah kuadrat (JK) sisaan pada regresi linier analog dengan devians pada regresi logistik [5]. Devians merupakan model log *likelihood* dari k variabel dikali-2, yaitu :

$$D = -2 \ln L_k,$$

sedangkan JK total analog dengan $-2 \ln L_0$, sehingga JK regresi pada regresi linier biasa dapat dikatakan analog dengan statistik uji G yaitu :

$$\begin{aligned} G &= -2 \ln L_0 - (-2 \ln L_k), \\ &= -2 \ln\left(\frac{L_0}{L_k}\right). \end{aligned}$$

Dengan demikian, R^2 pada regresi logistik yang analog dengan R^2 pada regresi linier biasa adalah :

$$\begin{aligned} R^2 &= \frac{-2 \ln L_0 - (-2 \ln L_k)}{-2 \ln L_0}, \\ &= 1 - \ln\left(\frac{L_k}{L_0}\right). \end{aligned}$$

Bentuk lain dari R^2 untuk model regresi logistik yaitu :

$$R^2 = 1 - \left(\ln \frac{L_0}{L_k}\right)^{2/n}.$$

Sementara bentuk modifikasi dari R^2 diatas, seperti berikut :

$$\bar{R}^2 = \frac{R^2}{\max R^2},$$

dengan $\max (R^2) = 1 - (\ln L_0)^{2/n}$ [3]. Nilai R^2 merupakan proporsi keragaman data yang dapat diterangkan oleh model. Jika R^2 semakin mendekati 1, maka semakin baik model yang diperoleh.

3. Metode Penelitian

Data yang digunakan dalam penelitian ini merupakan data sekunder. Data berasal dari Profil Kesehatan Indonesia 2011 dan 2012 yang dipublikasikan oleh Kementerian Kesehatan Republik Indonesia [4]. Pendataan yang dilakukan oleh Kementerian Kesehatan Republik Indonesia memiliki perlakuan yang sama dari tahun ketahun.

Kabupaten atau kota yang diambil menjadi sampel adalah wilayah regional 2 Indonesia (Sumatera) kemudian dipilih dengan menyaring data untuk kejadian DBD sehingga didapat 73 kabupaten atau kota sebagai unit sampling.

Variabel data yang digunakan dalam penelitian ini yaitu kejadian DBD pada kabupaten atau kota di wilayah regional 2 Indonesia (Sumatera) sebagai variabel respon (Y), dimana $Y = 0$ jika kabupaten atau kota mengalami kejadian luar biasa DBD, $Y = 1$ jika kabupaten atau kota tidak mengalami kejadian luar biasa DBD dan variabel prediktor (X) meliputi :

- (1) Kepadatan penduduk (X_1)
- (2) Persentase rumah atau bangunan bebas jentik nyamuk Aedes (X_2).
- (3) Persentase rumah sehat (X_3).
- (4) Persentase rumah tangga berperilaku hidup bersih dan sehat atau ber-PHBS (X_4).
- (5) Persentase keluarga menggunakan sumur terlindung (X_5).
- (6) Persentase institusi membina sarana pelayanan kesehatan (X_6).

Adapun langkah-langkah analisis data yang dilakukan dalam penelitian ini adalah sebagai berikut:

- (1) Membuat statistika deskriptif untuk variabel kejadian DBD pada kabupaten atau kota (Y) dan keenam variabel prediktor.
- (2) Untuk mengetahui hubungan faktor-faktor penyebab kejadian DBD dengan menggunakan metode analisis regresi logistik biner dengan tahapan sebagai berikut :
 - (a) Membentuk model dugaan regresi logistik dengan melibatkan seluruh variabel prediktor.
 - (b) Melakukan uji signifikansi keseluruhan model dugaan regresi logistik biner dengan menggunakan uji G.
 - (c) Memilih variabel prediktor yang signifikan berpengaruh terhadap variabel respon dengan uji Wald.
 - (d) Mencari nilai odds ratio untuk masing-masing variabel prediktor yang berpengaruh.
 - (e) Menginterpretasikan model regresi logistik biner.
 - (f) Menentukan keakuratan model.

4. Hasil dan Pembahasan

4.1. Pembentukan Model Terbaik

Dengan memasukkan semua variabel prediktor dan melakukan uji signifikansi model, diperoleh tiga variabel prediktor yang berpengaruh signifikan terhadap kejadian DBD. Setelah dimasukkan tiga variabel prediktor tersebut, diperoleh model dugaan sebagai berikut.

$$\ln \left(\frac{\hat{p}}{1 - \hat{p}} \right) = 3,678 + (-0,033)X_2 + (-0,026)X_4 + (-0,02)X_5$$

Uji Signifikansi Bersama

Tabel 1. Uji Signifikansi Secara Keseluruhan Tahap 2

<u>-2 Log likelihood</u>	<u>Cox & Snell R Square</u>	<u>Nagelkerke R Square</u>
80,501	0,193	0,264

Nilai R^2 – *Cox and Snell* sebesar 0,193 diketahui bahwa proporsi keragaman data yang dapat diterangkan oleh model ini sebesar 19,3%. Parameter determinan regresi logistik yaitu 0,264 sehingga dapat dikatakan kontribusi variabel prediktor terhadap variabel respon adalah 26,4%. Nilai uji G lebih besar dari χ^2 tabel ini berarti bahwa paling sedikit ada satu variabel yang berpengaruh signifikan kejadian DBD.

Uji Signifikansi Terpisah

Tabel 2. Hasil Pengujian Variabel Prediktor Tahap 2 ($\chi^2_{(0,1;1)} = 2,706$)

<u>Variabel</u>	<u>Para Meter</u>	<u>Duga-an</u>	<u>Standar Error</u>	<u>Wald</u>	<u>P -value</u>	<u>Exp(β)</u>	<u>Kesimpulan</u>
<u>Konstanta</u>	β_0	3,678	1,286	8,182	0,004	39,567	-
<u>Angka bebas jentik</u>	β_2	-0,033	0,014	5,746	0,017	0,968	<u>Berpengaruh</u>
<u>ber-PHBS</u>	β_4	-0,026	0,013	3,859	0,049	0,974	<u>Berpengaruh</u>
<u>Sumur Terlindung</u>	β_5	-0,020	0,011	3,410	0,065	0,980	<u>Berpengaruh</u>

Jika nilai Wald lebih besar dari χ^2 tabel maka variabel prediktor berpengaruh terhadap variabel respon dan sebaliknya. Sehingga dapat disimpulkan bahwa variabel angka bebas jentik nyamuk AEDES, ber-PHBS, dan sumur terlindung berpengaruh signifikan terhadap kejadian DBD.

4.2. Interpretasi Koefisien

Untuk menginterpretasikan seberapa besar angka bebas jentik nyamuk AEDES, ber-PHBS, dan sumur terlindung terhadap kejadian DBD digunakan nilai *odds ratio* yang nilainya diperlihatkan pada Tabel 3 berikut :

Tabel 3. Nilai *Odds Ratio* Model Regresi Logistik

Variabel	Rata-rata	Exp(β)
Angka bebas Jentik	-0,033	0,968
ber-PHBS	-0,026	0,974
Sumur Terlindung	-0,020	0,980

Nilai *odds ratio* untuk angka bebas jentik yaitu 0,968 artinya setiap kenaikan 1 persen angka bebas jentik memiliki resiko lebih kecil kejadian DBD sebesar 0,968 kali. Hal ini menginformasikan bahwa semakin berkurang angka bebas jentik nyamuk akan meningkatkan resiko kejadian DBD.

Nilai *odds ratio* untuk ber-PHBS yaitu 0,974 artinya setiap kenaikan 1 persen ber-PHBS memiliki resiko lebih kecil kejadian DBD sebesar 0,974 kali. Hal ini menginformasikan bahwa semakin berkurang rumah tangga ber-PHBS akan meningkatkan resiko kejadian DBD.

Nilai *odds ratio* untuk sumur terlindung yaitu 0,980 artinya setiap kenaikan 1 persen sumur terlindung memiliki resiko lebih kecil untuk kejadian DBD sebesar 0,959 kali. Hal ini menginformasikan bahwa semakin tidak terlindung sumur di lingkungan tempat tinggal akan meningkatkan resiko kejadian DBD.

Berdasarkan nilai koefisien β_0 dikatakan bahwa jika tidak terjadi KLB pada kabupaten atau kota bernilai 0, maka peluang terjadinya KLB adalah :

$$\begin{aligned} \ln\left(\frac{p}{1-p}\right) &= 3,267 \\ &= \frac{1}{1 + \exp(-3,678)} \\ &= 25 \end{aligned}$$

atau sebesar 25 %.

Daftar Pustaka

- [1] Bain, L. J dan Engelhardt, M. 1992. *Introduction to Probability and Mathematical Statistics Second Edition*. Duxbury Press : California.
- [2] Collect, D. 1991. *Modelling Binary Data Second Edition*. Chapman and Hall : UK.
- [3] Cox, D. R and Snell, E. J. 1998. *Analysis of Binary Data Second Edition*. Chapman and Hall : London.
- [4] Kementerian Kesehatan Republik Indonesia. 2013. *Profil Data Kesehatan Republik Indonesia 2012*. Kementerian Kesehatan Republik Indonesia : Jakarta.
- [5] Myers, R.1989. *Classical and Modern Regression with Applications*. PWS Kent : Bonston.
- [6] Nachrowi, D. H dan Usman. 2005. *Penggunaan Teknik Ekonometri*. Raja Grafindo Persada : Jakarta.
- [7] Sucipto, C. D. 2011. *Vektor Penyakit Tropis*. Gosyan Publishing : Jakarta.