

PEMODELAN FAKTOR-FAKTOR YANG MEMPENGARUHI STATUS DAERAH DI INDONESIA MENGUNAKAN METODE REGRESI LOGISTIK BINER

RIZKI FITRAH, YUDIANTRI ASDI,* MAIYASTRI

*Program Studi S1 Matematika,
Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Andalas,
Kampus UNAND Limau Manis Padang, Indonesia.
email : rizkifitrah20@gmail.com, yudiantriasdi@sci.unand.ac.id, maiyastri@sci.unand.ac.id*

Diterima 15 Desember 2020 Direvisi 29 Desember 2020 Dipublikasikan 12 Januari 2021

Abstrak. Status daerah dapat dikategorikan menjadi dua, yaitu daerah tertinggal dan daerah tidak tertinggal. Pada penelitian ini akan dianalisa faktor-faktor yang mempengaruhi status daerah di Indonesia. Analisis dilakukan dengan analisis regresi logistik biner. Data yang digunakan pada penelitian ini adalah data dari Survei Sosial Ekonomi Nasional (SUSENAS) yang dilaksanakan oleh Badan Pusat Statistik (BPS) pada tahun 2018. Penelitian ini menggunakan empat variabel prediktor, yaitu persentase penduduk miskin, angka melek huruf, angka harapan hidup, dan jumlah puskesmas. Langkah awal dalam penelitian adalah membuat statistika deskriptif untuk variabel respon dan keempat variabel prediktor. Kemudian dilakukan analisis Regresi Logistik Biner untuk mengetahui model akhir. Analisis regresi logistik biner dilakukan dengan pengujian signifikansi parameter secara serentak dan parsial. Hasil dari pengujian signifikansi parameter menunjukkan bahwa dari keempat variabel prediktor, variabel persentase penduduk miskin, angka harapan hidup, dan jumlah puskesmas berpengaruh secara signifikan terhadap status daerah di Indonesia.

Kata Kunci: Status daerah, regresi logistik biner

1. Pendahuluan

Status daerah dapat diukur dalam enam hal yaitu ekonomi, sumber daya, infrastruktur, kapasitas keuangan daerah, aksesibilitas, dan karakteristik daerah [4]. Dalam prakteknya, dalam menentukan apakah sebuah daerah termasuk ke dalam daerah tertinggal atau bukan daerah tertinggal, digunakan suatu indeks komposit yang dihitung dengan suatu metode tertentu dari banyak variabel. Sebahagian diantara variabel tersebut tidak dipublikasikan, sehingga menjadi tidak mudah untuk memperkirakan apakah suatu daerah merupakan daerah tertinggal atau bukan daerah tertinggal. Pada penelitian ini, akan digunakan variabel-variabel yang telah

*penulis korespondensi

dipublikasi untuk memperkirakan status kota/kabupaten di Indonesia, termasuk daerah tertinggal atau bukan. Pada penelitian ini akan dibentuk suatu model agar diketahui faktor-faktor apa saja yang mempengaruhi status daerah di Indonesia. Karena status daerah adalah variabel respon yang bersifat biner dengan kategori 0 untuk daerah tidak tertinggal dan kategori 1 untuk daerah tertinggal, maka pada kasus ini salah satu cara untuk memodelkannya adalah dengan menggunakan regresi logistik biner. Model regresi logistik biner pernah diterapkan oleh Zaynita [5] untuk menguji mengenai kasus stroke berdasarkan jenisnya. Kemudian Andriyana [1] menguji faktor-faktor yang mempengaruhi kepemilikan asuransi di provinsi Papua.

2. Landasan Teori

2.1. Regresi Logistik Biner

Regresi logistik biner merupakan metode yang digunakan untuk menganalisa hubungan antara beberapa variabel prediktor dengan suatu variabel respon dimana variabel respon ini bersifat biner yang hanya memiliki dua kemungkinan nilai yaitu sukses atau gagal. Dengan analisis regresi biasa hubungan tersebut dapat dinyatakan dalam model [2]:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \varepsilon_i \quad ; \begin{cases} i = 1, 2, 3, \dots, n \\ Y_i = 0, 1. \end{cases} \quad (2.1)$$

karena Y_i hanya memiliki dua kemungkinan nilai yaitu 0 dan 1 akibatnya ε_i juga memiliki kondisi yang sama, dimana:

$$\varepsilon_i = \begin{cases} 1 - \beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i} - \dots - \beta_k X_{ki} & ; Y_i = 1 \\ -\beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i} - \dots - \beta_k X_{ki} & ; Y_i = 0. \end{cases}$$

Dalam keadaan seperti ini ε_i tidak dapat lagi diasumsikan berdistribusi normal, melainkan berdistribusi menurut distribusi Bernoulli yang memiliki distribusi peluang Y_i sebagai berikut:

Tabel 1. Distribusi Peluang Y_i

Y_i	Peluang($P(Y_i = y_i)$)
0	$q_i = 1 - p_i$
1	p_i

dengan nilai tengah $E(Y_i)$ yaitu:

$$\begin{aligned} E(Y_i) &= 0(q_i) + 1(p_i) \\ &= p_i. \end{aligned} \quad (2.2)$$

Dari Persamaan 2.1 diperoleh nilai tengah $E(Y_i)$, yaitu:

$$\begin{aligned} E(Y_i) &= E(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \varepsilon_i) \\ &= E(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki}) + E(\varepsilon_i). \end{aligned}$$

dari asumsi diketahui $E(\varepsilon_i) = 0$, maka:

$$E(Y_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki}. \quad (2.3)$$

Dengan mensubstitusikan Persamaan 2.2 ke Persamaan 2.3 diperoleh:

$$p_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} \quad ; \quad 0 \leq p_i \leq 1. \quad (2.4)$$

Persamaan 2.4 disebut model peluang linier (*linear probability model*). Model ini menunjukkan bahwa $E(Y_i)$ dapat dinyatakan sebagai peluang sukses p_i , sehingga diperoleh distribusi peluang $\varepsilon_i (\varepsilon_i = Y_i - E(Y_i))$, sebagai berikut:

Tabel 2. Distribusi Peluang ε_i

Y_i	ε_i	Peluang($P(\varepsilon_i = e_i)$)
0	$-p_i$	$q_i = 1 - p_i$
1	$q_i = 1 - p_i$	p_i

Dari Tabel 2 dapat ditentukan:

$$\begin{aligned} E(\varepsilon_i) &= \sum_{i=0}^1 e_i P(\varepsilon_i = e_i) \\ &= -p_i P(\varepsilon_i = -p_i) + (1 - p_i) P(\varepsilon_i = 1 - p_i) \\ &= (-p_i)(1 - p_i) + (1 - p_i)p_i \\ &= 0 \end{aligned}$$

dan

$$\begin{aligned} E(\varepsilon_i^2) &= \sum_{i=0}^1 e_i^2 P(\varepsilon_i = e_i) \\ &= (-p_i)^2 P(\varepsilon_i = -p_i) + (1 - p_i)^2 P(\varepsilon_i = 1 - p_i) \\ &= (-p_i)^2 (1 - p_i) + (1 - p_i)^2 p_i \\ &= p_i^2 q_i + q_i^2 p_i \\ &= p_i q_i (p_i + (1 - p_i)) \\ &= p_i q_i. \end{aligned}$$

dengan nilai ragam galat:

$$\begin{aligned} Var(\varepsilon_i) &= E(\varepsilon_i^2) - [E(\varepsilon_i)]^2 \\ &= p_i q_i - 0 \\ &= p_i q_i. \end{aligned}$$

karena $Var(\varepsilon_i)$ memiliki nilai yang bergantung pada peluang sukses p_i , maka akibatnya asumsi kehomogenan ragam galat ε_i tidak terpenuhi.

Dari uraian di atas dapat diketahui bahwa analisis regresi biasa tidak dapat digunakan untuk memodelkan hubungan antara variabel respon biner dengan beberapa variabel prediktor. Salah satu pendekatan yang dapat digunakan untuk mengatasi masalah ini adalah analisis regresi logistik. Tetapi sebelum melakukan analisis ini perlu dilakukan transformasi logit yang dapat menjamin nilai peluang sukses p_i akan selalu berada dalam selang $[0,1]$.

2.2. Model Regresi Logistik Biner

Model regresi logistik adalah model regresi variabel respon biner yang melibatkan transformasi logit. Model regresi logistik diperoleh dari fungsi logistik dengan persamaan sebagai berikut.

$$p_i = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)}. \quad (2.5)$$

Persamaan (2.5) adalah fungsi logistik dengan satu variabel prediktor. Sedangkan untuk dua atau lebih variabel prediktor maka Persamaan (2.5) dapat diperluas menjadi:

$$p_i = \frac{\exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki})}{1 + \exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki})}. \quad (2.6)$$

Jika $\eta_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki}$ maka Persamaan (2.6) dapat ditulis menjadi:

$$p_i = \frac{1}{1 + e^{-\eta_i}} \quad ; \quad i = 1, 2, \dots, n. \quad (2.7)$$

Persamaan (2.7) adalah bentuk umum dari model logistik atau disebut juga sebagai fungsi logit. Jika $\eta_i \rightarrow -\infty$ maka $p_i \rightarrow 0$ dan jika $\eta_i \rightarrow \infty$ maka $p_i \rightarrow 1$. Dengan demikian, dapat dijamin bahwa nilai p_i akan selalu berada pada selang $[0,1]$.

Kemudian untuk membentuk model regresi logistik maka dilakukan transformasi terlebih dahulu terhadap Persamaan 2.4 untuk mempertahankan struktur linier dari model. Transformasi yang dilakukan adalah transformasi logit, dengan didefinisikan sebagai berikut:

$$\text{Logit}(p_i) = \ln \left(\frac{p_i}{1 - p_i} \right). \quad (2.8)$$

Dari Persamaan (2.7) diperoleh:

$$\begin{aligned} (1 + e^{-\eta_i})p_i &= 1 \\ e^{-\eta_i} &= \frac{1}{p_i} - 1 \\ \frac{1}{e^{\eta_i}} &= \frac{1 - p_i}{p_i} \\ e^{\eta_i} &= \frac{p_i}{1 - p_i}. \end{aligned}$$

sehingga diperoleh:

$$\eta_i = \ln \left(\frac{p_i}{1 - p_i} \right). \quad (2.9)$$

berdasarkan Persamaan (2.8) diperoleh:

$$\eta_i = \text{Logit}(p_i).$$

karena $\eta_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki}$ maka diperoleh model linier sebagai berikut:

$$\text{Logit}(p_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} ; -\infty \leq \text{Logit}(p_i) \leq \infty \quad (2.10)$$

Persamaan (2.10) merupakan hasil transformasi logit dan disebut sebagai model regresi logistik.

2.3. Pendugaan Parameter 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].

Jika x_i dan y_i adalah pasangan variabel prediktor dan respon pada pengamatan ke- i dan diasumsikan bahwa setiap pasangan pengamatan saling bebas dengan pasangan pengamatan lainnya, $i = 1, 2, \dots, n$ maka fungsi peluang untuk setiap pasangan adalah sebagai berikut.

$$f(x_i) = p_i^{y_i} (1 - p_i)^{1-y_i} \quad ; \quad y_i = 0, 1 \quad (2.11)$$

dengan

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

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

$$l(\beta) = \prod_{i=1}^n f(x_i) = \prod_{i=1}^n p_i^{y_i} (1 - p_i)^{1-y_i}. \quad (2.12)$$

Fungsi *likelihood* tersebut lebih mudah dimaksimumkan dalam bentuk $\ln l(\beta)$ dan dinyatakan dengan $L(\beta)$. Dengan melogaritma naturalkan fungsi *likelihood* tersebut diperoleh:

$$L(\beta) = \sum_{i=1}^n \left[y_i \eta_i - \ln(1 + e^{\eta_i}) \right]. \quad (2.13)$$

dengan $\eta_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki}$. Sehingga turunan pertama dari $L(\beta)$ terhadap $(\beta_0, \beta_1, \dots, \beta_k)$ adalah:

$$\frac{\partial L(\beta)}{\partial \beta_k} = \sum_{i=1}^n x_{ji} (y_i - p_i). \quad (2.14)$$

Sedangkan turunan kedua dari $L(\beta)$ adalah:

$$\frac{\partial^2 L(\beta)}{\partial \beta_j^2} = - \sum_{i=1}^n x_{ji}^2 [p_i(1 - p_i)]. \quad (2.15)$$

Pendugaan untuk β pada Persamaan (2.15) tidak dapat diperoleh secara langsung, karena fungsi yang dihasilkan berbentuk implisit sehingga diperlukan metode iterasi Newton Raphson.

2.4. Pengujian Parameter Regresi Logistik Biner Bivariat

Pengujian signifikansi parameter dilakukan untuk mengetahui apakah parameter yang terdapat di dalam model berpengaruh secara signifikan atau tidak. Pengujian signifikansi parameter ini dilakukan secara keseluruhan (serentak) dan secara individu (parsial). Pengujian signifikansi secara serentak dilakukan untuk mengetahui apakah variabel prediktor secara bersama-sama mempengaruhi model atau tidak.

Pengujian secara serentak ini dilakukan menggunakan uji perbandingan *likelihood (ratio test)* atau disebut juga statistik uji G , dengan hipotesis sebagai berikut:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1 : \text{minimal ada satu } \beta_j \neq 0, \text{ dengan } j = 1, 2, \dots, k$$

Rumus untuk uji G ini adalah [3]:

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

dengan : $L_k = \text{likelihood}$ model yang terdiri dari k peubah.

$L_0 = \text{likelihood}$ model yang terdiri atas β_0 (konstanta).

Statistik uji G mengikuti sebaran khi-kuadrat (χ^2) dengan derajat bebas k dan taraf nyata α dengan aturan pengambilan keputusan yaitu tolak H_0 jika nilai $G > \chi^2_{(\alpha, k)}$ dimana nilai $\chi^2_{(\alpha, k)}$ dapat diperoleh dari tabel *Chi-square*.

Pengujian signifikansi parameter secara parsial dilakukan untuk mengetahui apakah variabel prediktor secara individu mempengaruhi model atau tidak. Pengujian secara parsial ini dilakukan menggunakan uji Wald, dengan hipotesis sebagai berikut.

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0 \quad ; \quad j = 0, 1, 2, \dots, k$$

dengan rumus umum untuk uji Wald adalah [3]:

$$W_j = \left(\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \right)^2 \quad ; \quad j = 0, 1, 2, \dots, k.$$

dimana $\hat{\beta}_j$ merupakan nilai pendugaan koefisien regresi ke- j dan $SE(\hat{\beta}_j)$ merupakan nilai galat baku (*standar error*) pendugaan koefisien regresi ke- j . Statistik uji ini berdistribusi khi-kuadrat (χ^2) dengan derajat bebas 1 dan taraf nyata α , dengan aturan pengambilan keputusan yaitu tolak H_0 jika nilai $W_j > \chi^2_{\alpha(1)}$.

2.5. Status Daerah

Daerah tertinggal merupakan suatu daerah kota/kabupaten yang masyarakat dan wilayahnya relatif kurang berkembang dibandingkan daerah lain dalam skala nasional. Terdapat beberapa indikator yang diperkirakan terkait dengan status daerah tertinggal atau bukan daerah tertinggal, beberapa diantaranya adalah:

- (1) Persentase Penduduk Miskin
- (2) Angka Melek Huruf
- (3) Angka Harapan Hidup
- (4) Jumlah Puskesmas

3. Pembahasan

3.1. Pengujian Signifikansi Parameter Regresi Logistik Biner Secara Serentak

Pengujian signifikansi parameter secara serentak dilakukan untuk mengetahui apakah variabel prediktor secara bersama-sama mempengaruhi model atau tidak. Hipotesis signifikansi parameter secara serentak sebagai berikut.

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1 : \text{minimal ada satu } \beta_j \neq 0, \text{ dengan } j = 1, 2, \dots, k$$

Hasil pengujian signifikansi parameter secara serentak bisa dilihat pada Tabel berikut.

Tabel 3. Pengujian Signifikansi Parameter Regresi Logistik Biner Secara Serentak

<i>Chi-square</i>	db	P-value
102,909	4	0,000

H_0 ditolak apabila nilai *Chi-square* lebih besar dari nilai *Chi-square* tabel dimana nilai *Chi-square* tabel dengan taraf nyata 0,05 dan derajat bebas 4 yaitu 9,488. Berdasarkan Tabel 3 dapat dilihat bahwa nilai *Chi-square* sebesar 102,909 dimana nilai tersebut lebih besar dari nilai *Chi-square* tabel. Selain itu diperoleh juga nilai *P-value* sebesar 0,000 dimana nilai ini lebih kecil dari taraf nyata. Hal ini menyatakan bahwa H_0 ditolak, yang artinya minimal ada satu variabel prediktor yang mempengaruhi model.

3.2. Pengujian Signifikansi Parameter Regresi Logistik Biner Secara Parsial

Setelah dilakukan pengujian signifikansi parameter secara serentak, selanjutnya dilakukan pengujian signifikansi parameter secara parsial untuk mengetahui variabel prediktor yang berpengaruh signifikan terhadap model. Hasil pengujian signifikansi parameter secara parsial bisa dilihat pada Tabel berikut.

Berdasarkan Tabel 4 dapat dilihat bahwa variabel angka melek huruf tidak berpengaruh terhadap model. Hal ini ditunjukkan dari nilai Wald yang lebih kecil dari nilai χ^2 tabel.

Tabel 4 juga menunjukkan bahwa variabel persentase penduduk miskin, angka harapan hidup, dan jumlah puskesmas berpengaruh terhadap model. Hal ini ditunjukkan dari nilai Wald yang lebih besar dari nilai χ^2 tabel.

Berdasarkan hasil pengujian signifikansi parameter secara parsial tersebut, diperoleh variabel yang berpengaruh terhadap model yaitu variabel persentase penduduk miskin, angka harapan hidup dan jumlah puskesmas. Model akhir regresi logistik biner yang terbentuk adalah sebagai berikut.

$$\ln \left(\frac{p_i}{1 - p_i} \right) = 28,131 + 0,197X_1 - 0,388X_3 - 0,064X_4$$

Tabel 4. Pengujian Signifikansi Parameter Secara Parsial ($\chi^2_{(0,05;1)} = 3,841$)

Variabel	Parameter	Koefisien	Wald	db	P-value	Exp(β)
Konstanta	β_0	28,131	13,474	1	0,000	1,648
Persentase Penduduk Miskin (X_1)	β_1	0,197	17,221	1	0,000	1,218
Angka Melek Huruf (X_2)	β_2	-0,043	3,595	1	0,058	0,958
Angka Harapan Hidup (X_3)	β_3	-0,388	16,006	1	0,000	0,678
Jumlah Puskesmas (X_4)	β_4	-0,064	4,792	1	0,029	0,938

4. Kesimpulan

Berdasarkan hasil pengujian signifikansi parameter regresi logistik biner, diperoleh variabel prediktor yang berpengaruh terhadap variabel respon adalah persentase penduduk miskin (X_1), angka harapan hidup (X_3), dan jumlah puskesmas (X_4).

5. Ucapan Terima kasih

Terimakasih kepada Ibu Dr. Ferra Yanuar, Bapak Narwen, M.Si dan Bapak Budi Rudianto, M.Si yang telah memberikan masukan dan saran sehingga penelitian ini dapat diselesaikan dengan baik.

Daftar Pustaka

- [1] Andriyana, N.A., 2016, Pemodelan Faktor-Faktor Yang Mempengaruhi Kepemilikan Asuransi Kesehatan di Provinsi Papua, *Skripsi*, tidak diterbitkan, Institut Teknologi Sepuluh Nopember
- [2] Collet, D., 1991, *Modelling Binary Data*, Second Edition, Chapman & Hall, UK.
- [3] Hosmer, DW., Lemeshow, S., 2000, *Applied Logistic Regression*, Edisi ke-2, John Wiley & Sons, Inc, Canada
- [4] Republik Indonesia, 2015, *Peraturan Presiden Republik Indonesia No. 131 Tahun 2015 Tentang Penetapan Daerah Tertinggal Tahun 2015-2019*, Sekretariat Negara, Jakarta.
- [5] Zaynita, A.A., 2017, Pemodelan Kasus Stroke Berdasarkan Jenisnya Menggunakan Analisis Regresi Logistik Biner di Rumah Sakit Umum Haji Surabaya. *Skripsi*, tidak diterbitkan, Institut Teknologi Sepuluh Nopember