

Model Data Kepemilikan Asuransi Kesehatan di Indonesia Berdasarkan Status Pekerjaan Melalui Analisis Regresi Logistik Biner Dua Level

Marsya Anggun Prisila¹, Anna Islamiyati^{1,*}, Andi Kresna Jaya¹

¹Departemen Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Hasanuddin, Indonesia

*Corresponding author: annaislamiyati701@gmail.com

Abstract. Two-level binary logistic regression is a regression analysis method used to analyze the relationship between one response variable in the form of dichotomous qualitative data and several predictor variables, from data with a hierarchical structure. This study aims to obtain a data model of health insurance ownership in Indonesia based on employment status through a two-level binary logistic regression analysis. The method used is a two-level binary logistic regression with a random intercept model using maximum likelihood estimation on health insurance ownership data in Indonesia. Based on the results of the model estimation, it is found that employment status affects health insurance ownership in Indonesia and is 2.99 times more likely to have health insurance than people who do not have jobs.

Keywords: *binary logistics, health insurance, mle, random intercept, two levels.*

1 Pendahuluan

Regresi logistik biner merupakan metode analisis regresi yang digunakan untuk menganalisis hubungan antara satu variabel respon dan beberapa variabel prediktor, dengan variabel respon yang berupa data kualitatif dikotomi yaitu bernilai 0 atau gagal dan bernilai 1 atau sukses [1]. Namun seiring perkembangan zaman, data respon yang bersifat biner (dikotomi) tidak lagi hanya digunakan pada regresi logistik saja, tetapi telah berkembang pada metode regresi lainnya.

Model multilevel merupakan teknik statistik yang telah mengalami pengembangan dari regresi klasik/sederhana. Pengembangan model tersebut didasari karena dalam penelitian di berbagai kasus sering kali dijumpai perbedaan latar belakang pada responden yang diteliti, sehingga data tersebut memiliki struktur yang bertingkat/hirarki dan berkelompok (*cluster*). Goldstein [2] memperkenalkan pengembangan dari regresi biasa untuk mengatasi permasalahan yang disebabkan dari data yang berstruktur hirarki yaitu analisis multilevel. Analisis multilevel dapat digunakan untuk menyelesaikan permasalahan data yang berstruktur hirarki. Apabila variabel respon yang digunakan berupa data biner, maka analisis dilakukan dengan menggunakan regresi logistik biner multilevel [3].

Menurut hasil penelitian Aliyudin dan Budyandra [4] mengenai faktor-faktor yang mempengaruhi komplikasi persalinan Wanita Usia Subur (WUS) di Indonesia yaitu pada tingkat kepercayaan 95% terdapat efek acak yang signifikan pada kejadian komplikasi persalinan WUS di Indonesia. Signifikansi ini memberikan makna bahwa model regresi logistik multilevel lebih baik dibandingkan dengan regresi logistik biasa (satu level) dalam memodelkan data yang ada. Selain itu, diperoleh sebesar 5,1% keragaman dari komplikasi persalinan pada WUS di Indonesia dipengaruhi oleh adanya perbedaan karakteristik antarprovinsi. Dari penelitian tersebut dapat diketahui bahwa model multilevel (dua level) tepat digunakan dalam struktur data yang berjenjang.

Salah satu metode yang digunakan dalam mengestimasi parameter model regresi logistik biner dua level yaitu metode *Maximum Likelihood Estimation* (MLE). Menurut Nurlaila dkk. [5], metode tersebut mengestimasi parameter β dengan cara memaksimalkan fungsi *likelihood* dan mensyaratkan bahwa data harus mengikuti suatu distribusi tertentu. Pada regresi logistik, setiap pengamatan mengikuti distribusi Bernoulli, sehingga dapat ditentukan fungsi *likelihood*-nya [6].

Model dua level digunakan ketika data berbentuk kelompok dan berhirarki yang biasanya ditemui pada kasus di bidang sosial, kesehatan dan pendidikan. Salah satu kasus di bidang kesehatan yaitu kasus kepemilikan asuransi kesehatan. Asuransi kesehatan merupakan salah satu program jaminan pemeliharaan kesehatan yang ditawarkan kepada masyarakat. Tujuan dari asuransi kesehatan yaitu untuk memeralihkan risiko sakit dari tertanggung kepada penanggung, sehingga penanggung berkewajiban memberikan pelayanan (biaya) perawatan kesehatan kepada tertanggung apabila sakit. Sistem utama asuransi kesehatan yaitu untuk melindungi masyarakat dari kesulitan (ekonomi) dalam pembiayaan pelayanan kesehatan [7]. Namun pada beberapa kalangan masyarakat masih belum memiliki asuransi kesehatan. Menurut Pusat Pengelolaan Risiko Fiskal Badan Kebijakan Fiskal, persentase masyarakat yang belum memiliki jaminan kesehatan sebesar 37 % [8].

2 Analisis Regresi Logistik Biner Dua level

Model dua level merupakan sebuah model yang digunakan pada data berjenjang/hirarki. Data berjenjang seringkali ditemukan pada penelitian *survey* yang unit-unit analisisnya berasal dari kelompok-kelompok (*cluster*), atau data yang diambil melalui penarikan contoh bertahap (*cluster sampling*). Misalnya dalam pengambilan sampel menggunakan metode sampling satu tahap (*Single Stage Cluster Sampling*) unit-unit analisis yang berasal dari kelompok (*cluster*) diperhitungkan keberadaanya dalam analisis, sehingga dalam hal ini model yang cocok adalah model dua level. Unit-unit sampling yang ada dalam kelompok disebut unit level 1 dan kelompok-kelompok (*cluster*) disebut unit level 2. Banyaknya unit-unit analisis dalam kelompok bisa sama atau berbeda untuk setiap kelompok [9].

Analisis dua level digunakan untuk mengetahui pengaruh variabel prediktor yang tingkatannya berbeda. Analisis dua level dapat digunakan untuk menyelesaikan permasalahan data yang memiliki struktur berjenjang. Apabila variabel respon yang digunakan berupa data biner, maka analisis dilakukan dengan menggunakan regresi logistik biner dua level [3]. Model dua level dikelompokkan menjadi 2 yaitu model dua level dengan *random intercept* dan model dua level dengan *random slope* [4]. Model dua level dengan *random intercept* merupakan model *intercept* yang dimodelkan sebagai *random effect* dari variabel pada level 2 dengan asumsi bahwa setiap kelompok memiliki *intercept* yang berbeda-beda (tidak *fixed* seperti regresi biasa), namun memiliki kemiringan atau *slope* yang sama sehingga pengaruh setiap variabel prediktor terhadap variabel respon sama untuk tiap-tiap kelompok [4].

Model pada level 1:

$$y_{ij} = \beta_{0j} + \sum_{p=1}^r \beta_p x_{pij} + \varepsilon_{ij} \quad (1)$$

dengan:

$j = 1, 2, \dots, k$ dan $i = 1, 2, \dots, n_j$

y_{ij} : variabel respon pengamatan ke- i dalam unit ke- j

β_{0j} : intersep acak untuk unit ke- j

β_p : efek tetap untuk variabel prediktor ke- p

x_{pij} : variabel prediktor ke- p ($p = 1, 2, \dots, r$) di level 1 untuk pengamatan ke- i dalam unit ke- j

ε_{ij} : galat untuk pengamatan ke- i dalam unit ke- j

Model pada level 2:

$$\beta_{0j} = \beta_0 + u_{0j} \quad (2)$$

dengan:

β_0 : intersep rata-rata keseluruhan pada peubah y

u_{0j} : galat untuk unit ke- j

Persamaan (1) disubstitusikan pada Persamaan (2) sehingga diperoleh model regresi dua level dengan *random intercept* yaitu sebagai berikut:

$$y_{ij} = \beta_0 + \sum_{p=1}^r \beta_p x_{pij} + u_{0j} + \varepsilon_{ij} \quad (3)$$

Pada model intersep acak, notasi $j = 1, 2, \dots, k$ menyatakan unit-unit level 2 dan $i = 1, 2, \dots, n_j$ menyatakan pengamatan-pengamatan pada level 1 yang tersarang dalam unit ke- j pada level 2. Sehingga total observasi level 1 dalam seluruh unit level 2 adalah:

$$n = \sum_{j=1}^k n_j$$

Selanjutnya Persamaan (3) dapat dinyatakan sebagai berikut:

$$\mathbf{y}_{ij} = \boldsymbol{\beta}\mathbf{x}_{ij} + \mathbf{u}_j + \boldsymbol{\varepsilon}_i \quad (4)$$

dengan:

$$\mathbf{x}_{ij} = \begin{bmatrix} 1 & x_{1ij} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

dan u_j adalah error untuk level dua yang independen terhadap ε_i . Diketahui u_j berdistribusi normal dengan rata-rata 0 dan varians σ_u^2 .

Diketahui y_{ij} berdistribusi Bernoulli untuk setiap observasi tunggal, sehingga fungsi kepadatan peluang untuk setiap observasi seperti berikut:

$$f(y_{ij}) = \left(\pi(\mathbf{x}_{ij}) \right)^{y_{ij}} \left(1 - \pi(\mathbf{x}_{ij}) \right)^{1-y_{ij}} \quad (5)$$

Untuk model dengan fungsi penghubung logit model pada Persamaan (4) dapat dituliskan dalam bentuk matriks sebagai berikut:

$$\pi(\mathbf{x}_{ij}) = \boldsymbol{\beta}\mathbf{x}_{ij} + \mathbf{u}_j \quad (6)$$

dengan \mathbf{x}_{ij} adalah vektor variabel prediktor, $\boldsymbol{\beta}$ adalah vektor koefisien regresi dan \mathbf{u}_j adalah vektor error untuk level dua.

Selanjutnya fungsi *likelihood* dari persamaan y pada Persamaan (5) dinyatakan sebagai berikut:

$$\begin{aligned} L &= \prod_{i=1}^n f(y_{ij}) \\ &= \prod_{i=1}^n \pi(\mathbf{x}_{ij})^{y_{ij}} \left(1 - \pi(\mathbf{x}_{ij}) \right)^{1-y_{ij}} \end{aligned}$$

Langkah selanjutnya adalah membuat \ln dari fungsi *likelihood* untuk memaksimumkan fungsi *likelihood* sebagai berikut:

$$\ln L(\boldsymbol{\beta}) = \left(\prod_{j=1}^k \frac{1}{\sqrt{\pi}} w_p \left[\prod_{i=1}^n \left(\mathbf{x}_{ij}\boldsymbol{\beta} + \left(\frac{2\rho}{(1-\rho)} \right)^{\frac{1}{2}} v_p \right)^{y_{ij}} \left(1 - \left(\mathbf{x}_{ij}\boldsymbol{\beta} + \left(\frac{2\rho}{(1-\rho)} \right)^{\frac{1}{2}} v_p \right) \right)^{1-y_{ij}} \right] \right)$$

Selanjutnya memaksimumkan $\ln L(\boldsymbol{\beta})$ dengan menurunkan fungsi \ln dari fungsi *likelihood* terhadap $\boldsymbol{\beta}$, sehingga diperoleh nilai $\boldsymbol{\beta}$ sebagai berikut:

$$\hat{\boldsymbol{\beta}}_{(m+1)} = \hat{\boldsymbol{\beta}}_{(m)} - (\mathbf{H}(\boldsymbol{\beta}))^{-1} \mathbf{g}(\boldsymbol{\beta})$$

Iterasi berhenti jika $\|\hat{\boldsymbol{\beta}}_{(m+1)} - \hat{\boldsymbol{\beta}}_{(m)}\| < \varepsilon$, ε bilangan positif yang sangat kecil (0.001).

Secara runtut, tahapan analisis regresi logistik biner dua level adalah sebagai berikut [1]:

- a. Pengujian Signifikansi *Random Effect (Likelihood Ratio Test)*
- b. Pengujian Signifikansi Parameter secara Simultan
- c. Pengujian Signifikansi Parameter secara Parsial
- d. Pemilihan model terbaik
- e. Interpretasi Parameter
- f. *Interclass Correlation (ICC)*

3 Model Data Kepemilikan Asuransi Kesehatan

Data yang digunakan adalah data hasil dari *Indonesian Family Life Survey* gelombang 5 (IFLS-5). IFLS-5 dilaksanakan oleh *RAND Labor and Population* tahun 2014-2015. Adapun variabel respon (y) adalah kepemilikan asuransi kesehatan yang dikategorikan menjadi dua kategori yaitu tidak memiliki asuransi kesehatan diberi kode '0' dan memiliki asuransi kesehatan diberi kode '1'. Variabel prediktor dalam level 1 adalah tingkat individu untuk status pekerjaan dari responden yang diberi kode '0' untuk yang tidak bekerja dan '1' untuk yang bekerja. Selanjutnya, variabel prediktor dalam level 2 adalah tingkat propinsi yang terdiri atas 23 propinsi di Indonesia, yaitu Nanggroe Aceh Darussalam, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Lampung, Kep. Bangka Belitung, Kep. Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Kalimantan Barat, Kalimantan Tengah, Kalimantan Timur, Kalimantan Selatan, Sulawesi Barat dan Sulawesi Selatan.

3.1 Pengujian Signifikansi *Random Effect*

Uji signifikansi *random effect* dilakukan untuk melihat adanya pengaruh signifikan yang disebabkan oleh efek acak yang dalam hal ini adalah variabel level-2. Berdasarkan *output* STATA, diperoleh p -value sebesar 0.0000 dengan $\alpha = 5\%$ (p -value $< \alpha$) dan nilai *likelihood ratio* sebesar $162.23 > \text{nilai } \chi_{(0.05,1)} = 3.8415$. Dengan demikian, dapat disimpulkan bahwa pada tingkat kepercayaan 95% terdapat efek acak yang signifikan pada kasus kepemilikan asuransi kesehatan di Indonesia. Signifikansi tersebut memberikan makna bahwa model regresi logistik biner dua level lebih baik dibandingkan dengan regresi logistik biner biasa (satu level) dalam memodelkan data.

3.2 Pengujian Parameter secara Simultan

Uji signifikansi parameter secara simultan dilakukan untuk melihat pengaruh secara bersama-sama seluruh variabel prediktor terhadap variabel respon menggunakan uji statistik G^2 (*likelihood ratio test*). Berdasarkan *output* STATA dapat dihitung nilai G^2 :

$$\begin{aligned}
G^2 &= -2 \ln \left(\frac{L_{(null\ model)}}{L_{(conditional\ model)}} \right) \\
&= -2(-6568.4167 + 6513.7831) \\
&= -2(-54.6336) \\
&= 109.2672
\end{aligned}$$

Diperoleh nilai *likelihood ratio* sebesar $109.2672 >$ nilai $\chi_{(0.05,1)} = 3.8415$. Hal itu berarti bahwa pada tingkat kepercayaan 95% terdapat minimal satu variabel prediktor yang mempengaruhi kepemilikan asuransi kesehatan di Indonesia.

3.3 Pengujian Parameter secara Parsial

Uji Wald digunakan untuk melihat pengaruh masing-masing variabel prediktor terhadap variabel respon. Hasil uji Wald untuk level satu adalah sebagai berikut:

Tabel 1 Hasil Pengujian Parameter Koefisien Regresi secara Parsial untuk Satu Level

Variabel	Koefisien	<i>Std. Error</i>	<i>z</i>	<i>P > z </i>
Status bekerja (x_1) Tidak bekerja Bekerja (<i>ref</i>)	1.11	0.12	9.18	0.000
Konstanta	-2.11	0.12	-17.72	0.000

Sumber: Data diolah, 2022

Berdasarkan **Tabel 1**, diperoleh nilai *p-value* kurang dari 0.05 sehingga dapat disimpulkan bahwa status bekerja signifikan mempengaruhi kepemilikan asuransi kesehatan di Indonesia. Dengan demikian, persamaan regresi logistik biner satu level yang terbentuk adalah:

$$y_{ij} = -2.11 + 1.11x_{1ij}$$

3.4 Pembentukan Model Regresi Logistik Biner Dua Level

Membentuk model regresi logistik biner dua level dilakukan setelah memasukkan variabel provinsi. Untuk melihat pengaruh variabel prediktor yang signifikan, dilakukan kembali uji Wald dengan memasukkan variabel level dua yang dalam hal ini adalah variabel provinsi. Hasil uji Wald adalah sebagai berikut:

Tabel 2 Hasil Pengujian Parameter Koefisien Regresi secara Parsial untuk Dua Level

Variabel	Koefisien	Std. Error	z	P > z
Status bekerja (x_1) Tidak bekerja Bekerja (<i>ref</i>)	1.09	0.12	8.98	0.00
Konstanta	-1.90	0.16	-11.53	0.00

Sumber: Data diolah, 2022

Berdasarkan **Tabel 2**, diperoleh nilai $p\text{-value} < 0.05$ sehingga dapat disimpulkan bahwa setelah menambahkan variabel level dua yaitu provinsi, status bekerja signifikan mempengaruhi kepemilikan asuransi kesehatan di Indonesia. Dengan demikian, persamaan regresi logistik biner satu level yang terbentuk adalah:

$$y_{ij} = -1.90 + 1.09x_{1ij}$$

Untuk mengetahui pengaruh masing-masing variabel prediktor terhadap kepemilikan asuransi kesehatan di Indonesia, maka digunakan nilai *odds ratio* seperti berikut.

Tabel 3 Nilai *Odds Ratio*

Variabel	$\hat{\beta}$	$Exp(\hat{\beta})$
Status bekerja (x_1) Tidak bekerja Bekerja (<i>ref</i>)	1.09	2.99

Sumber: Data diolah, 2022

Ket: (*ref*) merupakan kategori acuan

Berdasarkan **Tabel 3**, dapat diketahui bahwa dari hasil uji parsial, status bekerja berpengaruh signifikan terhadap kepemilikan asuransi kesehatan di Indonesia dengan nilai koefisien sebesar 1.09. Hal ini menunjukkan bahwa penduduk yang berstatus bekerja memiliki kecenderungan untuk mempunyai asuransi kesehatan sebesar 2.99 kali dibanding penduduk yang tidak memiliki pekerjaan.

Dilakukan perhitungan nilai *Interclass Correlation* untuk mengukur variasi kepemilikan asuransi kesehatan yang dapat dijelaskan oleh adanya perbedaan karakteristik antarprovinsi.

Tabel 4 Nilai ICC

Level	ICC	Std. Error
Provinsi	0.065	0.025

Sumber: Data diolah, 2022

Nilai ICC pada **Tabel 4**, diperoleh sebesar 0.065 atau 6.5% keragaman kepemilikan asuransi kesehatan di Indonesia dipengaruhi oleh perbedaan karakteristik antarprovinsi. Menurut Sorra dan Dyer (2010) nilai ICC di atas 5% mengimplikasikan bahwa perbedaan

karakteristik kelompok memiliki efek pada respon yang diberikan oleh individu-individu di dalamnya sehingga analisis multilevel diperlukan.

4 Kesimpulan

Model data kepemilikan asuransi kesehatan di Indonesia berdasarkan status pekerjaan melalui analisis regresi logistik biner dua level yaitu sebagai berikut:

$$y_{ij} = -1.90 + 1.09x_{1ij}$$

Berdasarkan model regresi logistik biner dua level, diperoleh hasil bahwa status pekerjaan berpengaruh terhadap data kepemilikan asuransi kesehatan di Indonesia dengan nilai koefisien sebesar **1.09**. Jika mengambil nilai odds ratio dari koefisien regresi, maka diperoleh odds ratio 2.99, artinya bahwa penduduk yang berstatus bekerja memiliki kecenderungan untuk mempunyai asuransi kesehatan sebesar 2.99 kali dibanding penduduk yang tidak memiliki pekerjaan.

5 Daftar Pustaka

- [1] Saragih, M. T., Inayah, A. W., Nooraeni, R., Aprilio, M., Sinsyi, M. M. dan Aprilia, Y. R., 2020, *Penerapan Regresi Logistik Biner Multilevel pada Partisipasi Angkatan Kerja di Provinsi Jawa Tengah Tahun 2018*, Eigen Mathematics Journal, **3**(1), 35-44.
- [2] Goldstein, H., 1995, *Multilevel Statistical Models*, London: Arnold Publisher.
- [3] Yulian, E. dan Pawitan, G., 2017, *Pemodelan Status Usaha (Pengusaha dan Pekerja/Karyawan) Menggunakan Regresi Logistik Multilevel*, Jurnal Matematika "MANTIK", **3**(1), 30-38.
- [4] Aliyudin, F. dan Budyandra, 2016, *Faktor-Faktor yang Memengaruhi Komplikasi Persalinan Wanita Usia Subur di Indonesia Menggunakan Data SDKI 2012 (Aplikasi Analisis Regresi Logistik Biner Multilevel)*, Jurnal Aplikasi Statistika & Komputasi Statistika, **8**(2), 33-46.
- [5] Nurlaila, D., Kusnandar, D. dan Sulistianingsih, E., 2013, *Perbandingan Metode Maximum likelihood Estimation (MLE) & Metode Bayes dalam Pendugaan Parameter Distribusi Eksponensial*, BIMASTER, **2**(1), 51-56.
- [6] Fobia, M. E. dan Setyawan, Y., 2020, *Perbandingan Metode Classification and Regression Tree (CART) dan Metode Regresi Logistik Biner dalam Mengklasifikasi Status Wanita Bekerja di Kota Kupang*, Jurnal Statistika Industri dan Komputasi, **5**(1), 66-74.
- [7] Suryono, A., 2009, *Asuransi Kesehatan Berdasarkan Undang-Undang Nomor 3 Tahun 1992*, Jurnal Dinamika Hukum, **9**(3), 213-221.
- [8] Arimbawa, P. E., 2018, *Hubungan Kepemilikan Asuransi Kesehatan dengan Penggunaan Obat Rasional (POR) pada Pasien Swamedikasi*, Jurnal Ilmiah Medicamento, **4**(2), 118-122.

- [9] Tantular, B. dan Jaya, I. G. N. M., 2008, *Pendekatan Analisis Multilevel Respon Biner dalam Menentukan Faktor-Faktor yang Memengaruhi Imunisasi Lengkap*, Semnas Matematika dan Pendidikan Matematika, Sumedang, Bandung.