

Penaksiran Parameter Model Regresi Probit Multinomial Menggunakan *Maximum Likelihood Estimation*

*Parameters Estimation of Multinomial Probit Regression Model
Using Maximum Likelihood Estimation*

Annisa Ayu Ardita^{a,1}, Asima Manurung^{a,2}

^aProgram Studi Matematika, Universitas Sumatera Utara, Medan, 20155, Indonesia
annisaayuardt@gmail.com¹, asimamanurung73@gmail.com²

ABSTRAK

Regresi probit multinomial (MNP) merupakan analisis regresi yang mengidentifikasi hubungan antara variabel dependen dengan satu atau lebih variabel independen yang sifatnya kategorikal, dimana model yang digunakan adalah model logistik yang link functionnya menggunakan invers dari fungsi kumulatif distribusi normal. Untuk membentuk suatu model MNP dilakukan penaksiran parameter guna mengukur pengaruh relatif dari setiap kategori independen terhadap kategori dependen dalam model MNP. Metode yang digunakan untuk menaksir model MNP adalah maximum likelihood estimation (MLE). Hasil uji dengan mengambil contoh kasus mengenai Maternal Health Risk menunjukkan bahwa model probit multinomial yang parameternya telah ditaksir menggunakan MLE dapat memberikan tingkat akurasi model sebesar 67,88% dimana model untuk Z_1 sebagai log perbandingan antara peluang kecenderungan risiko menengah untuk mengalami maternal mortality terhadap yang mengalami risiko rendah adalah $-7,834 + 0,048x_2 - 0,044x_3 + 0,047x_6$. Model untuk Z_2 sebagai log perbandingan antara peluang kecenderungan risiko tinggi untuk mengalami maternal mortality terhadap yang mengalami risiko rendah adalah $-40,489 + 0,138x_1 + 0,101x_2 + 1,791x_4 + 0,095x_6$.

Kata Kunci : Maximum Likelihood Estimation, Penaksiran Parameter, Regresi Probit Multinomial

ABSTRACT

Multinomial probit regression (MNP) is a regression analysis that identifies the relationship between the dependent variable and one or more independent variables that are categorical, where the model used is a logistic model whose link function uses the inverse of the normal distribution cumulative function. To form an MNP model, parameter estimation is carried out to measure the relative influence of each independent category on the dependent category in the MNP model. The method used to estimate the MNP model is maximum likelihood estimation (MLE). The test results by taking a case example of Maternal Health Risk show that the multinomial probit model whose parameters have been estimated using MLE can provide a model accuracy rate of 67.88% where the model for Z_1 as the log ratio between the chances of medium risk tendency to experience maternal mortality to those experiencing low risk is $-7,834 + 0,048x_2 - 0,044x_3 + 0,047x_6$. The model for Z_2 as the log of the ratio between the odds of high risk tendency to experience maternal mortality to those experiencing low risk is $-40.489 + 0.138x_1 + 0.101x_2 + 1.791x_4 + 0.095x_6$.

Keywords : Maximum Likelihood Estimation, Multinomial Probit Regression, Parameter Estimation

Disubmit: 09 Oktober

Info Artikel :
Direview: 08 November

Diterima : 20 Desember

Copyright © 2023 - Journal UPU. All rights reserved.

1. PENDAHULUAN

Dalam ruang lingkup statistika multivariat, analisis regresi menjadi salah satu metode statistik yang paling umum digunakan untuk memodelkan hubungan antara variabel dependen Y dengan satu atau lebih variabel independen X . Teknik analisis regresi ini pun terbagi atas beberapa faktor, seperti jumlah variabel dependen dan independen, jenis data, dan tujuan analisis itu sendiri. Untuk berbagai kepentingan ataupun tujuan luas, kombinasi dari teknik-teknik regresi ini pun dikembangkan. Misalnya, untuk menganalisis variabel dependen Y yang merupakan variabel kategorikal (nominal ataupun ordinal) dan

variabel independen X yang juga merupakan variabel kategorikal, maka salah satu metode statistik yang bisa digunakan adalah analisis regresi logistik (logit atau probit). Saat variabel dependennya terklasifikasi atas dua kelompok, maka inilah yang disebut dengan analisis regresi logistik multinomial.

Pada kebanyakan penelitian yang dilakukan, regresi logistik yang sering dipakai adalah model logit. Dengan perbedaan yang terletak pada *link function*, dimana logit menggunakan fungsi logistik sedangkan probit menggunakan invers dari fungsi kumulatif distribusi normal, tidak menjadikan dasar bahwa model logit lebih baik daripada model probit ataupun sebaliknya [5]. Dibuktikan dengan penelitian yang dilakukan oleh Fatma Hanani (2019) mengenai “Perbandingan Model Logit dan Model Probit Untuk Mengidentifikasi Faktor-Faktor yang Mempengaruhi Partisipasi Ekonomi Wanita Nelayan Pesisir Selatan Gunungkidul” yang menarik kesimpulan bahwa model probit lebih baik daripada model logit dengan tingkat ketepatan menggunakan model logit sebesar 71,1% dan model probit 71,7 %.

Untuk membentuk suatu model regresi logistik probit multinomial perlu dilakukan penaksiran parameter model untuk mengukur pengaruh relatif dari setiap kategori independen terhadap kategori dependen dalam model regresi probit multinomial (MNP). Hal ini membantu dalam menentukan variabel independen mana yang paling penting mempengaruhi variabel dependen yang berbeda [6]. Beberapa metode yang digunakan dalam penaksiran parameter model regresi logistik adalah *noniterative weighted least squares*, analisis fungsi diskriminan, dan *maximum likelihood estimation* (MLE).

Metode *noniterative weighted least squares* digunakan untuk menangani masalah heteroskedastisitas dan korelasi antar residual dalam model MNP. Namun, metode ini memiliki kelemahan yaitu tidak konsisten dalam menaksir parameter model jika asumsi heteroskedastisitas tidak terpenuhi dan metode ini kurang fleksibel dalam menangani data yang kompleks atau yang tidak memenuhi asumsi distribusi normal multivariabel. Untuk metode analisis fungsi diskriminan, metode ini terlalu sensitif terhadap asumsi normalitas sehingga koefisien yang ditaksir dari data yang berdistribusi bukan normal cenderung *overestimate* [9]. Metode MLE sendiri merupakan metode penaksiran parameter pada model yang telah diketahui distribusinya dengan prosedur memaksimalkan fungsi *likelihoodnya*, yaitu fungsi probabilitas yang menggambarkan kemungkinan terjadinya suatu sampel data berdasarkan nilai parameter yang 3 diperkirakan [1]. MLE adalah salah satu metode penaksiran yang sering digunakan khususnya pada model regresi logistik, dikarenakan pendekatan yang lebih umum dan fleksibel yang memungkinkan menghasilkan nilai taksiran parameter secara konsisten, terlepas dari asumsi tertentu tentang distribusi data.

Penelitian ini akan diterapkan pada contoh kasus *maternal mortality*, yaitu kematian seorang wanita yang terjadi selama masa kehamilan, persalinan, atau dalam 42 hari setelah persalinan atau terminasi kehamilan (terlepas dari penyebab kematian). Data *Maternal Health Risk*, yang diperoleh dari Kaggle, melibatkan enam variabel yang mempengaruhi seorang wanita memiliki risiko *maternal mortality* rendah, sedang, atau tinggi. Enam variabel tersebut adalah *age* (usia), *systolic blood pressure* (angka sistolik tekanan darah), *diastolic blood pressure* (angka diastolik tekanan darah), *blood glucose* (kadar gula dalam darah), *body temperature* (suhu tubuh), dan *heart rate* (detak jantung per menit).

2. METODE

A. Regresi Probit

Regresi probit adalah metode analisis yang digunakan untuk memeriksa hubungan antara variabel dependen yang bersifat kategori (kualitatif) dan variabel-variabel independen yang bersifat kualitatif maupun kuantitatif menggunakan *cumulative distribution function* (CDF) sebagai *link functionnya*. Model regresi probit dapat dituliskan sebagai berikut:

$$\pi_i = (Z_i) = (\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} + \varepsilon_i) \quad (1)$$

dimana $\Phi(\cdot)$ merupakan fungsi kumulatif distribusi normal.

Untuk memperoleh suatu perkiraan dari nilai peluang probit Z_i dilakukan invers $(\cdot)$ sehingga didapatkan:

$$Z_i = \Phi^{-1}(\pi_i) = (\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} + \varepsilon_i) \quad (2)$$

dimana Z_i adalah variabel dependen berdistribusi normal, β_0 adalah parameter intersep yang tidak diketahui, $\beta_i = (\beta_1, \beta_2, \dots, \beta_n)$ adalah koefisien parameter, $X_i = (X_1, X_2, \dots, X_n)$ variabel independen dan ε adalah error yang diasumsikan berdistribusi $N(0,1)$.

B. Distribusi Multinomial

Distribusi multinomial didefinisikan sebagai berikut:

Jika suatu percobaan menghasilkan k hasil $E_1, E_2, \dots, E_k$ dengan peluang $p_1, p_2, \dots, p_k$ maka distribusi peluang peubah acak $X_1, X_2, \dots, X_k$ yang menyatakan banyak terjadinya $E_1, E_2, \dots, E_k$ dalam n percobaan bebas adalah:

$$f(x_1, x_2, \dots, x_k; p_1, p_2, \dots, p_k, n) = \binom{n}{x_1, x_2, \dots, x_k} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}$$

dengan $\sum_{i=1}^k x_i = n$ dan $\sum_{i=1}^k p_i = 1$.

C. Regresi Probit Multinomial

Regresi probit multinomial merupakan analisis regresi untuk mengidentifikasi korelasi antara variabel dependen dan variabel independen. Variabel dependen biasa disimbolkan Y dengan skala pengukuran dikotomis (biner) dan variabel independen biasa disimbolkan X yang skala pengukurannya dapat bersifat dikotomis (biner), polikotomis (multikategori), atau kontinu. Bentuk umum model regresi probit multinomial (MNP) adalah sebagai berikut:

$$\Pr(\beta_i = b_n | Z_i) = \frac{\exp Z_i}{1 + \exp Z_i} \quad (3)$$

$$\pi_i = \frac{\exp Z_i}{1 + \exp Z_i}$$

$$\pi_i(1 + \exp Z_i) = \exp Z_i$$

$$\pi_i + \pi_i \exp Z_i = \exp Z_i$$

$$\pi_i = \exp Z_i - \pi_i \exp Z_i$$

$$\pi_i = [1 - \pi_i] \exp Z_i$$

$$\ln \left[\frac{\pi_i}{1 - \pi_i} \right] = \ln \exp Z_i$$

$$\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} + \varepsilon_i = Z_i \quad (4)$$

D. Maximum Likelihood Estimation

Penaksiran parameter untuk model regresi logistik biasanya menggunakan *metode maximum likelihood estimation*, atau singkatnya MLE [1]. MLE merupakan salah satu metode penaksiran parameter yang dapat digunakan untuk menaksir parameter suatu model yang diketahui distribusinya dengan cara memaksimumkan fungsi *likelihood*-nya. Fungsi *likelihood*-nya adalah:

$$L(\beta) = \prod_{i=1}^n \pi_i^{Y_i} (1 - \pi_i)^{1-Y_i} \quad (5)$$

$$L(\beta) = \prod_{i=1}^n [\Phi(\beta_0 + \beta_i X_i)]^{Y_i} [1 - \Phi(\beta_0 + \beta_i X_i)]^{1-Y_i} \quad (6)$$

sedangkan fungsi *log likelihood*-nya adalah:

$$L(\beta) = \ln \prod_{i=1}^n \pi_i^{Y_i} (1 - \pi_i)^{1-Y_i} \quad (7)$$

$$\ln L(\beta) = \sum_{i=1}^n Y_i \ln \left[\frac{\Phi(\beta_0 + \beta_i X_i)}{1 - \Phi(\beta_0 + \beta_i X_i)} \right] + \sum_{i=1}^n \ln(1 - \Phi(\beta_0 + \beta_i X_i)) \quad (8)$$

Lalu, untuk mendapatkan penaksir β dengan memaksimumkan $\ln$ fungsi likelihood dilakukan dengan cara menurunkan $\ln$ fungsi likelihood terhadap parameter β kemudian menyamakannya dengan nol, seperti berikut:

$$\frac{\partial L(\beta)}{\partial \beta_0} = \sum_{i=1}^n Y_i \left[\frac{\Phi(\beta_0 + \beta_i X_i)}{1 - \Phi(\beta_0 + \beta_i X_i)} \right] = 0 \quad (9)$$

Sedangkan untuk parameter koefisien β_i sebagai berikut:

$$\frac{\partial L(\beta)}{\partial \beta_i} = \sum_{i=1}^n X_i Y_i \left[\frac{\Phi(\beta_0 + \beta_i X_i)}{1 - \Phi(\beta_0 + \beta_i X_i)} \right] = 0 \quad (10)$$

3. HASIL DAN PEMBAHASAN

A. Analisis Regresi Probit Multinomial

Akan dibuktikan bahwa $Z_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{in} + \varepsilon_i$ dimana $Z = \ln \left[\frac{\pi_i}{1 - \pi_i} \right]$. Jika

$$\pi(x_i) = \frac{e^{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_n X_{in}}}{1 + e^{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_n X_{in}}}, \text{ maka:}$$

$$\begin{aligned} 1 - \pi(x_i) &= 1 - \frac{e^{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_n X_{in}}}{1 + e^{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_n X_{in}}} \\ &= \frac{1 + e^{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_n X_{in}}}{1 + e^{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_n X_{in}}} - \frac{e^{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_n X_{in}}}{1 + e^{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_n X_{in}}} \\ &= \frac{1}{1 + e^{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_n X_{in}}} \end{aligned}$$

sehingga diperoleh:

$$\frac{\pi(x_i)}{1 - \pi(x_i)} = \frac{\frac{e^{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_n X_{in}}}{1 + e^{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_n X_{in}}}}{\frac{1}{1 + e^{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_n X_{in}}}} = e^{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_n X_{in}}$$

$$\ln \left[\frac{\pi_i}{1 - \pi_i} \right] = \ln e^{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_n X_{in}}$$

$$\frac{\pi_i}{1 - \pi_i} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_n X_{in} \quad \text{TERBUKTI} \quad (11)$$

B. Maximum Likelihood Estimation

Misalkan terdapat sampel terdiri dari n observasi yang berpasangan (x_i, y_i) , $i = 1, 2, \dots, n$ dengan model regresi probit multinomialnya $\pi(x_i) = \frac{\Phi e^{Z_i}}{1 + \Phi e^{Z_i}}$ dan fungsi kepadatannya:

$$f(y; \pi_i) = \pi_i^{y_i} (1 - \pi_i)^{1-y_i}$$

$$f(y; \pi(x_i)) = \Phi(\pi(x_i))^{y_i} (1 - \Phi(\pi(x_i)))^{1-y_i}$$

dengan $i = 1, 2, \dots, n$, maka fungsi *likelihood*-nya adalah:

$$L(\beta) = \prod_{i=1}^n \pi_i^{Y_i} (1 - \pi_i)^{1-Y_i}$$

$$= \prod_{i=1}^n \Phi(\pi(x_n))^{y_n} (1 - \Phi(\pi(x_n)))^{1-y_n}, \quad i = 1, 2, \dots, n \quad (12)$$

Dari persamaan (12) di atas, selanjutnya akan dicari *log likelihoodnya*.

$$\begin{aligned} L(\beta) &= \ln \prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{1-y_i} \\ &= \ln \prod_{i=1}^n (\Phi(\pi(x_n)))^{y_n} (1 - \Phi(\pi(x_n)))^{1-y_n}, \quad i = 1, 2, \dots, n \\ &= \sum_{i=1}^n \ln(\Phi(\pi(x_i)))^{y_i} + \sum_{i=1}^n \ln(1 - \Phi(\pi(x_i)))^{1-y_i} \\ &= \sum_{i=1}^n (y_i \cdot \ln \Phi(\pi(x_i)) + 1 - y_i \cdot \ln(1 - \Phi(\pi(x_i)))) \\ &= \sum_{i=1}^n (y_i \cdot \ln \Phi(\pi(x_i)) + \ln 1 - \ln \Phi(\pi(x_i)) - y_i \cdot \ln 1 + y_i \cdot \ln \Phi(\pi(x_i))) \\ &= \sum_{i=1}^n (y_i \cdot \ln \Phi(\pi(x_i)) - \ln \Phi(\pi(x_i)) - y_i \cdot \ln 1 + y_i \cdot \ln \Phi(\pi(x_i))) \\ &= \sum_{i=1}^n (y_i \cdot \ln \Phi(\pi(x_i)) - \ln(1 + \Phi(\pi(x_i)))) \\ &= \sum_{i=1}^n (y_i \cdot \Phi(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip}) - \ln(1 + \Phi e^{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip}})) \end{aligned} \quad (13)$$

Untuk memperoleh nilai taksiran β_j yang memaksimumkan (β), persamaan (13) diturunkan terhadap β_j , $j = 0, 1, \dots, p$ seperti berikut:

$$\begin{aligned} L(\beta) &= \sum_{i=1}^n (y_i \cdot \Phi(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip}) - \ln(1 + \Phi e^{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip}})) \\ \frac{\partial L(\beta)}{\partial \beta_0} &= \sum_{i=1}^n (y_i \cdot \Phi(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip}) - \ln(1 + \Phi e^{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip}})) \\ &= \sum_{i=1}^n \partial \beta_0 (y_i \cdot \Phi(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip}) - \partial \beta_0 \ln(1 + \Phi e^{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip}})) \\ &= \sum_{i=1}^n \left[y_i - \left(\frac{e^{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip}}}{1 + e^{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip}}} \right) \right] \end{aligned} \quad (14)$$

Karena $i = 1, 2, \dots, n$ maka diperoleh:

$$\begin{aligned} \frac{\partial L(\beta)}{\partial \beta_j} &= \sum_{i=1}^n (y_i \cdot \Phi(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip}) - \ln(1 + \Phi e^{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip}})) \\ &= \sum_{i=1}^n \partial \beta_j (y_i \cdot \Phi(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip}) - \partial \beta_j \ln(1 + \Phi e^{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip}})) \\ &= \sum_{i=1}^n \left[x_{ij} y_i - \left(\frac{e^{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip}}}{1 + e^{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip}}} \times x_{ij} \right) \right] \end{aligned} \quad (15)$$

C. Contoh Kasus

Berdasarkan data mengenai *Maternal Health Risk*, akan dianalisis bagaimana penaksiran parameter menggunakan MLE dapat menghasilkan pemodelan yang signifikan antara nilai prediksinya dengan data sebenarnya. Dilakukan analisis statistika deskriptif untuk memberikan gambaran umum tentang karakteristik dari masing-masing variabel penelitian. Perhatikan Tabel 1 berikut.

Table 1. Statistika Deskriptif

Variabel Penelitian	Keterangan	N	Rata-Rata (Mean)	Median
Y	Low Risk	195	-	-
	Mid Risk	171		
	High Risk	101		
X1	Age	467	26,98	27,00
X2	Systolic Blood Pressure	467	114,5	120,00

Variabel Penelitian	Keterangan	N	Rata-Rata (Mean)	Median
X3	<i>Diastolic Blood Pressure</i>	467	76,46	75,00
X4	<i>Blood Glucose</i>	467	7,851	7,2
X5	<i>Body Temperature</i>	467	98,46	98,00
X6	<i>Heart Rate</i>	467	74,41	76,00

Berdasarkan informasi pada Tabel 1, dari jumlah data dari tiap kategori variabel dependen Y, tidak ada satu kategori yang mendominasi secara signifikan atau memiliki frekuensi yang sangat jauh lebih besar atau lebih kecil daripada kategori lainnya atau dapat dikatakan proporsinya mendekati seragam.

Lalu sebelum dilakukan analisis lebih lanjut, dilakukan uji independensi untuk mengetahui apakah terdapat hubungan antara variabel independen dengan variabel dependen berdasarkan pengambilan keputusan atas hipotesis H_0 , yaitu tidak terdapat hubungan antara variabel independen X dengan variabel dependen Y . Pengambilan keputusan ini didasarkan pada nilai p -value dengan kriteria jika p -value > 0,05, maka H_0 diterima.

Table 2. Hasil Uji Independensi Variabel Independen

Variabel Independen	p -value	Keputusan
X1	0,000	Terdapat perubahan yang signifikan
X2	0,000	Terdapat perubahan yang signifikan
X3	0,000	Terdapat perubahan yang signifikan
X4	0,000	Terdapat perubahan yang signifikan
X5	0,588	Tidak terdapat perubahan yang signifikan
X6	0,000	Terdapat perubahan yang signifikan

Berdasarkan Tabel 2, diketahui bahwa variabel independen X yang signifikan berpengaruh terhadap variabel dependen Y adalah X1, X2, X3, X4 dan X6. Selanjutnya akan dilakukan penaksiran dan pengujian parameter dari variabel-variabel independen X tersebut.

Hasil dari penaksiran parameter menggunakan metode MLE melalui bantuan *software* RStudio dapat dilihat pada Tabel 3 berikut ini.

Table 3. Koefisien dan Standar Error Parameter Menggunakan MLE dengan Kategori *Low Risk* Sebagai Kategori Pembanding

Risk Level		Coefficient	Standard Error
Mid Risk	(Intercept)	-7,834	2,174
	X1	0,035	0,024
	X2	0,048	0,009
	X3	-0,044	0,011
	X4	0,156	0,177
	X6	0,047	0,016
High Risk	(Intercept)	-40,489	4,730
	X1	0,138	0,046
	X2	0,101	0,019
	X3	0,037	0,020
	X4	1,791	0,276
	X6	0,095	0,031

Setelah diperoleh koefisien dari parameter pada Tabel 3, selanjutnya akan diuji parameter tersebut secara simultan dan parsial. Untuk uji simultan, akan dibentuk sebuah *model reduced* dimana model ini tidak memuat variabel X_4 . Berdasarkan hasil *p-value* untuk uji simultan pada Gambar 1, diperoleh *p-value* sama dengan 0, dimana ini lebih kecil dari nilai signifikansi.

Likelihood ratio tests of Multinomial Models						
Response: Y						
	Model	Resid. df	Resid. Dev	Test	Df LR stat.	Pr(Chi)
1	X1 + X2 + X3 + X6	924	798.8910			
2	X1 + X2 + X3 + X4 + X6	922	668.9812	1 vs 2	2 129.9098	0

Gambar 1. Hasil Likelihood Ratio Test

Sehingga dapat disimpulkan bahwa model *full* lebih baik daripada model *reduced*. Uji parsial dilakukan dengan menghitung uji Wald. Hasil uji Wald dan keputusan mengenai variabel independennya dapat dilihat pada Tabel 4.4.

Table 4. Hasil Uji Wald

Risk Level	Variabel Independen	Coefficient	Standard Error	W	Chi-Square	Keputusan
Mid Risk	X1	0,035	0,024	2,126	3,841	H0 diterima
	X2	0,048	0,009	28,444	3,841	H0 ditolak
	X3	-0,044	0,011	16	3,841	H0 ditolak
	X4	0,156	0,177	0,776	3,841	H0 diterima
	X6	0,047	0,016	8,628	3,841	H0 ditolak
High Risk	X1	0,138	0,046	9	3,841	H0 ditolak
	X2	0,101	0,019	28,257	3,841	H0 ditolak
	X3	0,037	0,020	3,4225	3,841	H0 diterima
	X4	1,791	0,276	42,108	3,841	H0 ditolak
	X6	0,095	0,031	9,391	3,841	H0 ditolak

H_0 ditolak mengindikasikan bahwa variabel independen X berpengaruh secara signifikan terhadap variabel dependen. Dari hasil uji parsial di atas, untuk variabel $Y = 1$, variabel independen yang berpengaruh adalah X_2 , X_3 , dan X_6 . Sedangkan untuk variabel $Y = 2$, variabel independen yang berpengaruh adalah X_1 , X_2 , X_4 , dan X_6 .

Sebelum melakukan pemodelan regresi probit multinomial lebih lanjut, dilakukan uji apakah model yang terbentuk sesuai dengan data (tidak terdapat perbedaan antara hasil pengamatan dan kemungkinan hasil prediksi model). Mengikuti perhitungan dari hasil *likelihood ratio*, maka ditetapkan untuk model *full* memuat seluruh variabel independen (X_1, X_2, X_3, X_4 dan X_6) sedangkan untuk model *reduced* tidak memuat variabel independen X_4 . Berdasarkan perolehan nilai *likelihood* dari kedua model, maka:

$$LRT = -2[LL_{reduced} - LL_{full}]$$

$$LRT = -2[-406.63294 - (-340.17709)]$$

$$LRT = -2[-406.63294 + 340.17709]$$

$$LRT = 132,9117$$

Nilai LRT ini mengikuti distribusi chi-square dengan derajat kebebasannya adalah 1. Perolehan LRT $132,9117 \geq \chi^2_{df;\alpha}$, dimana nilai chi-square sebesar 3,841. Maka dari itu, keputusannya adalah H_0 ditolak yang artinya model telah sesuai.

Dari perolehan penaksiran parameter dan proses pengujian hipotesis yang telah dilakukan, maka dapat dibentuk model regresi probit multinomial terbaik pada contoh kasus risiko *maternal mortality* pada data *Maternal Health Risk* dengan variabel independen yang signifikan. Sehingga diperoleh:

$$Z_1 = -7,834 + 0,048X_2 - 0,044X_3 + 0,047X_6$$

dan

$$Z_2 = -40,489 + 0,138X_1 + 0,101X_2 + 1,791X_4 + 0,095X_6$$

Dimana Z_1 merupakan log perbandingan antara peluang kecenderungan mengalami risiko menengah untuk mengalami *maternal mortality* terhadap yang mengalami risiko rendah dan Z_2 merupakan log perbandingan antara peluang kecenderungan mengalami risiko tinggi untuk mengalami *maternal mortality* terhadap yang mengalami risiko rendah.

Selanjutnya akan dilihat seberapa akurat model probit yang dihasilkan dengan melihat perolehan nilai untuk tabel konfusi pada Gambar 2.

Confusion Matrix and Statistics			
	0	1	2
0	153	85	7
1	39	84	14
2	3	2	80
Overall Statistics			
Accuracy : 0.6788			
95% CI : (0.6343, 0.721)			
No Information Rate : 0.4176			
P-Value [Acc > NIR] : < 2.2e-16			
Kappa : 0.4935			
McNemar's Test P-Value : 4.271e-06			
Statistics by Class:			
	Class: 0	Class: 1	Class: 2
Sensitivity	0.7846	0.4912	0.7921
Specificity	0.6618	0.8209	0.9863
Pos Pred Value	0.6245	0.6131	0.9412
Neg Pred Value	0.8108	0.7364	0.9450
Prevalence	0.4176	0.3662	0.2163
Detection Rate	0.3276	0.1799	0.1713
Detection Prevalence	0.5246	0.2934	0.1820
Balanced Accuracy	0.7232	0.6561	0.8892

Gambar 2. Hasil Tabel Konfusi

Diperoleh jumlah data variabel dependen Y yang tepat diklasifikasikan ke tiap kategori dari model regresi probit multinomial yang terbentuk. Untuk *true positive* ada sebanyak 153 data, *true negative* ada sebanyak 164 data. Untuk *false positive* ada sebanyak 106 data, dan *false negative* ada sebanyak 44 data. Dengan demikian perhitungan nilai APER dengan E sebanyak 150 adalah:

$$APER = \frac{FP + FN + 2(N - (TP + TN))}{TP + TN + FP + FN}$$

$$APER = \frac{106 + 44 + 2(150)}{153 + 164 + 106 + 44} = \frac{450}{467} \approx 0,964$$

Dari *overall statistics* yang tertera juga pada Gambar 2, disimpulkan bahwa tingkat keakurasian model regresi probit multinomial ini adalah sebesar 0,6788 atau 67,88%.

4. KESIMPULAN

Persamaan model regresi probit multinomial $Z_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} + \varepsilon_i$ ditaksir menggunakan metode *maximum likelihood estimation* dengan persamaan *likelihoodnya*:

$$\sum_{i=1}^n \left[y_i - \left(\frac{e^{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_n X_{np}}}{1 + e^{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_n X_{np}}} \right) \right] = 0$$

$$\sum_{i=1}^n \left[x_{ij} y_i - \left(\frac{e^{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_n X_{np}}}{1 + e^{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_n X_{np}}} \times x_{ij} \right) \right] = 0$$

Penaksiran parameter dari model regresi probit multinomial menggunakan metode *maximum likelihood estimation* menghasilkan dua model untuk data *Maternal Health Risk*, yaitu:

$$Z_1 = -7,834 + 0,048X_2 - 0,044X_3 + 0,047X_6$$

$$Z_2 = -40,489 + 0,138X_1 + 0,101X_2 + 1,791X_4 + 0,095X_6$$

dimana tingkat keakurasian dari model yang dihasilkan sebesar 67,88%.

Pada penelitian ini terbatas pada cara menaksir parameter menggunakan metode *maximum likelihood*. Agar pengembangan literatur mengenai penaksiran parameter lebih luas lagi, penelitian selanjutnya dapat dilakukan dengan menggunakan metode yang berbeda, seperti *penalized maximum likelihood*.

UCAPAN TERIMA KASIH

Puji syukur saya ucapkan kepada Allah SWT karena atas berkas rahmat dan izin-Nya saya bisa melakukan penelitian ini hingga selesai. Shalawat dan salam kepada Nabi Muhammad SAW yang telah memberikan syafaat dan pedoman serta menjadi teladan bagi umatnya. Terima kasih saya ucapkan kepada keluarga besar saya dan Ibu Asima Manurung, M.Si selaku dosen pembimbing dan penasihat akademik. Tidak lupa juga kepada jurusan tercinta saya, Program Studi Sarjana Matematika FMIPA Universitas Sumatera Utara, atas kesempatan yang telah diberikan untuk mempublikasikan penelitian ini. Serta seluruh teman-teman yang mendukung saya selama masa pengerjaan penelitian ini.

REFERENSI

- [1] Aldrich, J. H., & Nelson, F. D. (1984). *Linear Probability, Logit, and Probit Models Series: Quantitative Applications in the Social Sciences*. Iowa: Sage Publications, Inc.
- [2] Altman, E. I. (1968). Financial Ratios, Discriminant Analysis and The Prediction Of Corporate Bankruptcy. *The Journal of Finance*, 589–609.
- [3] Draper, N. R., & Smith, H. (1992). *Analisis Regresi Terapan* (2nd ed.). (I. Sumantri, Trans.) Jakarta, Indonesia: PT Gramedia Pustaka Utama.
- [4] Eni, N., & Handayani, L. (2020). Regresi Probit Untuk Analisis Variabel-Variabel yang Mempengaruhi Perceraian di Sulawesi Tengah. *Jurnal Aplikasi Statistika dan Komputasi Statistik*.
- [5] Gujarati, D. (2006). *Basic Econometric*, 2nd Edition. New York: McGraw-Hill.
- [6] Hair Jr, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate Data Analysis Eight Edition*. Cengage Learning EMEA.
- [7] Hanani, F. (2019). Perbandingan Model Logit dan Model Probit Untuk Mengidentifikasi Faktor-Faktor yang Mempengaruhi Partisipasi Ekonomi Wanita Nelayan Pesisir Selatan Gunungkidul.
- [8] Hilbe, J. M. (2009). *Logistic Regression Models*. Tempe, USA: CRC Press.
- [9] Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression (Second Edition)*. New York: A Wiley-Interscience Publication.
- [10] Revou. (n.d.). *Statistika Deskriptif*. Retrieved from Revou.co: <https://revou.co/kosakata/statistika-deskriptif>
- [11] Rohman, M. A. (n.d.). Model Probit: Kelebihan dan Kekurangannya. Retrieved from Sekolah Stata: <https://sekolahstata.com/model-probit-kelebihan-dankekurangannya/>
- [12] Strang, Gilbert. (1988). *Linear Algebra and Its Applications*. Academic Press. Styaningsih, F. (2020). Perbandingan Regresi Logistik Ordinal Model Logit Dan Model Probit Pada Pengaruh Faktor Ibu Terhadap Berat Badan Lahir Rendah (Bblr) Di Indonesia Tahun 2017. Surabaya. Retrieved from <https://repository.unair.ac.id/95775/>
- [13] Styaningsih, F. (2020). Perbandingan Regresi Logistik Ordinal Model Logit Dan Model Probit Pada Pengaruh Faktor Ibu Terhadap Berat Badan Lahir Rendah (Bblr) Di Indonesia Tahun 2017. Surabaya. Retrieved from <https://repository.unair.ac.id/95775/>
- [14] Zahroh, Z. Z., & Zain, I. (2018). Analisis Regresi Logistik Multinomial Pada FaktorFaktor Yang Mempengaruhi Sumber Air Bersih Rumah Tangga di Jawa Timur. *Jurnal Sains dan Seni ITS*, 7(2).
- [15] Zmijewski, M. E. (1984). Methodological Issues Related to the Estimation of Financial Distress Prediction Models. *Journal of Accounting Research*, 59–82.