MATHunesa
Jurnal Ilmiah Matematika *Volume 8 No. 2 Tahun 2020*
ISSN 2716-506X

ESTIMASI PARAMETER MODEL REGRESI PROBIT MULTINOMIAL DENGAN METODE *MAXIMUM LIKELIHOOD*

Ika Nurwanitantya Wardani
Jurusan Matematika, FMIPA, Universitas Negeri Surabaya
e-mail : Ikaw16030214004@mhs.unesa.ac.id

A'yunin Sofro
Jurusan Matematika, FMIPA, Universitas Negeri Surabaya
e-mail: Ayuninsofro@unesa.ac.id

Abstrak

Regresi Probit Multinomial merupakan model non linier yang digunakan untuk menganalisis hubungan antar satu variabel respon dengan beberapa variabel bebas. Dimana, variabel respon nya berupa data biner kualitatif yaitu bernilai 0 dan 1. Penelitian ini membahas tentang model regresi Probit Multinomial untuk mengetahui faktor yang memengaruhi jumlah pendapatan terhadap pilihan memancing seseorang. Dalam mengestimasi parameter digunakan metode *Maximum Likelihood Estimation*. Setelah itu dilanjutkan dengan uji serentak menggunakan *Likelihood Ratio test,* dan uji parsial menggunakan uji Wald. Berdasar pada model Probit Multinomial yang diperoleh, diketahui variabel-variabel yang mempengaruhi secara signifikan jumlah pendapatan adalah jumlah memancing di kapal pribadi. Bahwa setiap penambahan 1 rasio jumlah memancing di kapal pribadi akan berpengaruh terhadap pendapatan.
Kata kunci: Regresi Probit Multinomial, *Maximum Likelihood Estimation* (MLE), *Likelihood Ratio test,* uji Wald

Abstract

Multinomial Probit Regression is a non-linear model used to analyze the relationship between one dependent variable with several independent variables where the dependent variable is a qualitative dichotomous data that is worth 0 and 1. This study examines the Multimomial Probit regression model to determine factors that affect the amount of income a person's fishing choices. In estimating parameters the Maximum Likelihood Estimation (MLE) method is used. Then proceed with the simultaneous test using the Likelihood Ratio test and a partial test using the Wald test. Based on the Multinomial Probit model obtained, it is known that the variables that significantly affect the amount of income are the amount of fishing on a private boat. That every addition of 1 ratio of the number of fishing on pobadi vessels will affect the income.
Keywords : Multinomial Probit Regression, Maximum Likelihood Estimation (MLE), Likelihood Ratio test, Wald test

1. PENDAHULUAN

Pada prinsipnya, analisis regresi merupakan suatu metode atau teknik yang digunakan untuk membentuk/membangun suatu persamaan yang menghubungkan antara variabel respon (Y) dengan satu atau lebih variabel prediktor (X). Sedangkan peubah itu sendiri dapat dibedakan atas dua macam yaitu peubah kuantitatif dan peubah kualitataif. Dengan demikian, berbagai kemungkinan kombinasi yang bisa terjadi, misalnya peubah bebas dan terikat keduanya kuantitatif, keduanya kualitatif atau salah satunya kualitatif.

Regresi dapat digunakan pada semua jenis data. Apabila data pada variabel respon adalah kategori, model regresi yang mampu menyelesaikannya adalah model logistik (logit) dan probit. Regresi logistik dan probit termasuk dalam *Generalized Linear Model* (GLM) di mana yang membedakan keduanya adalah adalah fungsi *link*. Regresi Logistik menggunakan fungsi logistik atau logit, sedangkan regresi probit menggunakan distribusi normal standar. Tetapi perbedaan ini tidaklah menjadikan model logit lebih baik daripada model probit atau sebaliknya (Gujarati, 2004).

Regresi probit yang merupakan singkatan dari *Probability Unit* berdasarkan fungsi sebaran peluang normal kumulatif baku dikenal sebagai model Normit atau *Normal Probability Unit* yakni suatu model regresi yang berhubungan dengan unit-unit probabilitas. Pada model regresi probit, variabel respon (Y) merupakan variabel *dummy* yang berdistribusi Bernoulli, yang memiliki arti yaitu 1 jika sukses dan 0 jika gagal. (Gujarat, 20016a:48). Model regresi probit multinomial merupakan metode analisis yang digunakan untuk menggambarkan hubungan antara variabel prediktor dan variabel respon lebih dari dua kategori. Adapun metode yang digunakan dalam pendugaan parameter model regresi probit multinomial yaitu *Maximum Likelihood Estimation* (MLE).

2. KAJIAN TEORI

Distribusi Keluarga Eksponensial

Sebuah konsep penting yang mendasari *Generalized Linear Model* (GLM) adalah distribusi keluarga eksponensial. Anggota keluarga distribusi eksponensial semuanya memiliki fungsi kepadatan probabilitas untuk respon yang teramati y yang dapat ditulis dalam bentuk berikut:

$$f(y;\theta,\phi) = exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right\} \quad (2.1)$$

dimana:

$a(.), b(.), c(.)$: fungsi spesifik

θ : natural parameter lokasi

ϕ : parameter disperse

Fungsi $a(\phi)$ umumnya berupa $a(\phi) = \phi.\omega$, dimana ω adalah konstanta yang diketahui (myers et al, 2010). Distribusi Binomial, Multinomial, Poisson, Gamma, Beta, dan Normal adalah anggota keluarga eksponensial. Dalam penelitian ini hanya akan membahas tentang distribusi multinomial. Karena distribusi keluarga eksponensial merupakan konsep yang mendasari *Generalized Linear Model* (GLM), maka *subsection* berikutnya akan menjelaskan tentang *Generalized Linear Model* (GLM).

Generalized Linier Model (GLM)

Generalized Linear Model (GLM) merupakan bentuk umum dari Model Linear. GLM juga merupakan suatu prosedur yang menyediakan analisis regresi dan analisis varian untuk satu variabel respon dengan satu atau lebih variabel prediktor yang bisa juga disebut dengan kovariat. GLM memungkinkan peneliti untuk meneliti interaksi antar respon dan juga pengaruh respon masing-masing, pengaruh kovariat, dan interaksi kovariat dengan respon. Misalkan, diketahui vektor y memiliki *n* komponen, yang berarti penerapan dari suatu matriks respon Y. Setiap komponen nya bebas dan berdistribusi dengan mean atau $\boldsymbol{E(Y)} = \boldsymbol{\lambda}$. Jika model yang terbentuk memiliki prediktor x, dengan beberapa parameter yang tidak diketahui $\boldsymbol{\beta_i}, \dots, \boldsymbol{\beta_n}$, maka modelnya adalah berupa kombinasi linear $\boldsymbol{\lambda} = \sum_{i=1}^{p} \boldsymbol{x_i \beta_i}$. Sebagai transisi/perubahan dari model Linear terhadap GLM, maka dijabarkan bentuk melalui tiga buah komponen, yaitu *Random Component*, berarti nilai pengamatan respon Y yang saling independen dan berdistribusi tertentu; *Systematic Component*, berarti kombinasi linear dari variabel x dengan parameter $\boldsymbol{\beta}$; *Link between random and Systematic/link function*, berarti suatu fungsi yang menjelaskan nilai harapan dari variabel respon (Y) yang menghubungkan dengan beberapa variabel prediktor melalui persamaan linier (Lindsey, 1997). Ada banyak kemungkinan pilihan *link function.* Jika dipilih:

$$\eta_i = \theta_i$$

maka η_i adalah *link* kanonik. *Link function* dari formulasi ini menghasilkan beberapa sifat teoritis yang akan ditunjukkan pada tabel 1 (Nielsen & Garcia, 2011).

Tabel 1. menujukkan *link* kanonik untuk pilihan distribusi yang paling umum digunakan pada GLM.

Tabel 1. *Link* Kanonik pada *Generalized Linear Model* (GLM)

DISTRIBUSI	*LINK* KANONIK
Normal	$\eta_i = \mu_i$
Binomial	$\eta_i = \ln\left(\frac{\mu_i}{1-\mu_i}\right)$
Multinomial	$\eta_i = \ln\left(\frac{\mu_i}{\mu_k}\right), i < k$
Poisson	$\eta_i = \ln(\mu_i)$
Eksponensial	$\eta_i = \frac{1}{\mu_i}$
Gamma	$\eta_i = \frac{1}{\mu_i}$

Ada *link function* lainnya yang dapat digunakan dengan GLM adalah (Nelder, 1983):

Link probit $\eta = \Phi^{-1}(\mu)$, dimana $\Phi(.)$ adalah fungsi distribusi kumulatif Normal. Pada *subsection* akan dijelaskan tentang distribusi Multinomial.

Distribusi Multinomial

Apabila n percobaan berulang dapat menghasilkan lebih dari 2 *outcome* yang mungkin dengan probabilitas masing-masing konstan pada setiap percobaan, akan dihasilkan distribusi multinomial. Distribusi peluang multinomial digunakan sebagai penentuan peluang hasil yang di kategorikan ke dalam lebih dari dua kelompok. Bila suatu usaha tertentu dapat menghasilkan *m* macam hasil $E_1, E_2, \dots E_m$ dengan peluang $p_1, p_2, \dots, p_3$, maka distribusi peluang peubah acak $X_1, X_2, \dots X_k$ yang menyatakan banyaknya kejadian $E_1, E_2, \dots E_k$ dalam *n* usaha bebas adalah:

$$P(x_1, x_2, \dots, x_m) = \frac{n!}{x_1!\, x_2! \dots x_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k} \quad (2)$$

dengan $\sum_{i=1}^{k} x_i = n$ dan $\sum_{i=1}^{k} p_i = 1$. Metode yang akan digunakan dalam penelitian ini adalah metode regresi probit multinomial yang akan dijelaskan pada *subsection* berikut ini.

Regresi Probit

Regresi probit merupakan analisis yang digunakan untuk mengetahui hubungan antar variabel respon yang bersifat kualitatif dan variabel-variabel bebas yang bersifat kualitatif maupun kuantitatif. Model regresi probit merupakan model non linier yang digunakan untuk menganalisis hubungan variabel respon dengan beberapa variabel prediktor, dimana variabel responnya berupa data kualitataif biner yaitu bernilai 0 dan 1. Metode regresi probit menggunakan fungsi kumulatif distribusi Normal untuk menjelaskan fungsi persamaannya. Karena variabel respon/dependen bersifat dikotomi/biner, maka variabel respon/dependen (y) mengikuti distribusi Binomial dengan fungsi peluang sebagai berikut:

$$f(y_i;\pi_i) = \pi_i^{y_i}(1-\pi_i)^{1-y_i},\ \ y_i = 0;1 \qquad (3)$$

dimana π_i = peluang kejadian ke-i untuk y = 1

$1-\pi_i$ = peluang kejadian ke-i untuk y = 0

Jika nilai Y adalah variabel acak Binomial dengan nilai mean $\mu = n\pi$ dan standar deviasi $\sigma = \sqrt{n\pi(1-\pi)}$, maka nilai Z untuk distribusi normal adalah:

$$Z = \frac{Y - n\pi}{\sqrt{n\pi(1-\pi)}}$$

dimana variabel *random* dari *standard normal distribution* dinotasikan dengan Z. *Standard normal distribution* memiliki nilai mean $\mu = 0$ dan standar deviasi $\sigma = 1$. Fungsi kepadatan peluang dari *standard normal distribution* variabel acak kontinu Z yaitu:

$$f_N(z;0,1) = P(Z \le z) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

Fungsi transformasi pada model regresi probit adalah fungsi kumulatif distribusi normal sebagai fungsi *link* pada GLM.

$$P(y_i = 1|x_i) = \boldsymbol{\Phi}(\boldsymbol{x_i'\beta}) = \int_{-\infty}^{x_i'\boldsymbol{\beta}} \phi(z)dz \qquad (4)$$

dimana $\boldsymbol{\Phi}(.)$ adalah fungsi kumulatif distribusi Normal dan $\phi(.)$ adalah fungsi kepadatan peluang distribusi normal. Dengan kata lain, untuk variabel respon/dependennya bersifat dikotomi/biner.

$$\Phi\big(g(x)\big) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{g(x)} e^{-\frac{z^2}{2}} dz$$

Model regresi probit secara umum dapat dinyatakan sebagai berikut:

$$\pi_i = \boldsymbol{\Phi}(Z_i) = \boldsymbol{\Phi}(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i)$$

dimana x_{ij} adalah variabel bebas/prediktor/independen yang bersifat stokastik. Adapun model peluang probit dapat dituliskan sebagai berikut:

$$Z_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i$$

Hal ini disebabkan karena model peluang probit berhubungan dengan fungsi kumulatif distribusi Normal. Sedangkan untuk memperoleh suatu perkiraan dari nilai peluang probit Z_i, maka dapat digunakan invers dari fungsi kumulatif ditribusi Normal sehingga dapat diperoleh:

$$Z_i = \boldsymbol{\Phi}^{-1}(\pi_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i$$

Model regresi probit adalah GLM dengan komponen acak binomial yang merupakan keluarga eksponensial dengan fungsi *link* model Φ^{-1}, mengubah peluang untuk *z-score* dari *standard normal distribution*. Kemudian akan dijelaskan estimasi parameter $\boldsymbol{\beta}$ pada regresi probit dengan menggunakan metode *Maximum Likelihood Estimation* (MLE) yang diasumsikan mengikuti distribusi binomial.

Estimasi Parameter Model Regresi Probit

Penaksiran paremeter pada model regresi probit menggunakan metode MLE. Metode MLE merupakan suatu metode yang dapat digunakan untuk menaksir parameter suatu model yang diketahui distribusinya. Penaksiran parameter regresi probit dengan metode MLE diawali dengan membuat fungsi *likelihood* nya. Fungsi *likelihood* untuk model regresi probit dapat ditulis:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} f(y_i;\pi_i) \qquad (5)$$

Fungsi transformasi pada model regresi probit yaitu fungsi kumulatif distribusi Normal sehingga persamaan (2.6) menjadi:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} [\Phi(\boldsymbol{x_i'\beta})]^{y_i}[1-\Phi(\boldsymbol{x_i'\beta})]^{1-y_i}$$

dimana $\Phi(.)$ adalah fungsi kumulatif distribusi Normal.

$$\Phi(\boldsymbol{x_i'\beta}) = \int_{-\infty}^{\boldsymbol{x_i'\beta}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

sehingga *log likelihood* nya diperoleh sebagai berikut:

$$\ell(\boldsymbol{\beta}) = \log L(\boldsymbol{\beta}) =$$

$$\sum_{i=1}^{n} y_i \log\big(\Phi(\boldsymbol{x_i'\beta})\big) + \sum_{i=1}^{n} (1-y_i) log[1-\Phi(x_i'\beta)] \qquad (6)$$

Regresi Multinomial

Misalkan variabel respon $\boldsymbol{w} = (w_1, \dots, w_n)'$ mengambil nilai $1, \dots, C$. Untuk mengestimasi probabilitas kelas kondisional $\Pr(w = i|x)$ untuk $i = 1, \dots, C$, didefinisikan C variabel laten $y_1, \dots, y_c$ sebagai

$$y_j = f_j(x) + e_{ij}, \qquad j = 1, \dots, C, \qquad (7)$$

dimana f_j adalah fungsi regresi C dan error $e = (e_1, \dots, e_c)' \sim N(0, \Sigma)$. Untuk $t = 1, \dots, n$, misalkan $y_{.t} = (y_{1t}, \dots, y_{Ct})'$ menjadi vektor dari variabel laten berhubungan pada observasi diskrit w_t ke t. Vektor laten $y_{.t}$ dihubungkan pada observasi diskrit w_t dengan mengharuskan jika $w_t = j$ maka $y_{jt} > y_{it}$ untuk semua

$i \neq j$. Model (2.7) tidak di identifikasi tanpa batasan lebih lanjut karena hanya mengamati w_t dan bukan $y_{.t}$. sehingga, jika $w_t = j$, maka diketahui bahwa $y_{jt} - y_{it} > 0$ untuk $i \neq j$, namun tidak diketahui variabel laten sebenarnya. Untuk mencapai identifikasi, diasumsikan tanpa kehilangan generalitas bahwa fungsi regresi $f_C(x)$ identic dengan 0 dan $\Sigma = I_C$. Asumsi kedua tidak dibutuhkan untuk identifikasi (Bunch 1991; keane 1992). Tetapi secara perhitungan yang tepat, saat mengestimasi kelas probabilitas. Ketika error e merupakan Gaussian, model (2.7) disebit model probit multinomial. Meskipun model (2.7) relatif tidak dikenal oleh ahli statistik, McFadden (1973) menunjukkan bahwa ketika error e_j, $j = 1, \dots, C$ independen dan didistribusikan secara identik dengan distribusi Gumbel, maka (2.7) sesuai dengan model regresi logistik yang digunakan secara luas dalam statistik. Oleh karena itu, diasumsikan bahwa matriks kovarian $\Sigma = I_C$ tidak terlalu ketat karena di ketahui bahwa probabilitas kelas yang sama diperoleh untuk regresi probit multinomial dengan $e \sim N(0,1)$ untuk regresi logistik (Yau, Kohn, & Wood, n.d., 2014).

Model Regresi Probit Multinomial

Regresi probit multinomial merupakan suatu analisis regresi untuk mengidentifikasi hubungan antara variabel respon dan variabel bebas. Variabel respon biasa disimbolkan Y dengan skala pengukuran dikotomus, dan variabel bebas biasa disimbolkan X yang skala pengukurannya bersifat dikotomus, polikotomus atau kontinu. Model probit multinomial sering menggunakan kerangka kerja variabel laten (Imbens, 2007).

$$U_i = \begin{pmatrix} U_{i0} \\ U_{i1} \\ \vdots \\ U_{iJ} \end{pmatrix} = \begin{pmatrix} X'_{i0}\beta + \epsilon_{i0} \\ X'_{i1}\beta + \epsilon_{i1} \\ \vdots \\ X'_{iJ}\beta + \epsilon_{iJ} \end{pmatrix}$$

$$\epsilon_i = \begin{pmatrix} \epsilon_{i0} \\ \epsilon_{i1} \\ \vdots \\ \epsilon_{iJ} \end{pmatrix} | X_i \sim \mathrm{N}(0, \Omega),$$

Untuk beberapa kovarians yang relatif tidak terbatas $(J+1) \times (J+1)$ matriks Ω (di luar normalisasi)

John F. Geweke, Michael P. Keane (1994) dan Hajivasilliou dan Mc-Fadden (1994) mengusulkan cara menghitung probabilitas dalam model probit multinomial yang memungkinkan peneliti untuk berurusan dengan set pilihan yang jauh lebih besar. Upaya sederhana untuk memperkirakan probabilitas adalah dengan menggambar i dari distribusi normal multivariat dan menghitung probabilitas pilihan j sebagai jumlah kali pilihan j sesuai dengan utilitas tertinggi.

Bagaimanapun, pendekatan probit multinomial yang tidak terbatas merupakan pilihan yang masuk akal terhadap jumlah parameter lebih dari satu: semua elemen $(J+1) \times (J+1)$ dalam dimensi Ω minus beberapa termasuk dalam batas normalisasi dan simetri.

Kita dapat memodelkan fungsi marginal untuk level individu yang bervariasi sebagai berikut:

$$U_{ij} = X'_{ij}\beta_i + \epsilon_{ij} \text{ atau } U_{ij} = X'_{ij}\bar{\beta} + v_{ij} \qquad (8)$$

Dimana $v_{ij} = \epsilon_{ij} + X_{ij} \cdot (\beta_i - \bar{\beta})$ yang tidak lagi merupakan pilihan independen.

Satu kemungkinan untuk mengimplementasikan metode ini adalah dengan mengasumsikan keberadaan dari sejumlah jenis individu, mirip dengan model campuran yang digunakan oleh J.Heckman and B. Singer (2016) dalam durasi pengaturan:

$$\beta_i \in \{b_0, b_1, \dots, , b_K\},$$

dengan $\Pr(\beta_i = b_k | Z_i) = p_k$ atau$\Pr(\beta_i = b_k | Z_i) = \frac{\exp(\mathrm{Z}'_i \gamma k)}{1 + \sum_{l=1}^{K} \exp(\mathrm{Z}'_i \gamma l)}$. Secara alternatif dapat ditulis:

$$\beta_i | Z_i \sim N(\mathrm{Z}'_i \gamma, \Sigma) \qquad (9)$$

dimana menggunakan campuran distribusi normal (Imbens, 2007).

Maximum Likelihood Estimation

Metode *Maximum Likelihood Estimation* (MLE) merupakan suatu metode penaksiran parameter yang digunakan untuk menaksir parameter pada model yang diketahui distribusinya. Metode MLE juga merupakan metode yang memaksimumkan fungsi *i*. Langkah-langkah yang digunakan pada metode metode maksimum likelihood, sebagai berikut:

a. Menentukan sampel acak sebanyak *n*.
b. Membentuk/membangun fungsi *likelihood* dari *n* sampel acak, karena pada umumnya metode *maximum likelihood* memberikan perkiraan β dengan memaksimalkan su atu fungsi *likelihood*. Fungsi *likelihood*, seperti di bawah ini:

$$l(\beta) = \prod_{i=1}^{n} [\pi_0(x_i)^{y_{0i}} \pi_1(x_i)^{y_{1i}} \dots \pi_n(x_i)^{y_{ni}}$$

c. Melakukan operasi perhitungan berbentuk ln, berfungsi untuk memaksimalkan fungsi *likelihood*

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

$$= ln \left[\prod_{i=1}^{n} [\pi_0(x_i)^{y_{0i}} \pi_1(x_i)^{y_{1i}} \pi_2(x_i)^{y_{2i}}] \right]$$

$$= \sum_{i=1}^{n} ln\pi_0(x_i)^{y_{0i}} + ln\pi_1(x_i)^{y_{1i}} + ln\pi_2\,(x_i)^{y_{2i}}$$

$$= \sum_{i=1}^{n} y_{1i} g_1(x_i) + y_{2i} g_2(x_i) - \ln[1 + \exp\left(g_1(x_i)\right) + \exp\left(g_1(x_i)\right)$$

d. Mendapatkan nilai β yang memaksimumkan nilai ln dengan melakukan diferensiasi $L(\beta)$, sebagai berikut:

$$\frac{\partial L}{\partial \beta} = 0, \text{dan } \frac{\partial^2 L}{\partial^2 \beta} < 0$$

Dalam melakukan uji hipotesis untuk parameter pada regresi probit multinomial terdapat dua uji sebagai berikut:

1) Uji serentak

Uji ini dilakukan guna mengetahui signifikansi koefisien β terhadap variabel respon secara serentak, dengan hipotesis uji sebagai berikut:

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

(artinya tidak ada satupun peubah bebas berpengaruh terhadap Y)

H_1 : paling sedikit satu $\beta_i \neq 0, i = 1,2, \dots, k$

(artinya paling tidak ada satu peubah yang berpengaruh terhadap respon Y)

Statistik uji G atau *likelihood ratio test* yang digunakan sebagai berikut:

$$G = -2\ln\left[\frac{\left(\frac{n_1}{n}\right)^{n_1}\left(\frac{n_2}{n}\right)^{n_2}\left(\frac{n_3}{n}\right)^{n_3}}{\prod_{i=1}^{n}[\pi_0(x_i)^{y_{0i}}\pi_1(x_i)^{y_{1i}}\pi_2(x_i)^{y_{2i}}]}\right] \quad (10)$$

dimana,

$n_1 = \sum_{i=1}^{n} y_{1i}$, $n_2 = \sum_{i=1}^{n} y_{2i}$, $n_3 = \sum_{i=1}^{n} y_{3i}$, dan $n = n_1 + n_2 + n_3$

keterangan:

n_1 : banyaknya nilai observasi Y=1

n_2 : banyaknya nilai observasi Y=2

n_3 : banyaknya nilai observasi Y=3

Statistik uji G menggunakan distribusi *chi-square* sehingga tolak H_0 jika $G > X^2_{(n,\alpha)}$, dengan n adalah banyaknya variabel prediktor atau banyaknya parameter dalam model pada taraf signifikan α sebesar 0,05 atau *p-value* < α.

2) Uji Parsial

Uji parsial dilakukan untuk mengetahui pengaruh koefisien β secara individual atau masing-masing, dengan pengujian hipotesis sebagai berikut.

$H_0 : \beta_i = 0, i = 0, 1,2, \dots, k$

H_1 : paling sedikit satu $\beta_i \neq 0, i = 0, 1,2, \dots, k$

Statistik uji yang digunakan adalah uji Wald sebagai berikut:

$$W = \frac{\hat{\beta}_j}{SE\left(\hat{\hat{\beta}}_j\right)} \quad (11)$$

Dimana $\hat{\beta}_j$ adalah penaksir dari β_j dan $SE(\hat{\beta}_j)$ adalah penaksir galat baku β_j, W adalah nilau uji Wald mengikuti sebaran X^2 dengan derajat bebas satu. Uji wald mengikuti *standard normal distribution* dengan hasil tolak H_0 jika *p-value* < α atau nilai $W > X^2_{(n,\alpha)}$, dimana α adalah tingkat signifikansi sebesar 0,05. Sehingga dapat disimpulkan bahwa variabel independen berpengaruh terhadap variabel dependen.

3. METODE

Penelitian ilmiah ini menggunakan jenis penelitian terapan, yaitu penelitian yang bertujuan untuk menerapkan, menguji, serta mengevaluasi kemampuan suatu teori yang diterapkan pada pemecahan masalah secara efisien. Adapun metode kepustakaan yang dilakukan pada penelitian ini yaitu dengan mengkaji teori pendukung yang berkaitan, yaitu meliputi teori metode estimasi, multikolinieritas, analisis regresi, model regresi probit, dan regresi multinomial untuk memperoleh estimator dari parameter pada model regresi probit multinomial. Adapun variabel dependen (respon) yang digunakan pada penelitian ini adalah pantai, dermaga, privat (pribadi), dan kapal sewa. Variabel respon adalah variabel yang menjadi akibat karena adanya variabel independen/bebas. Sedangkan variabel independen (bebas) yang digunakan pada penelitian ini adalah pendapatan. Variabel independen/bebas merupakan variabel yang mempengaruhi atau yang menjadi sebab perubahan munculnya variabel dependen/respon.

Tahap pertama pada penelitian ini yaitu pengumpulan teori-teori atau informasi yang diperukan dan sekaligus mempelajarinya, seperti pengumpulan literatur, diskusi, dan pemahaman topik. Setelah itu, menyusun asumsi dan batasan masalah untuk topik yang akan dibahas agar ruang lingkup permasalahan lebih jelas. Kemudian, dapat dilakukan pemodelan regresi Probit Multinomial dengan tahapan yaitu menentukan variabel dependen dan variabel independen hingga mengestimasi parameternya. Kemudian dilanjutkan dengan menguji faktor-faktor yang memengaruhi pendapatan terhadap pilihan memancing seseorang berdasarkan nilai parameter yang signifikan dengan metode uji serentak dan uji parsial. Sehingga pada akhirnya model dapat diperoleh.

4. PEMBAHASAN

Regresi Probit Multinomial

Model regresi probit multinomial adalah sebagai berikut:

$$\Pr(\beta_i = b_k | Z_i) = \frac{\exp(Z_i' \gamma k)}{1 + \sum_{l=1}^{K} \exp(Z_i' \gamma l)}$$

Model regresi probit multinomial dapat ditransformasikan untuk mempermudah perhitungan. Bentuk transformasi dari model regresi probit multinomial sebagai berikut:

$$\pi_i = \frac{\exp(Z_i' \gamma k)}{1 + \sum_{l=1}^{K} \exp(Z_i' \gamma l)}$$

$$(\pi_i)(1 + \exp(Z_i' \gamma k)) = \exp(Z_i' \gamma k)$$

$$(\pi_i + \pi_i \exp(Z_i' \gamma k)) = \exp(Z_i' \gamma k)$$
$$\pi_i = \exp(Z_i' \gamma k) - \pi_i \exp(Z_i' \gamma k)$$
$$\pi_i = [1 - \pi_i]\exp(Z_i' \gamma k)$$
$$\frac{\pi_i}{1 - \pi_i} = \exp(Z_i' \gamma k)$$
$$\ln\left(\frac{\pi_i}{1 - \pi_i}\right) = \ln(\exp(Z_i' \gamma k))$$
$$\ln\left(\frac{\pi_i}{1 - \pi_i}\right) = Z_i' \gamma k$$

dimana,

$$Z_i = \mathbf{\Phi}^{-1}(\pi_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i$$

Maximum Likelihood Estimation (MLE)

Penentuan parameter MLE dengan menggunakan software R, dalam melakukan uji parameter untuk memperoleh *p-value* dengan MLE. Adapun hasil yang diperoleh disajikan dalam Tabel 2. berikut:

Tabel 2. Hasil Estimasi Parameter Model Regresi Probit Multinomial

Variabel	Estimasi	Standar Eror	Z_{value}	*p-value*
Intercept₁	-1,694	0,224	-7,562	$3{,}952\text{x}10^{-14}$
Intercept₂	-0,916	0,207	-4,421	$9{,}805\text{x}10^{-06}$
Intercept₃	-1,167	0,159	-7,338	$2{,}169\text{x}10^{-13}$
Intercept₄	0,527	0,178	2,965	0,0030
Pendapatan; pantai	0,033	0,050	0,661	05084
Pendapatan; dermaga	-0,094	0,050	-1,883	0,0596
Pendapatan; pribadi	0,123	0,029	4,287	$1{,}813\text{x}10^{-05}$
Pendapatan; sewa	-0,043	0,039	-1,084	0,2782

Setelah memperoleh estimasi parameternya, langkah selanjutnya yaitu melakukan uji parameter dengan serentak untuk mengidentifikasi adakah pengaruh dari variabel bebas pada variabel respon dengan uji hipotesis yaitu:

$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4$

H_1 : minimal ada satu $\beta_i \neq 0$, i = 1,2,3,4

Nilai G pada analisis ini adalah 1473,2 dan derajat kebebasan = 4, sedangkan $X^2{}_{(4)(0.05)} = 49{,}077$. Karena nilai $G > X^2{}_{(4)(0{,}05)}$ maka menolak H_0, yang berarti minimal ada satu variabel bebas yang berpengaruh pada variabel respon.

Selanjutnya, yang dilakukan yaitu uji parsial untuk mengidentifikasi pengaruh masing-masing variabel BEBAS tersebut. Berdasarkan Tabel 2. menunjukkan bahwa hasil uji adalah menolak H_0 yang berarti parameter β_0 adalah signifikan karena nilai *p-value* lebih kecil dari α = 0,05. Pengujian pada Parameter β_1 diperoleh hasil uji gagal menolak H_0, bahwa nilai *p-value* yaitu 0,5084 lebih besar dari α = 0,05, yang berarti variabel prediktor $X_{1.1}$ tidak berpengaruh terhadap variabel respon (pantai). Pengujian pada Parameter β_2 diperoleh hasil uji gagal menolak H_0, bahwa nilai *p-value* yaitu 0,0596 lebih besar dari α = 0,05, yang berarti variabel prediktor $X_{1.2}$ tidak berpengaruh terhadap variabel respon (dermaga). Pengujian pada Parameter β_3 diperoleh hasil uji menolak H_0, bahwa nilai *p-value* sebesar $1{,}813\text{x}10^{-05}$ lebih kecil dari α = 0,05, yang berarti variabel prediktor $X_{1.3}$ berpengaruh terhadap variabel respon (pribadi). Pengujian pada Parameter β_4 diperoleh hasil uji gagal menolak H_0, bahwa nilai *p-value* yaitu 0,2782 lebih besar dari α = 0,05, yang berarti variabel prediktor $X_{1.4}$ tidak berpengaruh terhadap variabel respon (sewa).

Sehingga, diperoleh model dari Regresi Probit Multinomial sebagai berikut:

$$Z_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i$$

dimana:

$\beta_{0(1)}$ = -1,694, $\beta_{0(2)}$ = -0,916, $\beta_{0(3)}$ = -1,167

$\beta_{0(4)} = 0{,}527$ $\beta_3 = 0.123$

Berdasarkan model tersebut, maka dapat di interprestasikan bahwa dalam setiap penambahan 1% variabel $X_{1.3}$ yang signifikan akan melipat gandakan variabel respon Y dengan rata-rata sebesar $\exp(0.123)$ = 1.130. Dengan maksud yaitu, penambahan 1 rasio jumlah memancing di kapal pribadi akan berpengaruh terhadap pendapatan yang diperoleh.

5. KESIMPULAN

Estimasi parameter dari model regresi probit multinomial menggunakan metode MLE. Adapun hasil yang telah diperoleh pada bab sebelumnya, model regresi probit multinimoal yang terbentuk yaitu:

$$Z_i = -1{,}694 - 0{,}916 - 1{,}167 + 0{,}527 + 0.123X_3$$

Berdasarkan model probit multinomial di atas, maka diperoleh hasil variabel yang berpengaruh signifikan terhadap pendapatan adalah jumlah memancing di kapal pribadi. Bahwa setiap penambahan penambahan 1 rasio jumlah memancing di kapal pribadi akan berpengaruh terhadap pendapatan.

DAFTAR PUSTAKA

Gujarati, DN. (2004), *Basic Econometric*, 4th edition, McGraw-Hill, New York.

Gujarati, DN. (2006b), *Basic Econometric*, 2nd edition, McGraw-Hill, New York.

Hajivassiliou and Mc.Fadden, D. 1994. A. A method of simuated scores for the estimation of LDV models with application to external debt croses. Cowles Foundation

Imbens, G. (2007). *What is New in Econometrics Lecture 11: Discrete Choi ce Models*.

John F. Geweke, Michael P. Keane, D. E. R. (1994). *Statistical Inference iin the Multinomial Multiperiod Probit Model*. Federal Reserve Bank of Minneapolis.

Lindsey, J. K. (1997). *Applying Generalized Linear Models*. New York: Springer.

Nelder, P. M. and J. A. (1983). *Generalized Linear Models* (2nd ed.). London, New York.

Nielsen, F., & Garcia, V. (2011). *Statistical exponential families : A digest with flash cards ∗ A digest of exponential families*. *2011*.

Singer, J. H. and B. (2016). *A Method for Minimizing Impact of Distribution Assumptions in Econometric Models for Duration Data*. *52*(2), 271–320.

Yau, P., Kohn, R., & Wood, S. (n.d.). *Journal of Computational and Bayesian Variable Selection and Model Averaging in High-Dimensional Multinomial Nonparametric Regression* .