

MODEL REGRESI PROBIT BIVARIAT

NURFIDAH DWITIYANTI

Program Studi Teknik Informatika

Fakultas Teknik, Matematika dan Ilmu Pengetahuan Alam

Universitas Indraprasta PGRI

Jl. Nangka No. 58 C, Tanjung Barat, Jagakarsa, Jakarta Selatan 12530

Email: nurfidah.pulungan@gmail.com

Abstrak. Model regresi probit bivariat merupakan model regresi yang digunakan untuk menganalisis hubungan antara dua buah variabel respon dengan satu atau lebih variabel prediktor, dengan variabel respon berupa data kualitatif biner. Variabel galat pada model ini diasumsikan berdistribusi normal bivariat. Metode kemungkinan maksimum dengan bantuan iterasi Newton Raphson digunakan untuk mendapatkan taksiran parameter regresi probit bivariat. Selanjutnya, dilakukan pengujian signifikansi pada model dengan menggunakan uji perbandingan kemungkinan untuk menguji signifikansi parameter secara simultan dan uji Wald untuk menguji signifikansi parameter secara parsial. Uji Lagrange Multiplier digunakan untuk menguji ada atau tidaknya korelasi galat antara masing-masing variabel respon. Model probit bivariat diterapkan pada kasus kepercayaan seseorang terhadap layanan internet yang dikaitkan dengan penggunaan internet beserta faktor-faktor yang mempengaruhi keduanya. Berdasarkan hasil analisis yang dilakukan, diperoleh bahwa nilai peluang untuk kepercayaan seorang responden terhadap layanan internet dan penggunaan internet secara simultan mempunyai pengaruh terhadap keduanya dengan faktor-faktor yang mempengaruhinya adalah pendidikan, pendapatan, usia, dan jenis kelamin. Namun, dari hasil uji Wald diperoleh bahwa faktor pendidikan dan usia berpengaruh secara signifikan terhadap kepercayaan seseorang dalam layanan internet. Sedangkan untuk penggunaan internet, faktor yang berpengaruh adalah pendidikan dan pendapatan.

Kata Kunci: model probit bivariat, distribusi normal bivariat, metode kemungkinan maksimum

Abstract. Bivariate probit regression model is used to analyze the relationship between two binary response variables with one or more predictor variables. An error variable on this model is assumed a bivariate normal distribution. Maximum likelihood method using Newton Raphson iteration is used to estimate the regression coefficients. Furthermore, the likelihood ratio test is used for testing the significance of parameters simultaneously and the Wald test is used for testing the significance of parameter partially. The Lagrange Multiplier test is used to test a correlation between error of the both response variables respectively. This bivariate probit model is applied for the case study of people's trust for internet service and internet usage belong the influence factors. Based on the results, the probability of respondents who trust the internet services and the internet usage simultaneously have influence both of them with the influence factors namely education, income, age, and gender. However, based on the Wald test is obtained that the factors of education and age significantly influence the people's trust for the internet service. Whereas, for the internet usage, the influential factors are education and income.

Keywords: bivariate probit model, bivariate normal distribution, maximum likelihood method

PENDAHULUAN

Secara terminologis, regresi mengandung arti menuju rata-rata. Artinya, teknik regresi sesungguhnya berfokus pada pengestimasi distribusi data sampel untuk memprediksi hubungan kasualitas antarvariabel yang diharapkan terjadi ketika data sampel variabel-variabel yang diuji tersebut terdistribusi mendekati rata-rata (Abdillah & Hartono, 2015). Pada analisis regresi, variabel tersebut dibedakan kedalam dua jenis variabel yaitu variabel respon Y dan variabel prediktor X . Pada umumnya, analisis regresi sering menggunakan data kuantitatif sebagai variabel responnya. Akan tetapi dalam kenyataannya banyak ditemukan kasus dengan variabel responnya berupa data kualitatif yang berbentuk biner, misalnya jenis kelamin, dan pengambilan keputusan ya atau tidak. Dalam analisis regresi, hubungan antara variabel respon bersifat kualitatif sedangkan variabel prediktornya bisa bersifat kuantitatif, kualitatif, atau gabungan keduanya yang dapat digambarkan kedalam suatu model yang dikenal sebagai model respon biner. Sebagai ilustrasi, model respon biner dapat ditemukan dalam kasus partisipasi angkatan wanita dewasa pada angkatan kerja sebagai fungsi dari rata-rata tingkat upah, pendapatan suami, umur, banyak anak usia sekolah, dst. Telah diketahui bahwa seseorang dapat termasuk kedalam angkatan kerja ataupun tidak. Oleh karena itu, variabel responnya yaitu partisipasi angkatan kerja wanita dewasa dapat diasumsikan bernilai 1 jika termasuk kedalam angkatan kerja dan 0 jika tidak termasuk kedalam angkatan kerja (Nachrowi & Usman, 2002). Selanjutnya, dari model ini dapat diamati hubungan antara pernah atau tidaknya wanita tersebut menjadi angkatan kerja beserta faktor-faktor yang mempengaruhinya.

Model respon biner ini terdiri atas dua model, yaitu model logit dan model probit. Untuk mengatasi masalah seperti ilustrasi di atas, model probit merupakan salah satu model yang dapat digunakan selain model logit yang telah banyak diterapkan pada berbagai penelitian. Pada umumnya analisis model probit yang diterapkan hanya menggunakan satu variabel respon. Pada kasus tertentu terdapat penelitian yang memerlukan dua variabel respon, sehingga diperlukan analisis yang dapat menyelesaikan permasalahan tersebut (Trisna Pradnyantari, N, Gde Sukarsa, I dan Putu Suciptawati, N, 2015). Pengembangan model probit untuk dua buah variabel respon, model ini biasa dikenal dengan model probit bivariat. Pada model probit bivariat, variabel responnya terdiri atas dua buah variabel dikotomi sedangkan untuk variabel prediktornya dapat berupa data kuantitatif atau data kualitatif.

Selanjutnya akan dibahas mengenai model probit bivariat untuk penerapan pada kasus tentang faktor-faktor yang berpengaruh terhadap kepercayaan seseorang (dalam hal ini responden) dengan layanan internet yang dikaitkan dengan penggunaan internet pada 150 responden yang merupakan warga Amerika Serikat. Penulis mengambil data General Social Survey of NORC di Amerika Serikat. Regresi probit bivariat adalah suatu analisis regresi yang digunakan untuk menggambarkan hubungan antara dua buah variabel respon dengan satu atau lebih variabel prediktor, dengan variabel respon Y berupa data kualitatif biner, sedangkan variabel prediktor dapat berupa data kuantitatif, data kualitatif, atau gabungan keduanya yang disimbolkan dengan X . Pada model regresi probit bivariat, variabel galatnya ε diasumsikan saling bebas, berdistribusi identik dan mengikuti distribusi normal bivariat dengan korelasi ρ . Bentuk persamaan umum dari model regresi probit bivariat adalah:

$$Y_{i1}^* = \beta'_{i1} X_{i1} + \varepsilon_{i1}, i = 1, 2, 3, \dots, n \quad (1)$$

dengan

$$Y_{i1}^* = \begin{cases} 1, & Y_{i1}^* > 0 \\ 0, & \text{untuk yang lain} \end{cases}$$

$$Y_{i2}^* = \beta'_{i2} X_{i2} + \varepsilon_{i2}, i = 1, 2, 3, \dots, n \quad (2)$$

dengan

$$Y_{i2}^* = \begin{cases} 1, & Y_{i2}^* > 0 \\ 0, & \text{untuk yang lain} \end{cases}$$

dimana Y_{i1}^* dan Y_{i2}^* menyatakan variabel respon yang berukuran $n \times 1$. β'_{i1} dan β'_{i2} menyatakan vektor koefisien dari variabel bebas yang berukuran $p \times 1$. X_{i1} dan X_{i2} menyatakan vektor variabel bebas. Pada model regresi probit bivariat terdapat dua persamaan model regresi probit dengan masing-masing galat yang diasumsikan berdistribusi normal bivariat, sehingga perlu dilakukan pengujian untuk mengetahui apakah kedua galat pada masing-masing variabel responnya secara signifikan saling berkorelasi atau tidak.

Adapun langkah-langkah pengujian untuk mengetahui ada tidaknya korelasi antara galat masing-masing model dengan menggunakan uji Lagrange Multiplier adalah:

1. Perumusan Hipotesis

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

2. Besaran yang diperlukan

$$\text{Menghitung } g = \sum_{i=1}^n q_{i1} q_{i2} \frac{\phi(w_{i1})\phi(w_{i2})}{\phi(w_{i1})\phi(w_{i2})} \text{ dan } h = \sum_{i=1}^n \frac{[\phi(w_{i1})\phi(w_{i2})]^2}{\phi(w_{i1})\phi(-w_{i1})\phi(w_{i2})\phi(-w_{i2})}$$

Dengan bantuan Program Stata versi 10.

3. Statistik Uji

$$LM = \frac{g^2}{h}$$

4. Kriteria Pengujian

Dengan mengambil taraf signifikansi α , maka H_0 ditolak jika $p\text{-value} < \alpha$.

5. Kesimpulan

Penafsiran H_0 ditolak memberi arti bahwa korelasi antara galat masing-masing model adalah tidak sama dengan nol atau dengan kata lain bahwa kedua model persamaan secara signifikan saling berkorelasi satu sama lain.

Sedangkan metode yang digunakan untuk menaksir parameter yang tidak diketahui pada regresi probit bivariat adalah metode kemungkinan maksimum. Pada pengujian parameter model probit bivariat dilakukan uji secara simultan dan parsial. Uji hipotesis bagi koefisien regresi secara simultan dilakukan dengan maksud untuk mengetahui apakah variabel-variabel prediktor yang digunakan dalam model secara bersama-sama mempunyai pengaruh terhadap variabel yang ingin dijelaskan atau tidak.

Langkah-langkah yang perlu dilakukan dalam pengujian signifikansi parameter secara simultan dengan menggunakan uji perbandingan kemungkinan sebagai berikut:

1. Perumusan Hipotesis

$$H_0 : \beta_{i1} = \dots = \beta_{jp} = 0, \text{ untuk } j = 1, 2$$

$$H_1 : \text{sekurang-kurangnya terdapat satu } \beta_{jk} \neq 0, \text{ untuk } j = 1, 2, k = 1, 2, \dots, p$$

2. Besaran yang diperlukan

Menghitung $-2\log\left(\frac{\text{likelihood tanpa variabel prediktor}}{\text{likelihood dengan variabel prediktor}}\right)$
 dengan bantuan *software* STATA versi 10.

3. Statistik Uji

$$\chi^2_{hitung} = -2\log\left(\frac{\text{likelihood tanpa variabel prediktor}}{\text{likelihood dengan variabel prediktor}}\right)$$

4. Kriteria Pengujian

Dengan mengambil taraf signifikansi α , maka H_0 ditolak jika

$$\chi^2_{hitung} > \chi^2_{1-\alpha; p}$$

5. Simpulan

Penafsiran dari H_0 diterima atau ditolak.

Uji Wald digunakan untuk menguji signifikansi masing-masing parameter. Adapun langkah-langkah pengujian signifikansi parameter regresi secara parsial dalam uji Wald sebagai berikut:

1. Perumusan Hipotesis

$$H_0 : \beta_{jk} = 0, \text{ untuk } k = 1, 2, \dots, p$$

$$H_1 : \beta_{jk} \neq 0, \text{ untuk } k = 1, 2, \dots, p$$

2. Besaran yang diperlukan

Hitung $\hat{\beta}_{jk}$ dan $SE(\hat{\beta}_{jk})$

3. Statistik Uji

$$Z_{jk} = \frac{\hat{\beta}_{jk}}{SE(\hat{\beta}_{jk})}$$

4. Kriteria Pengujian

Dengan mengambil taraf signifikansi α , maka H_0 ditolak jika $|z_{jk}| > z_{\alpha/2}$ atau nilai $p_{value} < \alpha$.

6. Kesimpulan

Penafsiran dari H_0 diterima atau ditolak.

METODE

Data yang digunakan dalam penelitian ini adalah data sekunder yang diambil dari data *General Social Survey of NORC* di Amerika Serikat. Data tersebut diunduh pada http://www.indiana.edu/~statmath/stat/all/cdvm/gss_cdvm.csv. *General Social Survey of NORC* merupakan survei sosiologis yang dilakukan untuk mengumpulkan data tentang karakteristik demografi dan sikap warga negara Amerika Serikat. Data yang dikumpulkan dalam survei ini meliputi informasi demografi dan pendapat responden mengenai berbagai hal mulai dari pengeluaran pemerintah, gaya hidup, kepercayaan seseorang, keyakinan seseorang dalam menganut agama dan lain-lain. Model probit bivariat ini akan

diterapkan pada kasus kepercayaan seseorang terhadap layanan internet yang dikaitkan dengan penggunaan internet serta faktor-faktor yang menentukan kemungkinan seseorang percaya terhadap layanan internet dan penggunaan internet. Dalam hal ini diambil data sampel berukuran 150 responden yang merupakan warga Amerika Serikat. Adapun, variabel-variabel yang digunakan dalam penelitian ini adalah:

1. Variabel respon:
 - $y_1=1$; Responden yang percaya terhadap layanan internet.
 - = 0; Responden yang tidak percaya terhadap layanan internet.
 - $y_2=1$; Responden yang menggunakan internet.
 - =0; Responden yang tidak menggunakan internet.
2. Variabel prediktor:
 - x_1 : Pendidikan responden (dalam tahun).
 - x_2 : Pendapatan keluarga (dalam jutaan).
 - x_3 : Usia responden (dalam tahun).
 - x_4 : Jenis kelamin, yaitu:
 - Bernilai 1 untuk responden laki-laki.
 - Bernilai 0 untuk responden perempuan.

Langkah-langkah analisis pemodelan probit bivariat antara lain:

1. Mengumpulkan dan menyeleksi data;
2. Melakukan analisis statistik deskriptif terhadap semua variabel yang terdapat dalam penelitian;
3. Melakukan analisis model regresi probit bivariate;
4. Melakukan pengujian galat pada model regresi probit bivariate;
5. Melakukan pengujian parameter secara simultan model regresi probit bivariate;
6. Melakukan pengujian parameter secara parsial model regresi probit bivariate;
7. Interpretasi model.

HASIL DAN PEMBAHASAN

Pada penelitian ini dianalisis faktor-faktor yang menentukan peluang seseorang percaya terhadap layanan internet dan penggunaan internet di Amerika Serikat dengan menggunakan model regresi probit bivariat.

Tabel 1. Taksiran Parameter

Variabel	Koefisien	Standar Error
y1		
x1	0.238031	0.0507486
x2	0.0011757	0.020688
x3	0.0288553	0.0102938
x4	0.1566175	0.2314989
_cons	-5.066164	0.96011
y2		
x1	0.2339402	0.0550138
x2	0.0420102	0.0187224
x3	-0.0105218	0.0103145
x4	-0.1817653	0.2493678
_cons	-3.161891	0.9406781

Sumber: Hasil output STATA versi 10 data *General Social Survey of NORC*

Selanjutnya, berdasarkan hasil yang ditampilkan pada tabel 1, maka diperoleh taksiran model regresi probit bivariat sebagai berikut:

$$y_1^* = -5,066164 + 0,2380310 x_{11} + 0,0011757 x_{21} + 0,0288553 x_{31} + 0,1566175$$

$$y_2^* = 3,161891 + 0,2339402 x_{12} + 0,0420102 x_{22} - 0,0105218 x_{32} - 0,1817653 x_{42}$$

Dengan menggunakan uji Lagrange Multiplier, Diperoleh nilai $\lambda = 0,326783$ yang menyatakan nilai untuk uji statistik LM dengan $p\text{-value} = 0,053$. Karena $0,053 > 0,05$, maka H_0 diterima. Nilai koefisien korelasi kedua persamaan tersebut tidak signifikan secara statistik. Artinya hubungan koefisien variabel prediktor dengan variabel responnya ditunjukkan oleh hasil kedua model probitnya masing-masing.

Sedangkan untuk uji perbandingan kemungkinan, diperoleh nilai $\chi^2_{hitung} = 50,8$ dengan bantuan *software* STATA versi 10. Nilai $\chi^2_{hitung} > \chi^2_{(0,05;8)}$ maka H_0 ditolak, artinya paling tidak terdapat satu nilai β_{jk} yang tidak sama dengan nol, sehingga secara bersama-sama (simultan) variabel prediktor yaitu pendidikan, pendapatan, usia, dan jenis mempunyai pengaruh terhadap peluang seorang responden yang percaya terhadap layanan internet dan responden yang menggunakan internet.

Tabel 2. Nilai z hitung

Variabel	Statistik z	Signifikan ($\alpha = 0,05$)
y1		
x1	4.69	Signifikan
x2	0.06	Tidak Signifikan
x3	2.80	Signifikan
x4	0.68	Tidak Signifikan
_cons	-5.28	Signifikan
y2		
x1	4.25	Signifikan
x2	2.24	Signifikan
x3	-1.02	Tidak Signifikan
x4	-0.73	Tidak Signifikan
_cons	-3.36	Signifikan

Sumber: Hasil output STATA versi data *General Social Survey of NORC*

Kriteria pengujiannya adalah dengan mengambil taraf signifikansi $\alpha = 0,05$, maka dari tabel distribusi normal baku diperoleh nilai $z_{\frac{1}{2}(1-0,05)} = z_{0,475} = 1,96$. Dengan melihat pengaruh variabel prediktor dalam mempengaruhi peluang kepercayaan seorang responden terhadap layanan internet dan responden yang menggunakan internet, maka diperoleh model regresi probit bivariat dengan parameter-parameter yang signifikan sebagai berikut:

$$y_1^* = 5,066164 + 0,238031 x_{11} - 0,0288553 x_{31}$$

$$y_2^* = 3,161891 + 0,2339402 x_{12} + 0,0420102 x_{22}$$

PENUTUP

Simpulan

Berdasarkan hasil dan pembahasan analisis model regresi probit bivariat pada kasus mengenai kepercayaan seseorang terhadap layanan internet yang dikaitkan dengan penggunaan internet, setelah dilakukan uji Lagrange Multiplier diketahui bahwa nilai galat untuk masing-masing variabel responnya tidak saling berkorelasi, sehingga perlu dilakukan analisis kembali untuk masing-masing persamaan regresi probitnya. Kemudian dari hasil analisis regresi probitnya dapat diketahui nilai peluang untuk setiap kepercayaan seorang responden terhadap layanan internet dan responden yang menggunakan internet. Diketahui bahwa secara bersama-sama (simultan) variabel prediktor yaitu pendidikan, pendapatan, usia, dan jenis kelamin mempunyai pengaruh

terhadap peluang seorang responden yang percaya terhadap layanan internet dan responden yang menggunakan internet. Sedangkan berdasarkan uji Wald, untuk seseorang yang percaya terhadap layanan internet diperoleh bahwa variabel pendidikan dan usia berpengaruh secara signifikan terhadap peluang kepercayaan terhadap layanan internet, selanjutnya untuk penggunaan internet diperoleh bahwa variabel pendidikan dan pendapatan mempunyai pengaruh yang signifikan terhadap peluang penggunaan internet.

Saran

Saran untuk penelitian selanjutnya dianjurkan selain model regresi probit bivariat ini, model regresi probit ini juga dapat dikembangkan lagi dengan beberapa variabel respon atau dikenal dengan model regresi probit multivariat (*multivariate probit regression model*). Pengolahan data untuk regresi probit bivariat ini dapat juga menggunakan bantuan software LIMDEP, selain software STATA. Pada uji Lagrange Multiplier, jika $\chi^2 < \chi^2_{(1-\alpha)}$ maka model regresi probit bivariat dapat dilanjutkan dengan menghitung nilai *marginal effect* untuk masing-masing variabel prediktornya.

DAFTAR PUSTAKA

- Abdillah, W., & Hartono, J. 2015. **Partial Least Square (PLS)-Alternatif Structural Equation Modeling (SEM) dalam Penelitian Bisnis**. Yogyakarta: Penerbit ANDI.
- Agresti, A.1990. **Categorical Analysis**. New York: John Wiley and Sons. Inc.
- Nachrowi, N. D dan Usman, H. 2002. **Penggunaan Teknik Ekonometri**. Jakarta: PT Raja Grafindo Persada.
- Park, H.M. 2009. **Regression Models for Binary Dependent Variables Using Stata, SAS, R, LIMDEP, and SPSS. Working Paper**. UITS Center for Statistical and Mathematical Computing: Indiana University.
- Scott Long, J and Freese, J. 2001. **Regression Models For Categorical Dependent Variables Using STATA**. Texas: A Stata Press Publication STATA Corporation.
- Trisna Pradnyantari, N, Gde Sukarsa, I dan Putu Suciptawati, N. 2015. **Penerapan regresi probit bivariat untuk menduga faktor-faktor yang mempengaruhi kelulusan mahasiswa (studi kasus: Mahasiswa Fakultas MIPA Universitas Udayana)**. *E-Jurnal Matematika*, 4(2), 49-53.
- Yong, B. 2003. **Penaksir maksimum likelihood bagi model probit dan model probit bivariat**. *INTEGRAL*, 8(1), 11-18.