

ANALISIS FAKTOR YANG MEMPENGARUHI LAMA MASA STUDI MAHASISWA FMIPA UNTAN MENGGUNAKAN MODEL PROBIT

Yude Anggreni, Dadan Kusnandar, Shantika Martha

INTISARI

Model probit (probability unit) adalah model yang digunakan ketika variabel dependen (Y) berbentuk dikotomi atau biner dan galatnya diasumsikan berdistribusi normal. Penelitian ini bertujuan untuk menganalisis faktor yang mempengaruhi lama masa studi mahasiswa FMIPA Untan dengan menggunakan model probit. Data yang digunakan adalah data kelulusan mahasiswa FMIPA Untan angkatan 2011- Untuk mengestimasi parameter model probit digunakan Maximum Likelihood Estimation (MLE). Kemudian lakukan uji kebaikan dengan tiga tahapan, yaitu uji serentak, uji parsial, serta koefisien determinasi. Hasil dari ketiga uji menunjukkan bahwa model probit dapat digunakan untuk menganalisis faktor yang mempengaruhi lama masa studi mahasiswa FMIPA Untan dengan variabel independennya jenis kelamin, TUTEP (Tanjungpura University Test of English Proficiency), IPK (Indek Prestasi Mahasiswa) dan Jurusan. Berdasarkan hasil dari pengujian menunjukkan bahwa faktor yang mempengaruhi lama masa studi FMIPA Untan adalah variabel independen IPK (X_3), dengan nilai McFadden R-square (R_{MF}^2) sebesar 0,1745.

Kata kunci: Model Probit, Uji Serentak, Uji Parsial, McFadden R-square.

PENDAHULUAN

Analisis regresi merupakan analisis yang biasa digunakan untuk tujuan peramalan atau pendugaan tentang nilai variabel dependen (Y). Analisis regresi berkaitan dengan hubungan stokastik antara variabel dependen (Y) dengan variabel independen (X). Tujuan penggunaan analisis regresi adalah untuk membangun suatu model probabilistik yang dapat digunakan untuk meramalkan atau menduga nilai variabel dependen (Y), berdasarkan pada nilai-nilai variabel independen (X). Analisis regresi linier terbagi menjadi analisis regresi sederhana dan analisis regresi berganda. Regresi linier sederhana, yaitu tentang penduga satu variabel Y oleh satu variabel X saja, sedangkan regresi berganda (*multiple regression*) merupakan perluasan dari regresi linier sederhana, yaitu regresi dengan dua atau lebih variabel X [1]. Pada model regresi linier, variabel dependen harus bersifat kuantitatif, namun kenyataannya sering ditemukan kasus dengan variabel dependen yang bersifat kualitatif, sehingga model regresi tidak dapat digunakan. Pada permasalahan tersebut dapat diatasi dengan beberapa cara diantaranya *linear probability model*, model logit dan model probit [2].

Regresi probit merupakan model regresi yang dapat menjelaskan hubungan antara variabel dependen diskrit berskala nominal/ordinal dengan variabel independen yang terdiri dari variabel kontinu, diskrit atau campuran antara keduanya. Regresi probit pertama kali diperkenalkan oleh Chester Bliss pada tahun 1935. Istilah probit dalam model regresi probit adalah singkatan dari *probability unit* sehingga dapat dikatakan bahwa model regresi probit merupakan suatu model regresi yang berkaitan dengan unit-unit probabilitas. Penelitian ini menggunakan model probit untuk studi kasus di Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Untan. Seorang mahasiswa membutuhkan waktu normal selama empat sampai lima tahun untuk menyelesaikan program sarjananya. akan tetapi dalam kenyataannya mahasiswa tidak selalu dapat menyelesaikan studinya dalam waktu yang normal. Penelitian ini bertujuan untuk mengetahui variabel-variabel yang berpengaruh terhadap lama masa studi mahasiswa FMIPA Untan dengan menggunakan model probit. Data yang digunakan adalah data mahasiswa lulusan FMIPA Untan angkatan 2011. Variabel yang

digunakan adalah lama masa studi mahasiswa sebagai variabel dependen (Y) yang dikategorikan menjadi dua kategori yaitu kurang dari sama dengan lima tahun yang dikategorikan dengan nol dan lebih dari lima tahun sampai tujuh tahun dengan satu, serta jenis kelamin (X_1) nol untuk perempuan, satu untuk laki-laki, nilai TUTEP (X_2) nol untuk ≥ 425 dan satu untuk < 425 , IPK (X_3) nol untuk ≥ 3 dan satu untuk < 3 , dan jurusan (X_4) nol untuk matematika, satu untuk kimia, dua untuk biologi, tiga untuk fisika dan empat untuk sistem komputer, sebagai variabel independen. Untuk pendugaan parameternya menggunakan *Maximum Likelihood* dan pengujian parameternya secara serentak menggunakan *likelihood ratio test*, dan pengujian parameter secara parsial menggunakan uji Wald serta untuk koefisien determinasinya menggunakan *McFadden-R²*.

UJI NORMALITAS DAN UJI MULTIKOLINEARITAS

Uji normalitas dilakukan untuk mengetahui apakah galat berdistribusi normal, karena model yang baik memiliki galat yang berdistribusi normal. Hipotesis yang digunakan dalam pengujian ini adalah:

H_0 : galat berdistribusi normal

H_1 : galat tidak berdistribusi normal

Kriteria pengambilan keputusan adalah terima H_0 jika *p.value* pada pengujian *Kolmogorov-Smirnov* lebih besar dari α . Adapun nilai KS_{hitung} diperoleh dengan rumus sebagai berikut [3].

$$KS_{hitung} = \max |F(z_i) - S(z_i)| \quad (1)$$

Dengan $z_i = \frac{X_i - \bar{X}}{S}$; \bar{X} merupakan rata-rata dan S merupakan simpangan baku sampel, dan

$F(z_i) = \frac{z_1, z_2, z_3, \dots, z_n \leq z_i}{n}$ serta $S(z_i) = P(z \leq z_i)$. Jika galat tidak berdistribusi normal maka dapat dilakukan transformasi Box-Cox terhadap data.

Uji multikolinearitas dilakukan untuk mengetahui apakah ada atau tidaknya hubungan yang linier antara variabel independen yang signifikan pada model regresi. Untuk mengetahui terjadinya gangguan multikolinearitas dapat dilihat dari besarnya nilai *Variance Inflation Factor (VIF)* [1]. Jika nilai *VIF*nya lebih besar dari 10, maka terjadi multikolinearitas. Adapun nilai *VIF* dapat dihitung dengan rumus sebagai berikut:

$$VIF_j = \frac{1}{1 - R_j^2} \quad (2)$$

Dengan R_j^2 merupakan koefisien determinasi yang didapat dari variabel independen X_j terhadap p variabel independen lainnya [3].

MODEL PROBIT

Model probit merupakan model *non*-linier yang digunakan untuk menganalisis hubungan antara satu variabel dependen dengan beberapa variabel independen, dengan variabel dependennya berupa data dikotomi yang bernilai satu atau nol. Analisis regresi probit merupakan analisis yang digunakan untuk meneliti hubungan variabel dependen yang bersifat kategori (kualitatif) dan variabel-variabel independen yang bersifat kualitatif maupun kuantitatif. Model probit menggunakan *Normal Cumulatif Distribution Function (CDF)* untuk menjelaskan fungsi persamaannya, karena regresi probit

menggunakan pendekatan distribusi normal [2]. Secara umum model probit dapat dinyatakan dalam bentuk sebagai berikut:

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon \quad (3)$$

dengan y_i merupakan fungsi peluang kumulatif, $\beta_0, \beta_1, \beta_2$ dan β_p sebagai parameter regresi, x_1, x_2 dan x_p merupakan variabel prediktor regresi dan ε merupakan faktor galat.

Pendugaan dalam model probit biasanya dilakukan dengan menggunakan metode maksimum *likelihood*. Metode maksimum *likelihood* pertamakali diperkenalkan oleh R.A. Fisher pada tahun 1912. Metode ini memberikan hasil pendugaan yang baik bagi θ , terutama jika sampelnya besar, sehingga metode ini biasanya digunakan untuk pendugaan [4]. Metode *maximum likelihood* (MLE) adalah metode estimasi yang memaksimumkan fungsi *likelihood*-nya, jika terdapat variabel acak X_1, X_2, \dots, X_n dengan fungsi kepadatan peluang adalah $f(x_i | \theta)$ dengan $i = 1, 2, 3, \dots, n$ dan θ adalah parameter yang diduga. Fungsi kepadatan peluang bersama dari variabel acak X_1, X_2, \dots, X_n disebut sebagai fungsi *likelihood* yang dinyatakan dengan $L(\theta)$.

$$\begin{aligned} L(\theta) &= f(x_1 | \theta) * f(x_2 | \theta) \dots * f(x_n | \theta) \\ &= \prod_{i=1}^n f(x_i | \theta) \end{aligned} \quad (4)$$

Variabel dependen dalam probit mengikuti distribusi Bernoulli karena hanya terdapat dua kemungkinan, yaitu ya ($y = 0$) dan tidak ($y = 1$), serta mengasumsikan galat berdistribusi normal [5]. Jika Ω interval terbuka dan $L(\theta)$ dapat diturunkan serta diasumsikan maksimum dalam Ω , maka penduga maksimum *likelihood*-nya adalah:

$$\frac{d}{d\theta} L(\theta) = 0 \quad (5)$$

Penduga maksimum bagi θ juga dapat diperoleh dengan memaksimumkan fungsi *log-likelihood*-nya yaitu dengan mencari turunan dari $\log L(\theta)$, dan hasil turunannya sama dengan nol.

$$\frac{d}{d\theta} \log L(\theta) = 0 \quad (6)$$

Uji Parameter Secara Serentak

Pengujian parameter secara serentak atau disebut juga uji kecocokan model secara bersama-sama digunakan *likelihood ratio test* atau uji simultan variabel independen. Model hipotesis yang digunakan adalah:

H_0 : Tidak ada pengaruh antara variabel independen dengan variabel dependen.

H_1 : Paling sedikit ada satu variabel independen yang berpengaruh terhadap variabel dependen.

Statistik uji yang digunakan dalam pengujian parameter secara serentak dirumuskan sebagai berikut: [2]

$$G^2 = -2 \log \left[\frac{L(H_0)}{L(H_1)} \right]; L(H_1) \neq 0 \quad (7)$$

dengan $L(H_0)$ = nilai fungsi *likelihood* dengan variabel independen (model penuh); $L(H_1)$ = nilai fungsi *likelihood* tanpa variabel independen (hanya terdapat sebuah konstanta). Kriteria pengujian adalah H_0 ditolak pada $G^2 > \chi^2_{[\alpha, df]}$

Uji Parameter Secara Parsial

Pengujian parameter secara parsial pada masing-masing parameter dapat diuji dengan *Wald test*. Hipotesis yang digunakan dalam pengujian ini adalah [2]:

H_0 : Variabel independen ke- j tidak berpengaruh terhadap variabel dependen.

H_1 : Ada pengaruh antara variabel independen ke- j dengan variabel dependen.

Statistik uji:
$$W = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \quad (8)$$

dengan $\hat{\beta}_j$ = nilai koefisien dari parameter β_j ; $SE(\hat{\beta}_j)$ = standard error bagi parameter β_j . Kriteria pengujian adalah tolak H_0 jika nilai $|W| \geq Z_{\alpha/2}$ atau $p.value < \alpha$.

Koefisien Determinasi

Dalam model regresi linier besarnya koefisien determinasi (R^2) seringkali digunakan untuk uji kebaikan model. Namun dalam model probit (R^2) kurang sesuai digunakan untuk melakukan uji kebaikan. Adapun uji kebaikan yang sesuai untuk model probit dimana variabel dependennya terbatas yang hanya bernilai nol dan satu adalah *Pseudo- R^2* . *Pseudo- R^2* yang sering digunakan adalah *Pseudo- R^2* dari *McFadden* yaitu *McFadden R-square* (R_{MF}^2) [6], yang didefinisikan sebagai:

Statistik uji:
$$R_{MF}^2 = 1 - \left[\frac{L(H_0)}{L(H_1)} \right]; L(H_1) \neq 0 \quad (9)$$

STUDI KASUS

Data penelitian yang digunakan adalah data sekunder yaitu data mahasiswa FMIPA Untan angkatan 2011, yang diperoleh dari Kasubbag Pendidikan dan Kemahasiswaan FMIPA Untan. Banyak data yang digunakan ada 148 data dengan variabel lama masa studi (Y) sebagai variabel dependen dan lima variabel independen yaitu jenis kelamin (X_1), nilai TUTEP (*Tanjungpura University Test of English Proficiency*) (X_2), IPK (X_3), jurusan (X_4). Berikut ini statistik deskriptif dari data penelitian.

Tabel 1 Statistik Deskriptif

Variabel	Rata-rata	Minimum	Maksimum	Simpangan baku	Koefisien keragaman
Y	5,63	3,49	7,00	1,01	18%
X_2	419,43	333,00	536,00	39,49	9%
X_3	3,12	2,37	3,95	0,33	11%

Tabel 2 Jumlah Mahasiswa

X_1	X_4					Total
	Matematika	Kimia	Biologi	Fisika	Rekayasa Sistem Komputer	
Perempuan	17	29	19	20	6	91
Laki-laki	15	6	10	5	21	57
Total	32	35	29	25	27	148

Dari Tabel 1 diketahui rata-rata lama studi mahasiswa FMIPA Untan angkatan 2011 adalah 5,63 tahun dengan rata-rata IPK 3,12. Dari Tabel 2 juga dapat dilihat bahwa rata-rata jumlah mahasiswa untuk setiap jurusan tidak terlalu jauh berbeda, dan kebanyakan adalah mahasiswi dibanding mahasiswa. Hasil *p.value* Kolmogorov-Smirnovnya $0,027 > \alpha$ (1%), yang artinya galat berdistribusi normal.

Tabel 3 Uji multikolinearitas

Variabel	X_1	X_2	X_3	X_4
VIF	1,039	1,205	1,213	1,039

Berdasarkan Tabel 3 diketahui bahwa nilai VIF untuk semua variabel independen kurang dari 10, maka dapat disimpulkan antar variabel independen tidak terjadi multikolinearitas, sehingga dapat dilakukan pengujian secara serentak.

Dari analisis diperoleh nilai $G^2 = 153,4187 \geq \chi^2_{0,01;3}$ (11,35), sehingga H_0 ditolak, yang artinya paling sedikit ada satu variabel independen yang berpengaruh terhadap lama masa studi. Untuk mengetahui variabel independen apa saja yang mempengaruhi lama masa studi dilakukan Uji Parsial.

Tabel 4 Uji Parsial

Variabel	X_1	X_2	X_3	X_4
W	-0,833	0,064	-4,726	2,514
$Z_{0,025}$	2,58	2,58	2,58	2,58
<i>p.value</i>	0,485	0,949	0,000	0,012

Berdasarkan Tabel 4 dapat dilihat pada variabel X_1 , X_2 dan X_4 memiliki nilai Z_{hitung} atau $|W|$ lebih kecil dari nilai Z_{tabel} , sehingga H_0 diterima, yang artinya tidak ada pengaruh antara variabel X_1 , X_2 dan X_4 dengan variabel Y , sedangkan X_3 memiliki nilai Z_{hitung} atau $|W|$ lebih besar dari nilai Z_{tabel} , sehingga H_0 ditolak, yang artinya ada pengaruh antara variabel X_3 dengan variabel Y .

Tabel 5 Pengujian Tanpa Variabel X_1 , X_2 dan X_4

Variabel	Coefficient	W	$Z_{0,005}$	<i>p.value</i>
C	7,143			
X_3	-2,149	-5,290	2,58	0,000

Berdasarkan Tabel 5 tanpa memasukan variabel X_1 , X_2 dan X_4 , maka diperoleh faktor-faktor yang mempengaruhi lama masa studi mahasiswa FMIPA Untan adalah X_3 (IPK) dengan model probitnya adalah $y_i = 7,143 - 2,149x_3$.

Untuk mengetahui berapa besar peluangnya dapat dilihat dengan menggunakan tabel Z .

Misalkan, jika seorang mahasiswa memiliki IPK = 3 maka,

$$y_i = 7,143 - 2,149(3) = 0,696 = 0,70.$$

Kemudian mentransformasikan nilai y_i menjadi peluang dengan melihat tabel-Z

$$y_i(0,32) = P(Z < 0,32) = 0,758 = 75,8\%.$$

Selanjutnya dikurangkan peluangnya dengan nilai satu, sehingga diperoleh $1 - 0,758 = 0,242$.

Jadi, peluang seorang mahasiswa lulus kurang dari sama dengan lima tahun, dengan IPK = 3 adalah sebesar 24,2%.

Lakukan hal yang sama untuk setiap nilai IPK:

Tabel 6 Peluang seorang mahasiswa lulus kurang dari sama dengan lima tahun.

X_3	2,5	2,75	3	3,25	3,5	3,75
$P(Z \leq 5)$	3,84%	10,93%	24,20%	43,64%	64,80%	82,12%

Dari Tabel 6 dapat dilihat bahwa semakin tinggi IPK seorang mahasiswa maka semakin besar peluangnya untuk lulus kurang dari sama dengan lima tahun.

KESIMPULAN

Berdasarkan analisis data diperoleh hasil uji parameter secara serentak menyatakan paling sedikit ada satu variabel independen yang berpengaruh terhadap variabel dependen, yang artinya model probit dapat digunakan untuk kasus lama masa studi mahasiswa FMIPA Untan. Dari hasil uji *Wald* diketahui bahwa faktor yang mempengaruhi lama masa studi mahasiswa FMIPA Untan adalah variabel IPK (x_3), dengan nilai koefisien determinasi dari *McFadden* (R_{MF}^2) sebesar 0,1745 yang artinya variabel IPK (x_3) mempengaruhi variabel lama masa studi mahasiswa (Y) sebesar 17,45%.

Model probit yang didapat adalah $y_i = 7,143 - 2,149x_3$. Dari model dapat diartikan bahwa setiap penurunan IPK (x_3) memberikan pengaruh negatif pada lama masa studi mahasiswa FMIPA Untan (semakin rendah IPK, maka semakin lama masa studi mahasiswa). Untuk mengetahui peluang setiap mahasiswa lulus dapat dilihat menggunakan tabel Z.

DAFTAR PUSTAKA

- [1]. Kusnandar, D., Debararaja, N. N., Mara, M. N., dan Satyahadewi. N. *Metode Statistika serta Aplikasinya dengan Minitab, Excel dan R*. Pontianak;2019.
- [2]. Mallidina, T., Tantular, B., Faidah, D. F., dan Darmawan, G. Analisis Regresi Probit untuk Menentukan Peluang Kemenangan Pemain dalam Permainan *Age of Empire 2*. *Prosiding Seminar Nasional MIPA 2016*. 2016.
- [3]. Suliyanto. *Ekonometrika Terapan: Teori & Aplikasi dengan SPSS*. Yogyakarta;2011.
- [4]. DeGroot, M. H. *Probability and Statistic*, Ed ke-2. Canada;1986.
- [5]. Greene, W. H. *Econometric Analysis 5th ed*. New York;2003.
- [6]. Daga, E. K. N., dan Suryowati, K. Penerapan Metode Regresi Logistik Ordinal dan Regresi Probit Ordinal untuk Mengestimasi Probabilitas Lama Masa Studi Mahasiswa IST Akprind Yogyakarta. *Jurnal Statistika Industri dan Komputasi*. 2017; 2(2); 104-114.

YUDE ANGGRENI :Jurusan Statistika FMIPA UNTAN,Pontianak
yudeanggreni@student.untan.ac.id
 DADAN KUSNANDAR :Jurusan Matematika FMIPA UNTAN,Pontianak
dkusnandar@math.untan.ac.id
 SHANTIKA MARTHA :Jurusan Matematika FMIPA UNTAN,Pontianak
shantika.martha@math.untan.ac.id