Generalized Ordinal Logistic Regression Model pada Pemodelan Data Nilai Pesantren Mahasiswa Baru FMIPA Universitas Islam Bandung Tahun 2017

Generalized Ordinal Logistic Regression Model on Data Modeling Student Pesantren Value FMIPA Islamic University of Bandung Year 2017

[1]Kustiani Rahmah, [2]Anneke Iswani Achmad, [3]Nusar Hajarisman,
[1,2,3]Prodi Statistika, FakultasMatematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Jl. Tamansari No.1 Bandung 40116
email: [1]kustiani2224@gmail.com, [2]nusarhajarisman@yahoo.com, [3]annekeiswani11@gmail.com

Abstract. Logistic regression is one method that can be used to get functional relationship between categorical response variable with predictor variable. There are several kinds of logistic regression one of them is Generalized Ordinal Logistic Regression Model. This regression describes the functional relationship between the response variable that is ordinal and the predictor variable. This paper will discuss the use of Generalized Ordinal Logistic Regression Model on new students' pesantren exam data at Faculty of Mathematics and Natural Sciences (F-MIPA) of Islamic University of Bandung in 2017. Parameter estimation is done using likelihood maximum method, then solving nonlinear equations from Log-likelihood functions are estimated by numerical methods using the Newton-Raphson iteration process. Test of model significance using likelihood ratio test.

Keywords: Logistic Regression, Generalized Ordinal Logistic Regression Model, Maximum Likelihood, Newton-Raphson, Likelihood Ratio.

Abstrak. Regresi logistik merupakan salah satu metode yang dapat digunakan untuk mendapatkan hubungan fungsional antara variabel respon yang bersifat kategorik dengan variabel prediktor. Ada beberapa macam regresi logistik salah satunya adalah *Generalized Ordinal Logistic Regression Model*. Regresi ini menggambarkan hubungan fungsional antara variabel respon yang bersifat ordinal dengan variabel prediktor. Makalah ini akan membahas penggunaan *Generalized Ordinal Logistic Regression Model* pada data nilai ujian pesantren mahasiswa baru di Fakultas Matematika dan Ilmu Pengetahuan Alam (F-MIPA) Universitas Islam Bandung tahun 2017. Penaksiran parameter dilakukan menggunakan metode maksimum *likelihood,* selanjutnya memecahkan persamaan non-linear dari fungsi log-*likelihood* yang ditaksir dengan metode numerik menggunakan proses iterasi Newton-Raphson. Uji signifikansi model menggunakan uji rasio *likelihood.*

Kata Kunci: Regresi Logistik, Generalized Ordinal Logistic Regression Model, Maksimum Likelihood, Newton-Raphson, Rasio Likelihood.

A. Pendahuluan

Regresi logistik merupakan salah satu metode yang dapat digunakan untuk mendapatkan hubungan fungsional antara variabel respon yang bersifat kategorik dengan variabel prediktor (Agresti, 2007). Regresi Logistik Biner adalah regresi dengan variabel respon yang mempunyai dua kategori atau dua kejadian, yaitu sukses (Y=1) atau gagal (Y=0). Ketika variabel respon Y memiliki lebih dari dua kategori dan dapat diperingkat, atau variabel respon bersifat ordinal, maka model yang digunakan adalah regresi logistik ordinal.

Model yang dapat dipakai untuk regresi logistik ordinal adalah model logit kumulatif. Pada model logit ini sifat ordinal dari respon Y dituangkan dalam peluang kumulatif, sehingga model didapatkan dengan membandingkan peluang kumulatif yaitu peluang kurang dari atau sama dengan kategori respon ke-j pada *p* variabel prediktor (Hosmer dan Lemeshow, 2000). Model *Proportional Odd* (PO) yang dikemukakan oleh McCullagh (1980) merupakan jenis regresi logistik ordinal yang umum digunakan. Model PO digunakan untuk menaksir peluang kumulatif pada level tertentu atau di bawah level tertentu dari variabel respon. Dalam model ini, pengaruh dari setiap variabel prediktor diasumsikan sama disemua kategori variabel respon ordinal. Hal ini disebut sebagai asumsi proporsional odds (PO). Asumsi proporsional odds untuk analisis regresi ordinal sering dilanggar karena dipengaruhi oleh ukuran sampel dan kovariat yang ada di dalam model, misalnya kovariat yang kontinu atau adanya interaksi di dalam model (Allison, 1999). Ketika asumsi PO tersebut dilanggar maka dapat menyebabkan kesalahan interpretasi pada model.

Pengaruh dari masing-masing variabel prediktor terhadap variabel respon ordinal mungkin saja tidak sama. Hal ini bertentangan dengan asumsi proporsional odds tersebut. Karena itu untuk mengatasi masalah ini digunakanlah metode *Generalized Ordinal Logistic Regression Model* (GOLRM) yang dikembangkan oleh Fu (1998) dan William (2006). penggunaan *Generalized Ordinal Logistic Regression Model* jarang digunakan karena terbatasnya perangkat lunak statistika saat ini. *Generalized Ordinal Logistic Regression Model* dapat dihitung dengan menggunakan SAS. Tetapi pada program SAS mensyaratkan restrukturisasi data sebelum analisis. Karena itu Fu dan William mengembangkan program gologit yang dapat digunakan pada aplikasi Stata. Dimana program gologit untuk *Generalized ordinal logistic regression model* mampu mengatasi asumsi proporsional odds dengan mengakomodasi pengaruh dari setiap variabel prediktor tanpa merestrukturisasi data.

Untuk mengaplikasikan metode *Generalized Ordinal Logistic Regression Model* maka digunakan data nilai ujian pesantren mahasiswa baru Fakultas Mtematika dan Ilmu Pengetahuan Alam (F-MIPA) di Universitas Islam Bandung (Unisba) tahun 2017. Hasil nilai ujian pesantren ini berupa huruf mutu yang bersifat ordinal yaitu dengan nilai terbaik sampai terendah secara berurutan adalah A,B,C,D & E (dimana D & E dinyatakan tidak lulus).

B. Landasan Teori

Regresi Logistik Ordinal

Regresi logistik ordinal adalah suatu analisis regresi yang digunakan untuk menggambarkan hubungan fungsional antara variabel respon dengan sekumpulan variabel prediktor, dimana variabel respon bersifat ordinal, dan setiap kategori dapat diperingkat (Hosmer & Lemeshow, 2000).

Jika diasumsikan terdapat variabel respon Y berskala ordinal dengan J kategori dan variabel prediktor sebanyak k, maka peluang dari variabel respon kategori ke-*j* pada variabel prediktor X tertentu dinyatakan dengan $P[(Y = j|x)] = \pi_j(x)$ dan peluang kumulatifnya adalah $P[(Y \leq j|x)] = \pi_1(x) + \ldots + \pi_j(x)$. Model logit kumulatif didefinisikan sebagai (Agresti 2002):

$$\begin{aligned} L_j(x) &= \text{logit } P[(Y \leq j|x)] \\ &= \log \frac{P(Y\leq j|x)}{1-P(Y\leq j|x)} \\ &= \beta_{0j} + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k \end{aligned} \quad \ldots(1)$$

Dengan j = 1,..., J-1

Peluang kumulatif digunakan untuk menduga parameter, maka metode *likelihood* dapat ditulis sebgai perkalian dari J-1 kategori, sehingga fungsi kumulatif peluang bersama dari Y = (Y_1, Y_2,..., Y_n) adalah sama dengan perkalian n fungsi multinomial. Fungsi *likelihood* untuk n ampel random adalah:

$$L(\beta) = \prod_{i=1}^{n} \pi_1(x_i)^{z_{i1}}\, \pi_2(x_i)^{z_{i2}} \ldots \pi_j(x_i)^{z_{ij}} \qquad \ldots(2)$$

Fungsi log *likelihood*-nya adalah:

$$l(\beta) = \sum_{i=1}^{n} z_{01} ln[\pi_1(x_i)] + \ldots + z_{ij} ln\left[\pi_j(x_i)\right] \qquad \ldots(3)$$

Dengan $z_{ij} = \begin{cases} 1\ untuk\ y = j \\ 0\ untuk\ y\ \neq j \end{cases}$

Pendugaan parameter pada model regresi logistik ordinal didapat dengan memaksimumkan fungsi log *likelihood* terhadap parameter (Hosmer dan Lemeshow, 2000). Maksimum log *likelihood* diperoleh dengan mendeferensialkan $l(\beta)$ terhap β dan menyamakan dengan nol. Untuk menaksir β diperlukan turunan kedua dari fungsi log *likelihood* dimana hasil dari turunan kedua ini bernilai negatif. Untuk mendapatkan nilai tersebut digunakan metode iterasi Newton Raphson (Agresti, 2002).

Generalized Ordinal Logistic Regression Model **(GOLRM)**

Model regresi logstik ordinal yang sering digunakan adalah model proporsional odds (PO). Dalam model ini, pengaruh dari setiap variabel prediktor diasumsikan sama disemua kategori variabel respon ordinal. Hal ini disebut sebagai asumsi proporsional odds (PO).

Pengaruh dari masing-masing variabel prediktor terhadap variabel respon ordinal mungkin saja tidak sama. Hal ini bertentangan dengan asumsi proporsional odds tersebut. Karena itu untuk mengatasi masalah ini digunakanlah metode Generalized Ordinal Logistic Regression Model (GOLRM) yang dikembangkan oleh Fu (1998) dan William (2006). Model logit untuk GOLRM adalah sebagai berikut:

$$\ln(Y_j) = ln\left(\frac{\pi_{ij}(x)}{1-\pi_{ij}(x)}\right) = \beta_{0j} + \beta_{jk} X_{ik} \qquad \ldots(4)$$

Model pada persamaan (2.23) dapat juga ditulis sebagai berikut (William, 2006):

$$logit[\pi(y > j|x_1, x_2, \ldots, x_k)] = \left(\frac{\pi(Y > j|x_1, x_2, \ldots, x_k)}{\pi(Y \leq j|x_1, x_2, \ldots, x_k)}\right)$$

$$= \beta_{0j} + \beta_{1j} X_1 + \beta_{2j} X_2 + \cdots + \beta_{kj} X_k \qquad \ldots(5)$$

Dimana dalam kedua model tersebut:

β_{0j} adalah intercept

β_{kj} adalah koefisien logit

X_k adalah variabel prediktor

Penaksiran Parameter *Generalized Ordinal Logistic Regression Model*

Metode pendugaan parameter yang dapat digunakan pada generalized ordinal logistic regression model adalah dengan metode penaksiran maksimum *likelihood*. Prinsip dari metode maksimum *likelihood* adalah mencari nilai β dengan memaksimumkan fungsi *likelihood*.

Misal J merupkan banyaknya kategori dari variabel respon, dimana $J \geq 2$. Variabel acak Y dapat mengambil salah satu dari nilai J yang mungkin. Setiap pengamatan saling bebas dan setiap y_i adalah variabel acak multinomial. N mewakili banyak populasi dan n merupakan kolom vektor dengan unsur n_i yang mewakili banyak pengamatan dalam populasi i untuk i=1 sampai N. Data dikumpulan ke dalam setiap populasi yang mempresentasikan satu kombinasi variabel-variabel prediktor.

Matriks y adalah matriks dengan N baris dan J-1 kolom. Untuk setiap populasi,

y_{ij} mewakili perhitungan pengamatan nilai ke j dari y_i. Demikian pula π merupakan matriks dari dimensi yang sama seperti y dimana setiap unsur π_{ij} menunjukan peluang pengamatan nilai ke-j variabel respon dalam populasi ke-i. Desain matriks variabel prediktor X, berukuran N baris dan (K+1) kolom, dimana K adalah banyaknya varabel prediktor dan unsur pertama pada setiap baris x_{i0} =1 sebagai intersep. β menjadi matriks dengan $K+1$ baris dan J-1 kolom, sehingga setiap unsur β_{kj} mengandung penaksiran parameter untuk kovariat ke-k dan nilai variabel respon ke-j. Untuk model logit sebagai berikut:

$$\ln\left(\frac{\pi_{ij}}{\pi_{iJ}}\right) = ln\left(\frac{\pi_{ij}}{1-\sum_{j=1}^{J-1}\pi_{iJ}}\right) = \sum_{k=0}^{K} x_{ik}\beta_{jk}$$

i = 1,2,...,N
j = 1,2,...,J-1
k = 1,2,...,K

dimana:

$$\pi_{ij} = \frac{exp\sum_{k=0}^{K} x_{ik}\beta_{jk}}{1+\sum_{j=1}^{J} exp\sum_{k=0}^{K} x_{ik}\beta_{jk}} \quad ...(6)$$

$$\pi_{iJ} = \frac{1}{1+\sum_{j=1}^{J} exp\sum_{k=0}^{K} x_{ik}\beta_{jk}} \quad ...(7)$$

Untuk setiap populasi, variabel respon mengikuti distribusi multinomial dengan tingkat J. Fungsi densitas gabungannya adalah:

$$f(y|\beta) = \prod_{i=1}^{N}\left[\frac{n_i!}{\prod_{j=1}^{J} y_{ij}!}\prod_{j=1}^{J}\pi_{ij}^{y_{ij}}\right] \quad ...(8)$$

maka fungsi *likelihoodnya* adalah:

$$L(\beta|y) = \prod_{i=1}^{N}\prod_{j=1}^{J-1}\left(\frac{\pi_{ij}}{\pi_{iJ}}\right)^{y_{ij}}\pi_{iJ}^{n_i} \quad ...(9)$$

Dengan mensubstitusikan nilai π_{ij} dan π_{iJ} pada persamaan (6) ke dalam persamaan (7), fungsi *likelihood* menjadi:

$$L(\beta|y) = \prod_{i=1}^{N}\prod_{j=1}^{J-1} exp y_{ij}\sum_{k=0}^{K} x_{ik}\beta_{jk}\left(1+\sum_{j=1}^{J-1} exp\sum_{k=0}^{K} x_{ik}\beta_{jk}\right)^{-n_i} \quad ...(10)$$

Fungsi log *likelihood*-nya adalah:

$$l(\beta) = \prod_{i=1}^{N}\prod_{j=1}^{J-1}\left(y_{ij}\sum_{k=0}^{K} x_{ik}\beta_{jk}\right) - n_i log\left(1+\sum_{j=1}^{J-1} exp\sum_{k=0}^{K} x_{ik}\beta_{jk}\right) \quad ...(11)$$

Berkaitan dengan penaksiran parameter selanjutnya akan menggunakan metode Newton-Raphson yang melibatkan turunan pertama dan kedua dari fungsi log-*likelihood*. Turunan pertama dari fungsi log-*likelihood* adalah:

$$\frac{\partial l(\beta)}{\partial \beta_{jk}} = \sum_{i=1}^{N} y_{ij}x_{ik} - n_i\pi_{ij}x_{ik} \quad ...(12)$$

Turunan pertama dari fungsi log-*likelihood* merupakan sebuah matriks berukuran (J-1)(K+1) yang akan diatur sama dengan nol. Selanjutnya turunan parsial kedua dapat diperoleh dari turunan pertama fungsi log-*likelihood* terhadap β_{jk}. Turunan kedua yaitu:

$$\frac{\partial^2 l(\beta)}{\partial \beta_{jk}\partial \beta_{j'k'}} = -\sum_{i=1}^{N} n_i\pi_{ij}\left(1-\pi_{ij}\right)x_{ik}x_{ik'} \quad ...\ (13)$$

turunan pertama dari fungsi *log-likelihood* adalah $l'(\beta) = X^T(y-\pi)$ dan turunan kedua dari fungsi *log-likelihood* adalah $l''(\beta) = -X^TWX$. Penaksir baru dapat dihitung dengan menggunakan rumus sebagai berikut:

$$\hat{\beta}^{(t+1)} = \hat{\beta}^{(t)} + \left(X^T W^{(t)} X\right)^{-1} X^T\left(y - \pi^{(t)}\right) \qquad ...(14)$$

Proses tersebut berlangsung secara terus menerus hingga kovergen, artinya tidak terdapat perubahan antara unsur β dari satu iterasi ke iterasi berikutnya. Apabila penaksiran kemungkinan maksimum dikatakan teah konvergen maka matriks $\left[X^T W^{(t)} X\right]^{-1}$ akan menjadi matriks varians-kovarians dari penaksiran parameter.

Pengujian Prameter Model Regresi Logistik

Pengujian parameter model dilakukan untuk memeriksa signifikansi variabel prediktor yang ada dalam model. Untuk mengetahui pengaruh variabel prediktor terhadap variabel respon secara serentak di dalam model regresi logistik maka digunakan statistik uji G yang merupakan uji rasio *likelihood* dengan hipotesis sebagai berikut:

H_0: $\beta_1 = \beta_2 = ... = \beta_j = 0$

H_1: minimal ada satu $\beta_j \neq 0$; j = 1,2,...,k dimana j adalah jumlah variabel prediktor.

Uji rasio *likelihood* menggunakan G statistik:

$$G = -2ln\left(\frac{L_o}{L_k}\right) \qquad ...(15)$$

Dimana L_o adalah fungsi *likelihood* tanpa variabel prediktor dan L_k adalah fungsi *likelihood* dengan variabel prediktor (Hosmer & Lemeshow, 2000). Kriteria ujinya yaitu H_0 ditolak bila $G > \chi^2_{(k,\alpha)}$ dimana *k* adalah jumlah prediktor dalam model, atau p-value < α yang dipilih.

Interpretasi *Generalized Ordinal Logistic Regression Model*

GOLRM dapat diinterpretasikan dengan menggunakan odds rasio (OR). Misalkan variabel respon dengan J kategori dan variabel prediktor. Odds rasio untuk penjelasan respon j(j=2,...,J) relatif terhadap referensi pertama, j = 1.

$$OR_j = \frac{\pi_{jp}/\pi_{ja}}{\pi_{1p}/\pi_{1a}} \qquad ...(15)$$

Logaritma dari odds rasio dapat ditulis sebagai:

$$logOR_j = log\left(\frac{\pi_{jp}}{\pi_{1p}}\right) - log\left(\frac{\pi_{ja}}{\pi_{1a}}\right) = \beta_{1j}$$

Dari persamaan diatas maka odds rasio dapat ditulis:

$$OR_j = exp\left(\beta_{1j}\right) \qquad ...(16)$$

Jika $\beta_{1j} = 0$ sehingga nilai $OR_j = 1$, maka faktor dalam variabel prediktor tidak berpengaruh.

C. Hasil Penelitian dan Pembahasan

Deskripsi Data

Data yang digunakan adalah data nilai pesantren mahasiswa baru FMIPA Unisba tahun 2017. Variabel prediktor dalam penelitian ini adalah jenis kelamin dan lama mengaji per hari, dimana data diperoleh dengan menggunakan kuesioner. Berdasarkan hasil perhitungan ukuran sampel minimum, didapatkan ukuran sampel minimumnya yaitu 187 mahasiswa. Dari sampel minimum tersebut didapatkan untuk program studi Statistika sampelnya sebanyak 57 mahasiswa, untuk program studi Farmasi sebanyak 109 mahasiswa, dan untuk program studi matematika sampelnya sebanyak 21 mahasiswa

Deskripsi nilai ujian pesantren berdasarkan variabel jenis kelamin dapat dilihat pada tabel 1.

Tabel 1. Deskripsi Nilai Ujian Pesantren Berdasarkan Variabel Jenis Kelamin

Variabel Jenis Kelamin	Nilai			Jumlah
	A	B&C	D&E	
Laki-laki	15	16	14	45
Perempuan	55	49	38	142
Jumlah	70	65	52	187

Berdasarkan tabel 1. dapat dilihat jumlah mahasiswa laki-laki adalah sebanyak 45 orang. Mahasiswa laki-laki yang mendapat nilai A adalah sebanyak 15 orang atau 33,33%, yang mendapatkan nilai B&C sebanyak 16 orang atau 35,56% dan yang mendapatkan nilai D&E sebanyak 14 orang atau 31,11%. Sedangkan untuk mahasiswa yang berjenis kelamin perempuan ada sebanyak 142 orang. Mahasiswa perempuan yang mendapat nilai A ada sebanyak 55 atau 38,73%, yang mendapatkan nilai B&C ada sebanyak 49 orang atau 34,51% dan yang mendapat nilai D&E ada sebanyak 26,76%.

Nilai minimum untuk lama mengaji per hari adalah sebesar 0 menit dan nilai maksimumnya adalah sebesar 150 menit.Berdasarkan hasil penelitian, waktu mengaji dikategorikan sebentar yaitu dengan lama mengaji 0-14 menit dimana mahasiswa yang lama mengaji pada waktu tersebut mendapatkan nilai D&E, sedangkan untuk waktu mengaji sedang adalah 15-30 menit dimana pada interval waktu tersebut mahasiswa kebanyakan mendapat nilai B&C dan untuk waktu mengaji lama adalah 31-150 menit dimana pada interval waktu tersebut mahasiswa mendapatkan nilai A. Pembagian waktu lama mengaji dapat dilihat pada tabel 2.

Tabel 2. Pembagian Kelompok Waktu Lama Mengaji

Waktu Lama mengaji	Kelompok
0 – 14 menit	Sebentar
15 – 30 menit	Sedang
31 – 150 menit	Lama

Selanjutnya deskripsi data variabel lama mengaji berdasarkan variabel jenis kelamin dapat dilihat pada tabel 3.

Tabel 3. Deskripsi Data Lama mengaji Berdasarkan Jenis Kelamin

Jenis Kelamin	Lama Mengaji			Jumlah
	Sebentar	Sedang	Lama	
Laki-laki	14	19	12	45
Perempuan	36	69	37	142
Jumlah	50	88	49	187

Dari tabel 3. dapat dilihat mahasiswa yang berjenis kelamin laki-laki untuk waktu mengaji sebentar ada 14 orang atau 31,11%, untuk waktu mengaji sedang ada 19 orang atau 42,22% dan untuk waktu mengaji lama ada 12 orang atau 26,67%. Sedangkan untuk mahasiswa perempuan dapat dilihat ada 36 orang atau 25,35% untuk waktu mengaji sebentar, ada 69 orang atau 48,59% untuk waktu mengaji sedang dan ada 37 orang atau 26,06% untuk waktu mengaji lama.

Selanjutnya deskripsi data nilai ujian pesantren mahasiswa berdasarkan lama mengaji per hari dapat dilihat pada tabel 4.

Tabel 4. Data Nilai ujian Pesantren Mahasiswa Baru Berdasaran Lama Mengaji

Lama mengaji	Nilai Ujian Pesantren			Jumlah
	A	B&C	D&E	
Sebentar	0	0	49	49
Sedang	21	65	2	88
Lama	50	0	0	50
Jumlah	71	65	51	187

Dari tabel 4. dapat dilihat persentase mahasiswa dengan waktu mengaji sebentar adalah 100% mendapatkan nilai D&E, untuk mahasiswa dengan waktu mengaji sedang dengan nilai A ada sebanyak 21 orang atau 23,86% ,yang mendapatkan nilai B&C ada sebanyak 65 orang atau 73,86%, dan mahasiswa yang mendapatkan nilai D&E ada 2 orang atau 0,23%. Sedanagkan persentse untuk waktu mengaji lama adalah 100% pada nilai A.

Generalized Ordinal Logistic Regression Model

Berdasarkan uraian pada bagian sebelumnya, data pengamatan yang akan diteliti adalah data nilai pesantren mahasiswa baru dengan variable prediktornya yaitu jenis kelamin (X_1) dan lama mengaji per hari (X_2). Penaksiran parameter menggunakan metode maksimum *likelihood* dan metode iterasi Newton-Raphson.

Perhitungan *Generalized Ordinal Logistic Regression* dengan bantuan *software* stata program gologit2 dapat dilihat hasilnya menggunakan tabel sebagai berikut.

Tabel 5. Taksiran Parameter untuk *Generalized Ordinal Logistic Regression Model* menggunakan stata gologit2

model	Variabel	Parameter	Taksiran	P-Value
$ln\left(\frac{\pi_1(x)}{1-\pi_1(x)}\right)$	Intercept	β_{01}	-9.7394	0,000
	Jenis Kelamin	β_{11}	-2,1777	1,000
	Lama mengaji per hari	β_{21}	0,8394	0,000
$ln\left(\frac{\pi_2(x)}{1-\pi_2(x)}\right)$	Intercept	β_{02}	-22.0492	0,020
	Jenis Kelamin	β_{12}	-1,2527	0,715
	Lama mengaji per hari	β_{22}	0,9145	0,001

Sumber: Hasil pengolahan software Stata

Berdasarkan hasil perhitungan yang tersaji pada table 5 dapat diperoleh nilai-nilai taksiran parameter untuk *Generalized Ordinal Logistic Regression Model*. Untuk model logit pertama nilai taksiran parameter *intercept* adalah -9.7394, nilai taksiran parameter koefisien jenis kelamin (X_1) adalah -2,1777 dan nilai taksiran parameter koefisien lama mengaji per hari (X_2) adalah 0,8394.

Untuk model logit kedua nilai taksiran parameter *intercept* adalah -22.0492, nilai taksiran parameter koefisien jenis kelamin (X_1) adalah -1,2527 dan nilai taksiran parameter koefisien lama mengaji per hari (X_2) adalah 0,9145. Model logit untuk kategori pertama dan kedua dapat ditulis sebagai berikut:

$$\ln(Y_1) = ln\left(\frac{\pi_1(x)}{1-\pi_1(x)}\right) = -9.7394 - 2,1777x_1 + 0,8394x_2$$

$$\ln(Y_2) = ln\left(\frac{\pi_2(x)}{1-\pi_2(x)}\right) = -22.0492 - 1,2527x_1 + 0,9145x_2$$

Persamaan untuk memperoleh nilai-nilai peluang kumulatif yaitu:

$$\pi_1(x) = P(Y_1 \leq 1|x) = \frac{1}{1 + \exp(-9.7394 - 2{,}1777x_1 + 0{,}8394x_2)}$$

$$\pi_2(x) = P(Y_2 \leq 2|x) = \frac{1}{1 + \exp(-22.0492 - 1{,}2527x_1 + 0{,}9145x_2)}$$

Sehingga peluang untuk masing-masing kategori nilai adalah sebagai berikut:

$$P(Y_1 = 1|x) = \frac{1}{1 + \exp(-9.7394 - 2{,}1777x_1 + 0{,}8394x_2)}$$

$$P(Y_2 = 2|x) = \frac{1}{1 + \exp(-22.0492 - 1{,}2527x_1 + 0{,}9145x_2)} - \frac{1}{1 + \exp(-9.7394 - 2{,}1777x_1 + 0{,}8394x_2)}$$

$$P(Y_3 = 2|x) = 1 - \frac{1}{1 + \exp(-22.0492 - 1{,}2527x_1 + 0{,}9145x_2)}$$

Pengujian Parameter *Generalized Ordinal Logistic Regression Model*

Selanjutnya akan dilakukan pengujian parameter secara serentak dengan menggunakan uji Rasio *Likelihood*, dengan hipotesis sebagai berikut:

H_0: $\beta_1 = \beta_2 = ... = \beta_j = 0$

H_1: minimal ada satu $\beta_j \neq 0$; j = 1,2,...,k dimana k adalah jumlah variabel prediktor.

Hasil perhitungan uji Rasio *Likelihood* berdasarkan output *software* stata gologit2 dapat dilihat pada tabel 6.

Tabel 6. Output Rasio Likelihood Generalized Ordinal Logistic Regression Model

Pengujian	Chi-Kuadrat	db	P-value
Rasio Kemungkinan	394,59	4	0,000

Sumber: Hasil Pengolahan Software Stata Gologit2

Berdasarkan tabel 6. dapat dilihat bahwa nilai p-value adalah 0,000. Kriteria pengujiannya adalah tolak H_0 jika nilai p-value kurang dari α yang dipiilih dimana nilai α = 0,05. Karena nilai p-value lebih kecil dari α = 0,05 maka H_0 ditolak artinya parameter regresi logistik berpengaruh secara signifikan dalam model.

Pengujian signifikansi parameter secara parsial dapat dilihat dari penaksiran parameter hasil perhitungan stata gologit2 yang disajikan pada tabel 7.

Tabel 7. Uji Keberartian Parameter untuk Generalized Ordnial Logistic Regression Model

Model	Variabel	Parameter	Taksiran	P-Value
$ln\left(\frac{\pi_1(x)}{1-\pi_1(x)}\right)$	Intercept	β_{10}	-9.7394	0,000*
	Jenis Kelamin	β_{11}	-2,1777	1,000
	Lama Mengaji	β_{12}	0,8394	0,000*
$ln\left(\frac{\pi_2(x)}{1-\pi_2(x)}\right)$	Intercept	β_{20}	-22.0492	0,020*
	Jenis Kelamin	β_{21}	-1,2527	0,715
	Lama Mengaji	β_{22}	0,9145	0,001*

*)signifikan pada α = 5%

Berdasarkan hasil perhitungan yang tersaji pada tabel 7 diperoleh nilai-nilai taksiran parameter untuk *Generalized Ordinal Logistic Regression Model*. Dapat dilihat bahwa variabel jenis kelamin pada kedua model logit tidak bepengaruh secara signifikan. Sedangkan *intercept* dan variabel lama mengaji berpengaruh secara signifikan terhadap nilai ujian pesantren mahasiswa baru.

Interpretasi Generalized Ordinal Logistic Regression Model

Model *Generalized Ordnial Logistic Regression Model* dapat diinterpretasikan menggunakan odds rasio, hasil perhitungan dari odds rasio dapat dilihat pada tabel 8.

Tabel 8. Hasil Generalized Ordinal Logistic Regression Model menggunakan stata gologit2 untuk model logit pertama dan model logit kedua

	Y > 1 vs Y≤ 1		**Y > 2 vs Y≤ 2**	
Variabel	b	OR	b	OR
Jenis Kelamin	-2,1777	0,1133	-1,2527	0,2857
Lama mengaji per hari	0,8394	2,315	0,9145	2,4955

Sumber: Hasil Pengolahan Software Stata Gologit2

Pada tabel 8 variabel Y terbagi menjadi tiga kategori yaitu Y=1 adalah untuk nilai D&E, Y = 2 untuk nilai B&C, dan Y = 3 untuk nilai A. Untuk variabel jenis kelamin pada model yang pertama yaitu jenis kelamin laki-laki (X_1=1) dan jenis kelamin perempuan (X_2=0) nilai odds rasionya adalah 0,1133 atau 1/9, maka mahasiswa dengan jenis kelamin perempuan akan memperoleh nilai D&E 9 kali dibandingkan mahasiswa laki-laki, untuk model yang kedua dengan variabel jenis kelamin nilai odds rasio sebesar 0,2857 atau 1/3, maka mahasiswa dengan jenis kelamin perempuan akan memperoleh nilai B&C 3 kali dibandingkan dengan mahasiswa laki-laki.

Sedangkan untuk variabel lama mengaji nilai odds rasionya 2,315 untuk model pertma, waktu lama mengaji dibagi kedalam tiga kelompok yaitu sebentar, sedang dan lama. Untuk waktu mengaji sebentar dibanding waktu mengaji lainnya (waktu mengaji sedang dan waktu mengaji lama) maka mahasiswa dengan waktu mengaji sebentar akan memperoleh nilai D&E lebih besar 2,315 kali dibandingkan dengan waktu mengaji lainnya. Untuk model yang kedua nilai odds rasio adalah 2,4955, untuk waktu mengaji sedang dibanding lainnya (waktu mengaji sebentar dan waktu mengaji lama) maka mahasiswa dengan waktu mengaji sedang akan memperoleh nilai B&C lebih besar 2,315 kali dibandingkan waktu lainnya.

Selanjutnya akan dilakukan perhitungan nilai peluang pada deskripsi data dengan lama mengaji yang dibagi kedalam tiga kelompok yaitu sebentar (1 – 14 menit), sedang (15 – 30 menit), dan lama (31 – 150).

Berikut merupakan perhitungan peluang untuk lama mengaji sedang, misalkan diambil 20 menit untuk waktu mengaji dan jenis kelamin laki-laki:

$$\ln(Y_1) = ln\left(\frac{\pi_1(x)}{1-\pi_1(x)}\right) = -9.7394 - 2{,}1777(1) + 0{,}8394(20) = 4{,}8709$$

$$\ln(Y_2) = ln\left(\frac{\pi_2(x)}{1-\pi_2(x)}\right) = -22.0492 - 1{,}2527(1) + 0{,}9145(20) = -5{,}0119$$

Sehingga diperoleh nilai-nilai peluangnya adalah sebagai berikut:

$$P(Y_1 = 1|x) = \frac{1}{1+\exp(4{,}8709)} = 0{,}0076$$

$$P(Y_2 = 2|x) = \frac{1}{1+\exp(-5{,}0119)} - \frac{1}{1+\exp(4{,}8709)} = 0{,}9858$$

$$P(Y_3 = 2|x) = 1 - \frac{1}{1+\exp(-5{,}0119)} = 0{,}0066$$

Sehingga dapat disimpulkan bahwa seorang mahasiswa yang berjenis kelamin laki-laki dan lama mengaji per hari yaitu 20 menit, maka akan mendapatkan nilai B atau C karena memiliki peluang yang paling besar.

D. Kesimpulan dan Saran

Kesimpulan

Penelitian ini memodelkan data dengan menggunakan *Generalized Ordinal Logistic Regression Model*. Dimana Pada penelitian ini variabel respon berupa nilai ujian pesantren mahasiswa baru FMIPA Unisba tahun 2017, variabel prediktornya yaitu jenis kelamin (X_1) dan lama mengaji per hari (X_2). Model *Generalized Ordinal Logistic Regression* yang terbentuk adalah:

$$\ln(Y_1) = ln\left(\frac{\pi_1(x)}{1-\pi_1(x)}\right) = -9.7394 - 2{,}1777x_1 + 0{,}8394x_2$$

$$\ln(Y_2) = ln\left(\frac{\pi_2(x)}{1-\pi_2(x)}\right) = -22.0492 - 1{,}2527x_1 + 0{,}9145x_2$$

Pengujian Parameter secara serentak menggunakan uji Rasio *Likelihood* dimana hasil perhitungan menunjukan parameter regresi logit berpengaruh secara signifikan dalam model. Pengujian secara parsial menunjukan bahwa variabel jenis kelamin (X_1) tidak berpengaruh secara signifikan, sedangkan variabel lama mengaji berpengaruh secara signifikan terhadap nilai ujian pesantren mahasiswa baru.

Saran

Saran yang dapat dikemukakan dalam skripsi ini adalah Bagi yang akan menganalisis masalah yang sama sebaiknya menambahkan variabel prediktor selain jenis kelamin dan lama mengaji. *Generalized Ordinal Logistic Regression Model* selain dengan stata gologit2 dapat juga menggunakan aplikasi *software* yang lain seperti SAS

Daftar Pustaka

Agresti, A. (2007), *An Introduction to Categorical Data Analysis*. New York: John Wiley and Sons.

Czepiel, S. A. (2011). Maximum likelihood estimation of logistic regression models: Theory and Implementation, (Online). (http://czep.net/contact.html , diakses pada tanggal 24 Juli 2017)

Hajarisman, N. (2009). *Analisis Data Kategorik*. Program Studi Statistika Universitas Islam Bandung.

Hosmer, D.W., dan Lemeshow, S. (2000). *Applied Logistic Regresion*. Second Edition. New York: John Wiley and Sons, Inc.

Liu, X., dan Koirala, H. (2012). Ordinal Regression Analysis: Using Generalized Ordinal Logistic Regression Models to Estimate Educational Data. *Journal of Modern Applied Statistical Methods*, Vol.11:Iss.1, Article 21.

McCullagh, P. (1980). Regression Models for Ordinal Data (with discussion), *Journal Royl Statistic Society*, B(42):109-142.

Williams, R. A. (2005). Gologit2: A program for Generalized Logistic Regression/partial proportional Odds Models for Ordinal Variable, (Online). (http://www.nd.edu/~rwilliam/stata/gologit2.pdf , diakses pada tanggal 1 Juli 2017)