

Penerapan Distribusi *Zero Inflated Binomial* untuk Mengatasi Masalah Overdispersi dalam Data Status Kelulusan Mata Kuliah Pada Mahasiswa FMIPA UNISBA Tahun Akademik 2016/2017

Application Of The Zero Inflated Binomial Distribution To Overcome The Problem Of Overdispersion In The Data On The Graduate Status Of The Courses In The Student Of FMIPA UNISBA 2016/2017 Academic Year

¹Sintia Sri Novianti,²Nusar Hajarisman

^{1,2}*Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Jl. Tamansari No.1 Bandung 40116*

Email: ¹sintiasri16.ssn@gmail.com, ²nusarhajarisman@yahoo.com

Abstract. Binary data is the data which values are categorized only with the symbol 1 or 0. In general, binary data will be binomial distributed. Binomial distribution has several assumptions that must be fulfilled, namely random variables that are mutually independent, and the chance of success or constant for each experiment. But in reality often these assumptions are violated, such as the correlation between observations. Which will produce a variance value that is greater than the variance under binomial sampling. Where this is an indication of an overdispersion problem. Overdispersion will be result in invalid variance estimation. Then the solution is using the Zero Inflated Binomial distribution, which is used in the data that contains a lot of zero and has overdispersion problems. In this thesis, will discuss the use of Zero Inflated Binomial distribution on the data of Course Graduation Status in FMIPA UNISBA Students 2016/2017 Academic Year. Parameter estimation is performed using the maximum likelihood method. And calculating the chi-square value is done to see if there is oversids. Estimating parameters by using the distribution Zero Inflated Binomial explains that the distribution is suitable for overcoming the problem of overdispersion in the data on the graduation status of the subjects at the FMIPA UNISBA 2016/2017 Academic Year students.

Keywords: Binary data, Distribusi Binomial Overdispersion, Zero Inflated Binomial Distribution, Chi-Kuadrat

Abstrak. Data biner adalah data yang nilainya dikategorikan hanya dengan simbol 1 atau 0. Pada umumnya data biner akan berdistribusi binomial. Distribusi binomial memiliki beberapa asumsi yang harus dipenuhi yaitu variabel acak yang saling bebas, dan peluang sukses sama atau konstan untuk setiap percobaan. Namun pada kenyataannya seringkali asumsi tersebut dilanggar, seperti adanya korelasi antar pengamatan. Yang mana akan menghasilkan nilai varians yang lebih besar dari varians dibawah sampling binomial. Dimana hal tersebut merupakan indikasi adanya masalah overdispersi. Overdispersi akan mengakibatkan hasil dari penaksiran varians tidak valid. Maka solusi yang digunakan adalah dengan menggunakan distribusi *Zero Inflated Binomial*, yang mana distribusi ini digunakan pada data yang berisi banyak nol dan mengalami masalah overdispersi. Dalam skripsi ini, akan membahas penggunaan distribusi *Zero Inflated Binomial* pada data Status Kelulusan Mata Kuliah Pada Mahasiswa FMIPA UNISBA Tahun Akademik 2016/2017. Penaksiran parameter dilakukan menggunakan metode kemungkinan maksimum. Dan dilakukan perhitungan nilai chi-kuadrat untuk melihat adanya oversidspersi. Penaksiran parameter dengan menggunakan distribusi *Zero Inflated Binomial* menjelaskan bahwa distribusi tersebut cocok digunakan untuk mengatasi masalah overdispersi pada data status kelulusan mata kuliah pada mahasiswa FMIPA UNISBA Tahun Akademik 2016/2017.

Kata kunci: Data biner, Distribusi Binomial, Overdispersi, Distribusi *Zero Inflated Binomial*, Chi-Kuadrat

A. Pendahuluan

Data yang biasa dijumpai dalam analisis statistika adalah berupa data biner. Data biner adalah data yang nilainya dikategorikan hanya dengan simbol 1 atau 0. Data biner biasanya akan mengikuti distribusi Bernoulli ketika hanya dilakukan dalam 1 kali percobaan. Tetapi jika percobaan diulang sebanyak n kali maka akan menjadi distribusi Binomial.

Distribusi Binomial memiliki beberapa asumsi yang harus dipenuhi, yaitu variabel acak yang saling bebas, dan peluang sukses sama/konstan untuk setiap pengamatan. Namun jika asumsi tidak terpenuhi, maka akan mengalami overdispersi. Akibatnya, data biasanya menunjukkan varians yang lebih besar daripada varians yang diizinkan oleh model Binomial.

Berdasarkan hal tersebut, maka perlu dicari suatu metode statistika yang tepat untuk dapat digunakan dalam menentukan taksiran peluang bagi parameter, dimana ada masalah overdispersi didalamnya. Salah satu cara untuk mengatasi masalah overdispersi dalam data berdistribusi Binomial adalah menggunakan distribusi "*Zero Inflated Binomial*". Pemilihan distribusi ini didasarkan pada karakteristik data yang mana digunakan pada saat data berisi banyak nilai 0. Data yang akan digunakan pada skripsi ini adalah data jumlah mata kuliah yang tidak lulus pada mahasiswa Tahun Akademik 2016/2017 Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung.

Berdasarkan latar belakang yang telah diuraikan, maka perumusan masalah dalam makalah ini adalah:

1. Bagaimana mendeteksi masalah overdispersi dalam data status kelulusan mata kuliah pada mahasiswa FMIPA UNISBA Tahun Akademik 2016/2017 dengan menggunakan distribusi Binomial
2. Bagaimana mengatasi masalah overdispersi dalam dalam data status kelulusan mata kuliah pada mahasiswa FMIPA UNISBA Tahun Akademik 2016/2017 dengan menggunakan distribusi *Zero Inflated Binomial*

B. Landasan Teori

Distribusi Binomial

Misalkan x_1, x_2, \dots, x_n menyatakan respons untuk n buah percobaan yang identik dan saling bebas sedemikian rupa sehingga :

$$P(x_i = 1) = \pi \text{ dan } P(x_i = 0) = 1 - \pi$$

Di sini digunakan label yang sangat umum untuk menyatakan angka '1' sebagai peristiwa 'sukses' dan '0' untuk menyatakan peristiwa 'gagal'. Percobaan yang identik mempunyai makna bahwa peluang sukses π adalah sama untuk setiap percobaan. Sedangkan percobaan yang saling bebas mempunyai makna bahwa $\{x_i\}$ merupakan variabel acak yang saling bebas.

Fungsi massa peluang untuk variabel acak x_1, x_2, \dots, x_n yang berdistribusi Binomial adalah sebagai berikut :

$$f(x) = \binom{n}{x} \pi^x (1 - \pi)^{n-x}, \text{ untuk } x = 0, 1, 2, \dots, n \quad (1)$$

Dimana koefisien Binomial $\binom{n}{x} = \frac{n!}{x!(n-x)!}$

Distribusi Binomial untuk $X = \sum_{i=1}^n x_i$ mempunyai rata-rata dan varians sebagai berikut:

$$E(X) = n \pi \quad (2)$$

dan

$$Var(X) = n\pi(1 - \pi) \quad (3)$$

Penulisan notasi dari peubah acak X yang berdistribusi Binomial adalah $B(X; n, \pi)$, artinya peubah acak X berdistribusi Binomial dengan banyak pengulangan eksperimen sebanyak n kali, peluang terjadi peristiwa sukses sebesar π , dan banyak peristiwa sukses terjadi ada sebanyak x .

Metode Kemungkinan Maksimum Untuk Parameter Binomial

Bagian dari fungsi kemungkinan maksimum yang menyangkut parameter disebut kernel. Oleh karena memaksimumkan fungsi kemungkinan terhadap parameter, maka bagian lainnya menjadi tidak penting untuk dibahas.

Untuk mengilustrasikan hal tersebut, perhatikan distribusi Binomial dalam pers.(2.4). Dimana koefisien $\binom{n}{x}$ tidak berpengaruh pada maksimum terhadap parameter π . Jadi, untuk selanjutnya koefisien Binomial dapat diabaikan, dan memperoleh kernel sebagai fungsi kemungkinan. Fungsi kemungkinan Binomial adalah sebagai berikut:

$$L(\pi) = \prod_{i=1}^n [\pi^{x_i} (1 - \pi)^{n - x_i}] = \pi^{\sum x_i} (1 - \pi)^{n^2 - \sum x_i} \quad (4)$$

Selanjutnya fungsi log-kemungkinan Binomial adalah sebagai berikut :

$$l(\pi) = \ln[\pi^{\sum x_i} (1 - \pi)^{n^2 - \sum x_i}] = \sum x_i \ln(\pi) + (n^2 - \sum x_i) \ln(1 - \pi) \quad (5)$$

Kemudian, dengan menurunkan terhadap π akan menghasilkan :

$$\frac{\partial l(\pi)}{\partial \pi} = \frac{\sum x_i}{\pi} - \frac{(n^2 - \sum x_i)}{(1 - \pi)} = \frac{(\sum x_i - n^2 \pi)}{\pi(1 - \pi)} \quad (6)$$

Bentuk di atas disamakan dengan 0, maka akan memberikan solusi dari persamaan kemungkinan sebagai $\hat{\pi} = \frac{\bar{x}}{n}$, proporsi sukses sampel dari n buah percobaan.

Overdispersi

Fenomena overdispersi terjadi jika nilai varians yang dihasilkan lebih dari varians pada distribusi Binomial yang berlaku atau $Var(x) > n\pi(1 - \pi)$. Overdispersi terjadi pada data yang memiliki korelasi antar pengamatan, atau tidak adanya kebebasan antar pengamatan atau variabel ataupun juga jika terdapat peluang yang berbeda satu sama lain dalam 1 kelompok.

Pada kasus Binomial dengan overdispersi dapat ditandai dalam dua momen pertama. Misalkan X adalah variabel acak Binomial yang mewakili jumlah keberhasilan dari uji coba sebanyak n. Kemudian,

$$E(X) = n\pi \quad (7)$$

dan

$$Var(X) = n(1 - \pi)\{1 + \rho^2(n - 1)\} \quad (8)$$

dimana π mewakili probabilitas keberhasilan setiap percobaan Bernoulli.

Perhatikan bahwa nilai harapan atau rata-rata dari X cocok dengan model Binomial, akan tetapi varians yang diberikan dalam persamaan (8) lebih besar daripada varians yang dimodelkan oleh model Binomial karena faktor $\{1 + \rho^2(n - 1)\}$ lebih besar dari satu. Variabel acak X dengan mean dan varians, seperti pada persamaan (7) dan (8), dikatakan memiliki overdispersi atau variasi ekstra relatif terhadap model Binomial.

Chi-kuadrat Pearson

Terdapat sejumlah alternatif untuk mengukur kecocokan model, dan salah satu yang paling populer adalah statistik chi-kuadrat pearson yang didefinisikan sebagai berikut:

$$\chi^2 = \sum_{i=0}^n \frac{(O_i - E_i)^2}{E_i} \tag{9}$$

Dimana O_i adalah nilai observasi ke i dan E_i adalah frekuensi harapan dari observasi ke i . Pada saat nilai χ^2 dibagi dengan derajat bebasnya mempunyai nilai yang lebih besar dari 1, maka akan berarti bahwa distribusi tersebut tidak cocok untuk data dan data memiliki overdispersi. Dengan demikian, dapat dikatakan bahwa statistik χ^2 ini digunakan untuk mengukur kecocokan distribusi dan mendeteksi adanya overdispersi.

Distribusi Zero Inflated Binomial

Distribusi *Zero Inflated Binomial* merupakan distribusi yang digunakan untuk data yang berisi banyak 0 dan mengalami overdispersi, juga jika data mengikuti distribusi Binomial. Distribusi diskrit yang mendasarinya adalah Binomial($\pi; n$) pada (1), yang diperbaharui menjadi distribusi *Zero-inflated Binomial* (ZIB) dengan fungsi probabilitas $Pr(X = x) = \omega \binom{n}{x} \pi^x (1 - \pi)^{n-x} + (1 - \omega)I(x = 0)$, $x = 0, 1, 2, \dots, n$ (10)

dimana $I(\cdot)$ adalah fungsi indikator atau fungsi penghubung. Perhatikan bahwa distribusi ZIB adalah gabungan dari dua distribusi, yaitu Binomial ($\pi; n$) dengan probabilitas " ω ", dan distribusi indikator dengan probabilitas " $1-\omega$ ". Di bawah parameterisasi $\pi = P + \theta(1 - P)$ dan $\omega = \frac{P}{\pi}$ dengan $0 \leq \theta \leq 1$, fungsi probabilitas (10) dapat ditulis sebagai

$$Pr(X = x) = \frac{P}{P + \theta(1 - P)} Pr(X = x) + \frac{\theta(1 - P)}{P + \theta(1 - P)} I(x = 0), \quad x = 0, 1, \dots, n \tag{11}$$

dimana $X \sim$ Binomial $\{P + \theta(1 - P); n\}$. Bila $\theta = 0$, persamaan (11) akan menjadi distribusi Binomial ($P; n$). Di sisi lain, kasus $\theta = 1$ akan menimbulkan distribusi Zero Degenerate. Jika X memiliki fungsi probabilitas seperti pada (11), X akan didistribusikan sebagai distribusi ZIB($P, \theta; n$). Dapat ditunjukkan bahwa dua momen pertama dari distribusi ZIB adalah

$$E(X) = nP \tag{12}$$

dan

$$Var(X) = nP(1 - P)\{1 + \theta(n - 1)\} \tag{13}$$

Dengan demikian, distribusi *Zero-inflated Binomial* memiliki overdispersi yang relatif terhadap distribusi Binomial. Perhatikan bahwa parameter overdispersi θ pada (13) mewakili parameter ρ^2 pada persamaan (8).

Deskripsi Data

Data yang diperoleh adalah data hasil penilaian mata kuliah pada semester 1 dan 2 mahasiswa FMIPA Universitas Islam Bandung dengan jumlah mahasiswa sebanyak 267 orang. Presentase jumlah mata kuliah yang tidak lulus pada mahasiswa Fakultas MIPA dapat di lihat pada Tabel 1.

Tabel 1. Persentase Jumlah Mata Kuliah Yang Tidak Lulus Pada Semester 1 dan Semester 2 Mahasiswa FMIPA UNISBA Tahun Akademik 2016/2017

Jumlah Mata kuliah yang tidak lulus	0	1	2	≥ 3
Jumlah Mahasiswa	234	26	6	1
Persentase (%)	87,64	9,74	2,25	0,37

Dari Tabel 1 terlihat bahwa dari total 267 mahasiswa fakultas MIPA pada semester 1 dan 2 Tahun Akademik 2016/2017 terdapat 87,64% atau setara dengan 234

mahasiswa yang tidak memiliki mata kuliah yang tidak lulus atau belum tuntas; 9,74% atau setara dengan 26 mahasiswa memiliki 1 mata kuliah yang tidak lulus; 2,25% mahasiswa yang memiliki 2 mata kuliah yang tidak lulus; dan 0,37% mahasiswa yang memiliki lebih dari atau sama dengan 3 mata kuliah yang tidak lulus.

Hasil Penaksiran Parameter Dengan Distribusi Binomial

Penaksiran parameter dengan menggunakan distribusi Binomial diperoleh dari hasil perhitungan yang akan disajikan pada Tabel 2. Dimana perhitungan dilakukan menggunakan program SAS.

Tabel 2. Hasil Penaksiran Parameter Distribusi Binomial

	π
Penaksir	0.05119
Galat baku taksiran	0.007787
Db	4
t value	6,57
Pr > t 	0,0028
Batas Bawah	0,02957
Batas Atas	0,07281
-2 log likelihood	253,3

Dari hasil di atas didapat nilai taksiran dari π (peluang mata kuliah yang tidak lulus pada semester 1 dan semester 2 mahasiswa Tahun Akademik 2016/2017 FMIPA UNISBA) sebesar 0,05119. Besarnya nilai galat baku taksiran, yang digunakan untuk pengujian hipotesis, untuk π sebesar 0,007787, dan nilai t value sebesar 6,57. Dengan nilai t value sebesar 6,57 didapat nilai Pr > |t| sebesar 0,0028. Adapun nilai bagi selang kepercayaan dengan batas bawah untuk π sebesar 0,02957 dan nilai batas atas sebesar 0,07281.

Hasil Mendeteksi Adanya Overdispersi

Untuk memeriksa adanya overdispersi dalam data jumlah mata kuliah yang tidak lulus pada semester 1 dan semester 2 mahasiswa Tahun Akademik 2016/2017 Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Bandung dengan menggunakan distribusi Binomial, maka akan dihitung nilai chi kuadrat dengan menggunakan persamaan (9). Sehingga didapat nilai frekuensi harapan dan chi-kuadrat seperti yang disajikan pada Tabel 3.

Tabel 3. Nilai Frekuensi Harapan dan Chi-Kuadrat dari Distribusi Binomial

x	O_i	$\hat{\pi}(X = x)$	$E_i = n \cdot \hat{\pi}(X = x)$
0	234	0,8764	228,0628
1	26	0,0959	36,91017
2	6	0,0255	1,991207
3	1	0,0023	0,035807
χ^2		37,41362	
Db		2	
χ^2 / db		18,70681	

Berdasarkan Tabel 4.3. dapat dilihat pada distribusi Binomial dengan jumlah observasi ketika $t = 0$ adalah sebesar 234 dan nilai peluang (π) sebesar 0,8764 akan menghasilkan nilai frekuensi harapan (E_i) sebesar 228,0628. Yang artinya dari 234 mahasiswa FMIPA UNISBA Tahun Akademik 2016/2017 yang diketahui memperoleh nilai lulus semua pada mata kuliah semester 1 dan semester 2, dengan menggunakan distribusi Binomial dengan nilai peluang sebesar 0,8764 ditaksir ada 228,0628 mahasiswa FMIPA UNISBA Tahun Akademik 2016/2017 memperoleh nilai lulus semua pada mata kuliah semester 1 dan semester 2.

Didapat nilai $\chi^2_{(2)}$ dari distribusi Binomial sebesar 18,70681 yang berarti terdapat overdispersi pada data karena nilai $\chi^2_{(2)} > 1$. Sehingga distribusi Binomial tidak cocok untuk data jumlah mata kuliah yang tidak lulus pada semester 1 dan semester 2 Mahasiswa FMIPA UNISBA Tahun Akademik 2016/2017.

Hasil Penaksiran Parameter Dengan Distribusi *Zero Inflated Binomial*

Penaksiran parameter dengan menggunakan distribusi *Zero Inflated Binomial* diperoleh dari hasil perhitungan yang akan disajikan pada Tabel 4. Dimana perhitungan dilakukan menggunakan program SAS.

Tabel 4. Hasil Penaksiran Parameter Distribusi *Zero Inflated Binomial*

	P	θ
Penaksir	0.05119	0.1672
Galat baku taksiran	0.008995	0.07620
Db	4	4
t value	5,69	5,37
Pr > t 	0,0047	0,0058
Batas Bawah	0.02621	0.1973
Batas Atas	0.07616	0,6204
-2 log likelihood	239,9	

Dari hasil di atas didapat nilai taksiran dari P (peluang mata kuliah yang tidak lulus pada semester 1 dan semester 2 mahasiswa Tahun Akademik 2016/2017 FMIPA UNISBA) sebesar 0,05119 dan nilai penaksir untuk θ (korelasi antara satu mata kuliah dengan mata kuliah yang lain pada seorang mahasiswa) pada distribusi *Zero Inflated Binomial* sebesar 0,1672. Besarnya nilai galat baku taksiran, yang digunakan untuk pengujian hipotesis, untuk P sebesar 0,008995, dan nilai galat baku taksiran untuk θ pada distribusi *Zero Inflated Binomial* sebesar 0,07620.

Didapat pula nilai t value sebesar 5,69 dan $Pr > |t|$ sebesar 0,0047 untuk P , dan nilai t value sebesar 5,37 dan $Pr > |t|$ sebesar 0,058 untuk θ pada distribusi *Zero Inflated Binomial*. Nilai-nilai tersebut dapat digunakan untuk pengujian hipotesis bagi parameter-parameter π dan θ . Adapun nilai bagi selang kepercayaan dengan batas bawah untuk P sebesar 0,02621 dan nilai batas atas sebesar 0,07616. Dan Nilai batas bawah untuk θ pada distribusi *Zero Inflated Binomial* sebesar 0,01973 dan nilai batas atas sebesar 0,6204.

Setelah dilakukan penaksiran dengan menggunakan distribusi Binomial dan distribusi *Zero Inflated Binomial*, didapat nilai penaksir peluang baik menggunakan distribusi Binomial maupun distribusi *Zero Inflated Binomial* sebesar 0,5119. Karena besarnya nilai penaksir peluang bagi kedua distribusi sama besar, maka untuk melihat adanya perbedaan antara distribusi Binomial dan distribusi *Zero Inflated Binomial* akan dilihat perbandingan antara nilai selisih dari “-2 log likelihood” antara kedua distribusi dengan nilai $\chi^2_{1;0,05}$. Sehingga diperoleh hasil sebagai berikut :

$$D = |(-2 \log \text{likelihood}(\text{Binomial})) - (-2 \log \text{likelihood}(\text{Zero Inflated Binomial}))|$$

$$D = 253,3 - 239,9 = 13,4$$

Dengan nilai $\chi^2_{1;0,05} = 3,84146$, maka terlihat bahwa nilai $D > \chi^2_{1;0,05}$. Hal ini menunjukkan bahwa terdapat perbedaan antara distribusi Binomial dan distribusi *Zero Inflated Binomial*.

Pengujian Hipotesis bagi Parameter P dan θ

Terdapat 2 buah parameter yang didapat pada distribusi *Zero Inflated Binomial*, yaitu P dan θ . Akan dilakukan pengujian apakah parameter-parameter tersebut signifikan atau tidak signifikan secara statistik dengan menggunakan uji t .

a. Untuk P pada distribusi *Zero Inflated Binomial* dengan Hipotesis sebagai berikut:

$H_0: P = 0$; peluang seorang mahasiswa Tahun Akademik 2016/2017 FMIPA UNISBA memperoleh nilai mata kuliah tidak lulus pada semester 1 dan 2 dari distribusi *Zero Inflated Binomial* sama dengan nol

$H_1: P \neq 0$; peluang seorang mahasiswa Tahun Akademik 2016/2017 FMIPA UNISBA memperoleh nilai mata kuliah tidak lulus pada semester 1 dan 2 dari distribusi *Zero Inflated Binomial* tidak sama dengan nol

Dengan statistika uji pada $\alpha = 0,05$ dengan nilai $t = \frac{\text{nilai penaksir } P}{\text{Galat baku taksiran}} = \frac{0,05119}{0,008995} = 5,69$ dan diperoleh nilai $Pr > |t| = 0,0047$.

Dimana kriteria uji: tolak H_0 jika nilai “ $Pr > |t|$ ” $< \alpha$.

Sehingga didapat kesimpulan, H_0 ditolak karena $0,0047 < \alpha$, yang artinya bahwa peluang seorang mahasiswa FMIPA UNISBA Tahun Akademik 2016/2017 memperoleh nilai mata kuliah tidak lulus pada semester 1 dan 2 dari distribusi *Zero Inflated Binomial* tidak sama dengan nol dengan $\alpha = 5\%$.

b. Untuk θ pada distribusi *Zero Inflated Binomial* dengan Hipotesis sebagai berikut:

$H_0: \theta = 0$, tidak ada hubungan antara satu mata kuliah dengan mata kuliah yang lain dari seorang mahasiswa FMIPA UNISBA Tahun Akademik 2016/2017

$H_1: \theta \neq 0$, ada hubungan antara satu mata kuliah dengan mata kuliah yang lain dari seorang mahasiswa FMIPA UNISBA Tahun Akademik 2016/2017

Dengan statistika uji pada α sebesar 0,10 dengan nilai $t = \frac{\text{nilai penaksir } \theta}{\text{Galat baku taksiran}} = \frac{0,1672}{0,07620} = 5,37$ dan diperoleh nilai $\text{Pr} > |t|$ sebesar 0,0058.

Dimana kriteria uji: tolak H_0 jika nilai " $\text{Pr} > |t|$ " $< \alpha$.

Sehingga didapat kesimpulan, H_0 ditolak karena $0,0058 < \alpha$, yang artinya ada hubungan antara satu mata kuliah dengan mata kuliah yang lain dari seorang mahasiswa FMIPA UNISBA Tahun Akademik 2016/2017 dengan $\alpha = 10\%$.

Hasil Mendeteksi Overdispersi Pada Distribusi *Zero Inflated Binomial*

Untuk memastikan apakah dengan menggunakan distribusi *Zero Inflated Binomial* data jumlah mata kuliah yang tidak lulus pada semester 1 dan semester 2 mahasiswa FMIPA UNISBA Tahun Akademik 2016/2017 sudah tidak memiliki overdispersi, maka akan dihitung nilai chi kuadrat dengan menggunakan persamaan (9). Sehingga didapat nilai Frekuensi harapan dan chi-kuadrat seperti yang disajikan pada Tabel 5.

Tabel 5. Nilai Frekuensi Harapan dan Chi-Kuadrat dari Distribusi *Zero Inflated Binomial*

X	O_i	$\hat{\pi}(X = x)$	$E_i = n \cdot \hat{\pi}(X = x)$
0	234	0,8542	234
1	26	0,1382	25,60147
2	6	0,00746	6,797038
3	1	0,00013	0,601524
χ^2			0,363634
db			1
χ^2 / db			0,363634

Dari Tabel di atas dapat dilihat dengan jumlah observasi ketika $x = 0$ adalah sebesar 234 dan nilai peluang (π) sebesar 0,8542 akan menghasilkan nilai frekuensi harapan (E_i) sebesar 234. Yang artinya dari 234 mahasiswa FMIPA UNISBA Tahun Akademik 2016/2017 yang diketahui memperoleh nilai lulus semua pada mata kuliah semester 1 dan semester 2, dengan menggunakan distribusi Binomial dengan nilai peluang sebesar 0,8542 ditaksir ada 234 mahasiswa FMIPA UNISBA Tahun Akademik 2016/2017 memperoleh nilai lulus semua pada mata kuliah semester 1 dan semester 2.

Dengan menggunakan distribusi *Zero Inflated Binomial* nilai $\chi^2_{(1)}$ menjadi sebesar 0,36 yang mana lebih kecil dari 1, sehingga memiliki arti bahwa dengan menggunakan distribusi *Zero Inflated Binomial* data tidak memiliki overdispersi dan distribusi *Zero Inflated Binomial* lebih cocok untuk digunakan pada data tersebut.

Nilai Peluang Distribusi *Zero Inflated Binomial*

Untuk melihat nilai taksiran peluang dari distribusi *Zero Inflated Binomial*, maka akan disajikan Tabel 6. yang berisi nilai peluang untuk setiap x dengan menggunakan distribusi Binomial dan distribusi *Zero Inflated Binomial*.

Tabel 6. Nilai Peluang Untuk Setiap x dengan Menggunakan Distribusi *Zero Inflated Binomial*

x	<i>Zero Inflated Binomial</i>
0	0,861282441
1	0,090846735
2	0,036760081
≥ 3	0,011110743

Dilihat dari tabel di atas, didapat bahwa pada saat t semakin besar maka nilai peluang baik pada distribusi *Zero Inflated Binomial* semakin kecil. Hal ini menunjukkan bahwa dengan menggunakan distribusi *Zero Inflated Binomial*, peluang “sukses” bahwa mahasiswa FMIPA UNISBA Tahun Akademik 2016/2017 memperoleh nilai mata kuliah tidak lulus akan semakin kecil seiring bertambahnya jumlah mata kuliah yang diambil. Dapat dilihat pula, dengan menggunakan distribusi *Zero Inflated Binomial* peluang mahasiswa FMIPA UNISBA Tahun Akademik 2016/2017 memperoleh 1 nilai mata kuliah yang tidak lulus sebesar 0,090846735.

C. Kesimpulan dan Saran

Kesimpulan

1. Setelah dilakukan penaksiran peluang menggunakan distribusi Binomial didapat taksiran peluang sebesar 0,5119. Tetapi pada saat dilakukan perhitungan nilai χ^2 , didapat nilai $\chi^2_{(2)}$ pada distribusi Binomial sebesar 18,70681 yang berarti terdapat overdispersi pada data dengan distribusi Binomial karena $\chi^2_{(2)} > 1$. Yang artinya terdapat masalah overdispersi pada data status kelulusan mahasiswa FMIPA UNISBA pada semester 1 dan semester 2 Tahun Akademik 2016/2017 yang mengikuti distribusi Binomial, sehingga distribusi ini tidak cocok untuk data tersebut.
2. Setelah menggunakan penaksiran peluang dengan distribusi *Zero Inflated Binomial* didapat nilai taksiran untuk P atau peluang “sukses” sebesar 0,5119 dan nilai taksiran untuk θ sebesar 0,1672. Dilakukan pula pendeteksian overdispersi dengan menghitung nilai χ^2 untuk meyakinkan bahwa distribusi *Zero Inflated Binomial* memang cocok untuk data status kelulusan mahasiswa FMIPA UNISBA pada semester 1 dan semester 2 Tahun Akademik 2016/2017. Didapat nilai $\chi^2_{(1)}$ sebesar 0,36 yang mana lebih kecil dari 1, sehingga tidak terdapat masalah overdispersi pada data status kelulusan mahasiswa FMIPA UNISBA pada semester 1 dan semester 2 Tahun Akademik 2016/2017 yang mengikuti distribusi *Zero Inflated Binomial* dan distribusi *Zero Inflated Binomial* lebih cocok untuk digunakan pada data tersebut. Karena setelah dilakukan penaksiran dengan menggunakan distribusi Binomial dan distribusi *Zero Inflated Binomial*, didapat nilai penaksir peluang sebesar 0,5119. Maka untuk melihat adanya perbedaan antara distribusi Binomial dan distribusi *Zero Inflated Binomial* dihitung nilai perbandingan antara nilai selisih dari “-2 log likelihood” antara kedua distribusi dan didapat bahwa nilai $D > \chi^2_{1;0,05}$. Hal ini menunjukkan bahwa terdapat perbedaan antara distribusi Binomial dan distribusi *Zero Inflated Binomial*.

Saran

1. Selain distribusi *Zero Inflated Binomial*, masih banyak distribusi lain yang dapat digunakan sebagai solusi untuk mengatasi masalah overdispersi dalam data biner seperti distribusi *Random-Clumped Binomial*, distribusi Beta-Binomial, dan distribusi *A Generalized of The Binomial*. Untuk itu para pembaca dapat menggunakan distribusi-distribusi tersebut yang sesuai dengan karakteristik data dalam mengatasi masalah overdispersi.
2. Dalam skripsi ini penulis tidak membahas model, sehingga dapat diimplementasikan kedalam model dan dapat dilihat faktor-faktor apa saja yang berpengaruh terhadap model.

Daftar Pustaka

- Agresti, A. (1990). *Categorical Data Analysis*. New York: John Wiley and Sons.
- Collet, D. (1991). *Modeling Binary Data*. London: Chapman and Hall.
- Hajarisman, N. (1998), Kajian Perbandingan Model Regresi Beta-Binomial dengan Model Regresi Logistik dan Penerapannya untuk Menduga Pola Kelulusan Mahasiswa TPB-IPB.
- Hajarisman, N. (2009), Analisis Data Kategorik. Bandung: Pusat Penerbitan Unisba.
- Kusuma, Komalasari, dan Hadijati. 2013. *Model Regresi Zero Inflated Poisson Pada Data Overdispersion*.
<https://ojs.unud.ac.id/index.php/jmat/article/view/16570/10863>. Diakses Pada Tanggal 2 Mei 2018.
- McCullagh, P., and J.A. Nelder (1989). *Generalized Linear Models. (Second Edition)*. New York: Chapman and Hall. Wilson
- Morel, Jorge G., and Neerchal, Nagaraj K. (2011). *Overdispersion Models in SAS*. Cary, NC, USA: SAS Institute Inc.
- Sunendiari, S. (2012), Statistika Matematika II. Bandung: Pusat Penerbitan Unisba.