PEMILIHAN METODE *BETA-BINOMIAL* DAN *LOGISTIC-NORMAL* DALAM MENGATASI OVERDISPERSI PADA REGRESI LOGISTIK

SKRIPSI

Sebagai salah satu syarat untuk memperoleh gelar
Sarjana Sains dalam bidang Statistika

oleh:
DEVA RIZKY YUNIANA
105090513111003-95

PROGRAM STUDI STATISTIKA
JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS BRAWIJAYA
MALANG
2014

LEMBAR PENGESAHAN SKRIPSI

PEMILIHAN METODE *BETA-BINOMIAL* DAN *LOGISTIC-NORMAL* DALAM MENGATASI OVERDISPERSI PADA REGRESI LOGISTIK

oleh :

DEVA RIZKY YUNIANA

105090513111003

Setelah dipertahankan di depan Majelis Penguji
pada tanggal 26 Juni 2014
dan dinyatakan memenuhi syarat untuk memperoleh gelar
Sarjana Sains dalam bidang Statistika

Dosen Pembimbing,

Ir. Heni Kusdarwati, MS
NIP. 196112081987012001

Mengetahui,
Ketua Jurusan Matematika
Fakultas MIPA Universitas Brawijaya

Dr. Abdul Rouf Alghofari, M.Sc
NIP. 19670971992031001

LEMBAR PERNYATAAN
Saya yang bertanda tangan di bawah ini :

Nama	**: DEVA RIZKY YUNIANA**
NIM	**: 105090513111003**
Program Studi	**: STATISTIKA**
Penulis Skripsi Berjudul	**:**

PEMILIHAN METODE *BETA-BINOMIAL* DAN *LOGISTIC-NORMAL* DALAM MENGATASI OVERDISPERSI PADA REGRESI LOGISTIK

Dengan ini menyatakan bahwa

1. **Isi dari skripsi yang saya buat adalah benar-benar karya sendiri dan tidak menjiplak karya orang lain, selain nama-nama yang termaktub di isi dan tertulis di daftar pustaka dalam skripsi ini.**
2. **Apabila di kemudian hari ternyata skripsi yang saya tulis terbukti hasil jiplakan, maka saya akan bersedia menanggung resiko yang akan saya terima.**

Demikian pernyataan ini dibuat dengan segala kesadaran.

Malang, 26 Juni 2014

Yang menyatakan,

DEVA RIZKY YUNIANA
NIM. 105090513111003-95

PEMILIHAN METODE *BETA-BINOMIAL* DAN *LOGISTIC-NORMAL* DALAM MENGATASI OVERDISPERSI PADA REGRESI LOGISTIK

ABSTRAK

Analisis regresi logistik digunakan untuk mengetahui hubungan peubah prediktor (X) dengan peubah respon (Y) yang bertipe kategorik. Dalam regresi logistik, peubah respon (Y) diasumsikan berdistribusi binomial dengan ragam pengamatan sama dengan ragam dugaan. Apabila ragam pengamatan lebih besar dari ragam dugaan maka terjadi overdispersi. Konsekuensi adanya overdispersi adalah kesalahan dalam penarikan kesimpulan akibat galat baku yang *underestimate*. Salah satu metode yang dapat mengakomodasi masalah overdispersi adalah metode *Williams,* namun terdapat kelemahan dari metode tersebut. Sebagai alternatif model regresi *beta- binomial* dan regresi *logistic-normal* dapat mengakomodasi masalah overdispersi pada regresi logistik. Penelitian ini bertujuan untuk mengetahui data dari hasil rancangan yang sesuai dimodelkan menggunakan regresi *beta-binomial* dan regresi *logistic-normal*. Adapun data yang digunakan adalah 3 data sekunder, di mana data 1 merupakan hasil rancangan dengan pengelompokan sedangkan data 2 dan 3 merupakan hasil rancangan tanpa pengelompokan serta terindikasi overdispersi. Berdasarkan uji kesesuaian model dan nilai AIC dihasilkan bahwa data dari hasil rancangan dengan pengelompokan lebih sesuai dimodelkan menggunakan regresi *logistic-normal* dan data dari rancangan tanpa pengelompokan lebih sesuai dimodelkan menggunakan regresi *beta-binomial*.

Kata Kunci : Rancangan Data, Regresi Logistik, Overdispersi, Regresi *Beta-Binomial,* Regresi *Logistic-Normal*

CHOOSING THE BETA-BINOMIAL METHOD AND LOGISTIC-NORMAL METHOD TO SOLVE OVERDISPERSION ON THE LOGISTIC REGRESSION

ABSTRACT

The logistic regression used to know the relationship between predictor variable (X) and categorical response variable (Y). In logistic regression, response variable (Y) is assumed to be binomially distributed with observation variance is similar to its estimating variance. If observation variance is higher than estimating variance, overdispersion will occurs. Overdispersion will lead to the wrong conclusion because standard error is underestimated. One of method to solve overdispersion is *Williams* method. Conversely, there is a weakness of this. *Beta-binomial* and *logistic-normal* regression is an alternative to solve overdispersion in logistic regression. The study aimed to know the suitable data for *beta-binomial* and *logistic-normal* regression. The data used three data where the data 1 is chosen through clustered sampling while the data 2 and 3 are chosen without clustered sampling which was indicated overdispersion. Based on the goodness of fit testing and AIC value, it showed that choosing data through clustered was more suitable with the use of *logistic- normal*regression and data without clustered was more suitable with the use of *beta-binomial* regression.

Keywords : Data Design, Overdispersion, Beta- Binomial Regression, Logistic-Normal Regression

KATA PENGANTAR

Puji syukur kehadirat Allah Subhanahu wa Ta'ala, atas segala rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi berjudul "Pemilihan Metode *Beta-Binomial* dan *Logistic-Normal* Dalam Mengatasi Overdispersi Pada Regresi Logistik*"* sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains dalam bidang Statistika. Oleh karena itu penulis ingin mengucapkan terima kasih kepada :

1. Ibu Ir. Heni Kusdarwati, MS selaku Dosen Pembimbing yang telah memberikan pengarahan dan masukan dengan sabar kepada penulis selama penyusunan skripsi ini.
2. Ibu Eni Sumarminingsih, S.Si., MM dan ibu Dr. Ir. Ni Wayan Surya Wardhani, MS selaku Dosen Penguji yang telah memberikan pengarahan dan masukan dengan sabar kepada penulis selama penyusunan skripsi ini.
3. Bapak Dr. Abdul Rouf Alghofari, M.Sc selaku Ketua Jurusan Matematika Fakultas MIPA Universitas Brawijaya.
4. Bapak ibu Dosen Statistika atas ilmu yang diberikan selama kuliah.
5. Ayah tersayang Hantoyo, Ibu tercinta Sri Hartati, kakak-kakakku David Handiyanto, Uniek Dwi Handayani, Retno Indriyani, dan Devi Rizky Yuniani atas doa, dukungan dan semangatnya.
6. Motivatorku Zull Irawan, Oma Harefa, Muhammad Abid dan Dinda Rinai Vivit Senja yang selalu memberikan semangat.
7. Semua sahabat Nurma, Nurul, Mahdiyatus, Ika, Ulfalina, Arista, Anggono, Faizatur, Widya, Eka, Mukhlisin, Fandy yang telah memberikan semangat serta doa dalam penyusunan skripsi ini.
8. Teman-teman Statistika 2010 atas bantuan dan kerja sama selama penyusunan skripsi ini dan semua pihak yang telah membantu.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan mengingat keterbatasan kemampuan penulis, oleh karena itu kritik dan saran yang bersifat membangun sangat diharapkan demi tersusunnya penulisan yang lebih baik. Semoga penulisan Skirpsi ini dapat memberikan manfaat bagi pembaca.

Malang, Juni 2014

Penulis

DAFTAR ISI

BAB III METODOLOGI

BAB IV HASIL DAN PEMBAHASAN

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN

BAB I
PENDAHULUAN

1.1 Latar Belakang

Analisis regresi logistik digunakan untuk menganalisis hubungan peubah prediktor (X) dengan peubah respon (Y) yang bertipe kategorik. *Generalized Linear Models* (GLMs) merupakan perluasan dari model linear klasik untuk memodelkan peubah respon yang berdistribusi tidak normal. Dalam regresi logistik, peubah respon (Y) diasumsikan berdistribusi binomial dengan ragam pengamatan sama dengan ragam dugaan. Apabila ragam pengamatan lebih besar dari ragam dugaan maka terjadi overdispersi. Menurut Hinde dan Demetrio (2007), beberapa hal yang menjadi penyebab terjadinya overdispersi antara lain menghilangkan kovariat yang penting, kesalahan dalam penentuan *link function*, adanya korelasi antar pengamatan, adanya pengelompokan serta kompleksitas pengamatan yang tidak dipahami secara baik.

Hasil penelitian tentang overdispersi pada regresi logistik yang dilakukan oleh Kurnia, dkk (2002) menunjukkan bahwa nilai duga galat baku dari model bernilai lebih kecil dari seharusnya (*underestimates*). Sebagai konsekuensi dari nilai duga galat baku yang dihasilkan pada model regresi logistik tersebut, statistik uji t lebih sering menolak H_0. Hal ini juga menyebabkan jumlah peubah yang berpengaruh nyata lebih besar dari seharusnya dan penarikan kesimpulan menjadi tidak tepat.

Pada penelitian tentang pengaruh tingkat dispersi terhadap galat baku pada regresi logistik, Rahmawati (2010) menganalisis data yang mengalami kasus overdispersi menggunakan metode *Williams*. Hasil yang diperoleh menunjukkan nilai duga parameter regresi logistik tidak jauh berbeda dengan metode *Williams* namun terjadi perubahan pada nilai χ^2 *Pearson*. Metode *Williams* menghasilkan nilai χ^2 *Pearson* yang mendekati derajat bebas data sehingga metode *Williams* mampu untuk mengakomodasi data yang mengalami overdispersi. Berdasarkan nilai AIC, dihasilkan metode *Williams* memiliki nilai yang lebih kecil dari pada regresi logistik tanpa penanganan kasus overdispersi sehingga model yang ditangani dengan metode *Williams* lebih layak. Akan tetapi, metode *Williams* memiliki kelemahan yakni tidak dapat digunakan untuk mengatasi data yang memiliki tingkat dispersi lebih dari 4.

Sebagai alternatif, Senja (2013) telah melakukan analisis dalam mengatasi overdispersi menggunakan metode *beta-binomial* pada data yang memiliki peubah respon berupa proporsi. Dalam pemodelan menggunakan regresi *beta-binomial,* Agresti (2002) menyatakan parameter yang digunakan adalah efek tetap saja. Di sisi lain, Hinde dan Demetrio (2007) juga memperkenalkan metode lain dalam mengatasi overdispersi pada regresi logistik di mana terdapat efek acak kelompok selain efek tetap dalam model yaitu *logistic-normal* dengan efek acak kelompok diasumsikan menyebar normal. Metode *beta-binomial* dikenal juga dengan istilah adalah *William Type II* sedangkan metode *logistic-normal* disebut juga *William Type III*.

Untuk memperkuat teori tersebut, pada penelitian ini akan digunakan dua jenis data, yaitu data dari hasil rancangan dengan pengelompokan dan data dari hasil rancangan tanpa pengelompokan. Masing-masing data akan dimodelkan menggunakan dua metode yaitu *logistic-normal* dan *beta-binomial* yang selanjutnya akan dibandingkan berdasarkan kesesuaian model dan nilai AIC. Jika data dari rancangan dengan pengelompokan lebih sesuai dimodelkan menggunakan regresi *logistic-normal* maka nilai AIC akan bernilai lebih kecil, sebaliknya jika data dari hasil rancangan tanpa pengelompokan lebih sesuai dimodelkan menggunakan regresi *beta-binomial* maka nilai yang dihasilkan akan bernilai lebih kecil.

1.2 Rumusan Masalah

Permasalahan dalam penelitian ini adalah:

1. Bagaimana memodelkan data menggunakan metode *beta-binomial* (*william type II*)?
2. Bagaimana memodelkan data menggunakan metode *logistic-normal (william type III)*?
3. Bagaimana memilih metode *logistic-normal* dan *beta-binomial* dalam mengatasi overdispersi pada regresi logistik?

1.3 Batasan Masalah

Batasan masalah dari penelitian ini adalah:

1. Model yang diteliti adalah model penuh *(full model)*.
2. Pembentukan model *beta-binomial (william Type II)* dan *logistic-normal (william Type III)* tanpa memperhatikan ada tidaknya pencilan data.
3. Data yang digunakan bebas multikolinearitas dan terindikasi adanya overdispersi lebih dari 4.

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah:

1. Membentuk model *beta-binomial (william type II)*.
2. Membentuk model *logistic-normal (william type III)*.
3. Memilih metode *logistic-normal* dan *beta-binomial* dalam mengatasi overdispersi pada regresi logistik berdasarkan nilai AIC.

1.5 Manfaat Penelitian

Manfaat yang dapat diperoleh dari penelitian ini adalah dapat diketahui penggunaan dan pemodelan regresi *beta-binomial (william type II)* dan *logistic-normal (william type III)* dalam mengatasi overdispersi berdasarkan data hasil rancangan.

BAB II
TINJAUAN PUSTAKA

2.1 Generalized Linear Models (GLMs)

Generalized Linear Models (GLMs) merupakan perluasan dari model linier klasik untuk menghubungkan peubah respon yang memiliki distribusi tidak normal dan memodelkannya dengan nilai harapan peubah respon. Terdapat tiga komponen GLMs menurut Agresti (2002) yaitu:

1. *Random component* (komponen acak) terdiri dari peubah respon Y yang saling bebas antar pengamatan $\left(y_1,\cdots,y_n\right)$ di mana peubah respon berasal dari distribusi keluarga eksponensial. Fungsi kepekatan peluang keluarga eksponensial adalah sebagai berikut:
$$f\left(y_i;\theta_i\right)=a\left(\theta_i\right)b(y_i)\exp[y_iQ\left(\theta_i\right)] \tag{2.1}$$
nilai parameter θ_i berubah-ubah untuk $i=1,\cdots,n$ tergantung dari nilai peubah penjelas. $Q\left(\theta\right)$ disebut juga *natural parameter*.
2. *Systematic component* (komponen sistematik) berkaitan dengan vektor$\left(\eta_1,\cdots,\eta_n\right)$ yang merupakan model linier dari beberapa peubah bebas. Jika x_{ij} dinotasikan sebagai nilai peubah bebas j untuk pengamatan ke-i, maka:
$$\eta_i=\sum_j\beta_jx_{ij} \qquad i=1,\cdots,n \tag{2.2}$$
Apabila $x_{ij}=1$ untuk semua *i*, maka koefisien intersep dinotasikan α.
3. *Link function* (fungsi penghubung) yang menghubungkan *random component* (komponen random) dengan *systematic component* (komponen sistematik). Apabila $\mu_i=E\left(Y_i\right)$ untuk $i=1,\cdots,n$ maka model $\eta_i=g\left(\mu_i\right)$ yang artinya bahwa komponen sistematik sama dengan fungsi dari nilai harapan, dinotasikan sebagai:
$$g\left(\mu_i\right)=\eta_i \tag{2.3}$$

2.2 Regresi Logistik

Regresi logistik adalah salah satu tipe dari *Generalized Linear Models* (GLMs) yang dapat digunakan untuk mengetahui pengaruh beberapa peubah bebas terhadap peubah respon yang berdistribusi binomial. Suatu peubah respon biner memiliki dua kemungkinan yaitu sukses dan gagal yang mengikuti distribusi Bernoulli jika $n=1$ dengan $P(Y=1)=\pi$ dan $P(Y=0)=1-\pi$.

Jika banyak percobaan lebih dari 2, maka Y akan menyebar secara binomial (n,π) dengan peluang:

$$P(Y=y)=\binom{n}{y}\pi^{y}(1-\pi)^{n-y} \tag{2.4}$$

di mana : $y=0,1,2,\cdots,n$

$n=$ banyaknya percobaan

Peubah acak Y_i menggambarkan banyaknya sukses dari n_i percobaan yang memiliki rata-rata dan ragam sebagai berikut:

$$E(Y_i)=\mu_i=n_i\pi_i \tag{2.5}$$

$$V(Y_i)=n_i\pi_i(1-\pi_i) \tag{2.6}$$

Menurut Hinde and Demetrio (2007) fungsi penghubung untuk sebaran binomial berbentuk:

$$g(\mu_i)=\ln\left(\frac{\mu_i}{n_i-\mu_i}\right)=\ln\left(\frac{\pi_i}{1-\pi_i}\right)=\eta_i \tag{2.7}$$

Model logit (*log odds*) dapat dituliskan sebagai:

$$\ln\left(\frac{\pi_i}{1-\pi_i}\right)=\alpha+\sum_{j=1}^{p}\beta_j x_{ij} \tag{2.8}$$

$$\frac{\pi_i}{1-\pi_i}=\exp\left(\alpha+\sum_{j=1}^{p}\beta_j x_{ij}\right) \tag{2.9}$$

$$\pi_i = \frac{\exp(\alpha + \sum_{j=1}^{p} \beta_j x_{ij})}{(1 + \exp(\alpha + \sum_{j=1}^{p} \beta_j x_{ij})} \tag{2.10}$$

di mana:

p = banyaknya peubah prediktor

$\alpha, \beta_{1,} \beta_2, \cdots, \beta_p$ = koefisien regresi setiap peubah predikor

$x_{i1}, x_{i2}, \cdots, x_{ip}$ = pengamatan ke- i untuk peubah prediktor ke $1, 2, \cdots, p$

π_i = peluang sukses kejadian amatan ke-i

2.3. Estimasi Parameter Regresi Logistik

Regresi logistik merupakan regresi nonlinier yang pendugaan parameternya menggunakan metode *Maximum Likelihood*. Jika dilakukan pengamatan secara bebas, maka fungsi *likelihood*-nya merupakan perkalian dari masing-masing peluang untuk $Y_{i1}, Y_{i2}, \cdots, Y_{ip}$ sehingga didapatkan fungsi *likelihood* untuk sebaran binom yaitu:

$$L\left(\beta_j\right) = \prod_{i=1}^{N} \pi\left(x_i\right)^{y_i} (1 - \pi(x_i))^{n_i - y_i} \tag{2.11}$$

$$= \prod_{i=1}^{N} \left(1 - \pi\left(x_i\right)\right)^{n_i} \left(\frac{\pi(x_i)}{1 - \pi(x_i)}\right)^{y_i} \tag{2.12}$$

di mana:

y_i = banyaknya sukses pada amatan ke-i

n_i = total amatan ke-i

π_i = peluang sukses kejadian amatan ke-i

x_i = peubah prediktor ke-i

Berdasarkan persamaan (2.9) dan (2.10) maka fungsi kemungkinannya menjadi :

$$\frac{\pi_i}{1-\pi_i}=\exp\left(\alpha+\sum_j\beta_j x_{ij}\right)$$

$$l\left(\beta\right)=\left(\frac{1}{1+\exp(\alpha+\sum_j\beta_j x_{ij})}\right)^{n_i}+\left(\exp\left(\alpha+\sum_j\beta_j x_{ij}\right)\right)^{y_i} \tag{2.13}$$

Fungsi $\ln L(\beta)$ digunakan untuk mempermudah dalam menghitung maksimum dari fungsi kemungkinannya, sehingga fungsi $\ln L(\beta)$ dari persamaan (2.13) menjadi:

$$l(\beta)=\ln L\left(\beta\right)$$

$$\begin{aligned}
l\left(\beta\right)&=\ln\left(\frac{1}{1+\exp(\alpha+\sum_j\beta_j x_{ij})}\right)^{n_i}+\ln\left(\exp\left(\alpha+\sum_j\beta_j x_{ij}\right)\right)^{y_i}\\
&=\ln(1+\exp(\alpha+\sum_j\beta_j x_{ij}))^{-n_i}+\ln\left(\exp\left(\alpha+\sum_j\beta_j x_{ij}\right)\right)^{y_i}\\
&=-n_i\ln(1+\exp(\alpha+\sum_j\beta_j x_{ij}))+\sum_{i=1}^{N}y_i\left(\alpha+\sum_j\beta_j x_{ij}\right)
\end{aligned} \tag{2.14}$$

Agar nilai fungsi $l\left(\beta\right)$ mencapai maksimum maka turunan parsial pertama dari $l\left(\beta\right)$ terhadap β harus sama dengan nol.

$$\frac{\partial l\left(\hat{\beta}\right)}{\partial\hat{\beta}}=0$$

$$\frac{\partial l\left(\hat{\beta}\right)}{\partial\hat{\beta}}=\sum_{i=1}^{N}y_i x_{ij}-\sum_{i=1}^{n}n_i x_{ij}\frac{\exp(\hat{\alpha}+\sum_{j=1}^{p}\hat{\beta}_j x_{ij})}{1+\exp(\hat{\alpha}+\sum_{j=1}^{p}\hat{\beta}_j x_{ij})}$$

$$0=\sum_{i=1}^{N} y_i x_{ij} - \sum_{i=1}^{n} n_i x_{ij} \frac{\exp(\hat{\alpha}+\sum_{j=1}^{p}\hat{\beta}_j x_{ij})}{1+\exp(\hat{\alpha}+\sum_{j=1}^{p}\hat{\beta}_j x_{ij})} \qquad (2.15)$$

sehingga diperoleh persamaan skor kemungkinan

$$\sum_{i=1}^{N} y_i x_{ij} - \sum_{i=1}^{N} n_i \hat{\pi}(x_i) x_{ij} = 0 \qquad j=1,\ldots,p$$

Untuk menyelesaikan persamaan (2.15) tidaklah mudah karena $\hat{\beta}_j$ bersifat non linear sehingga dibutuhkan metode iterasi. Menurut Agresti (2002), metode iteratif Newton-Raphson dapat digunakan untuk menyelesaikan persamaan non linear seperti menentukan nilai maksimum suatu fungsi.

Pada metode Newton-Raphson, $\hat{\beta}$ ditentukan dengan memaksimumkan $l(\beta)$. Jika $\boldsymbol{u}'$ merupakan matriks turunan pertama $l(\beta)$ terhadap β_0 dan β_j yang dinotasikan sebagai:

$$\boldsymbol{u}' = \left(\frac{\partial l(\boldsymbol{\beta})}{\partial \beta_0}, \frac{\partial l(\boldsymbol{\beta})}{\partial \beta_1}, \ldots, \frac{\partial l(\boldsymbol{\beta})}{\partial \beta_p} \right) \qquad (2.16)$$

Jika **H** adalah matriks yang berisi unsur-unsur turunan kedua terhadap β_a dan β_b (a,b= 0,.., p) yaitu:

$$h_{ab} = \frac{\partial^2 l(\boldsymbol{\beta})}{\partial \beta_a \partial \beta_b} = -\sum_i x_{ia} x_{ib} n_i \pi(x_i)(1-\pi(x_i))$$

(2.17)

Anggap $\boldsymbol{u}^{(t)}$ dan $\boldsymbol{H}^{(t)}$ sebagai **u** dan **H** terevaluasi pada $\boldsymbol{\beta}^{(t)}$ yaitu penduga $\hat{\boldsymbol{\beta}}$ pada iterasi ke-t $(t=0,1,2,..)$. Nilai $\boldsymbol{\beta}^{(t)}$ pada setiap iterasi mengikuti bentuk deret Taylor orde kedua yaitu:

$$L(\boldsymbol{\beta}) \approx L(\boldsymbol{\beta}^{(t)}) + \boldsymbol{u}^{(t)}(\boldsymbol{\beta}-\boldsymbol{\beta}^{(t)}) + \frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}^{(t)})' \boldsymbol{H}^{(t)}(\boldsymbol{\beta}-\boldsymbol{\beta}^{(t)}) \qquad (2.18)$$

Dengan menyelesaikan

$$\frac{\partial L(\hat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}} \approx \boldsymbol{u}^{(t)} - \boldsymbol{H}^{(t)}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^{(t)}) = 0 \qquad (2.19)$$

maka $\hat{\boldsymbol{\beta}}$ pada iterasi berikut adalah:

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} - (\boldsymbol{H}^{(t)})^{-1}\boldsymbol{u}^{(t)}$$

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} - \left\{-X'diag\left[n_i\pi(x_i^{(t)})(1-\pi(x_i^{(t)}))\right]x\right\}^{-1} X'\left(y_i - n_i\pi(x_i^{(t)})\right) \qquad (2.20)$$

Kekonvergenan $\boldsymbol{\beta}^{(t)}$ yang mendekati $\hat{\boldsymbol{\beta}}$ terpenuhi jika pada setiap-j

$$\left|\beta_j^{(t+1)} - \hat{\beta}_j^{(t)}\right| \le c\left|\beta_j^{(t)} - \hat{\beta}_j^{(t-1)}\right|^2 \qquad (2.21)$$

Untuk $c > 0$ (Agresti, 2002).

2.4. Overdispersi

Dalam regresi logistik, diasumsikan bahwa ragam pengamatan peubah respon Y_i akan sama dengan ragam dari model binomial yang diduga seperti pada persamaan (2.6). Ketidaksesuaian antara ragam amatan dengan ragam dugaan pada regresi logisik disebut juga dengan dispersi. Menurut Hinde dan Demetrio (2007) adanya overdispersi pada regresi logistik menunjukkan ragam hasil pengamatan lebih besar dari pada ragam dugaan model distribusi binomial. Konsekuensi terjadinya overdispersi dalam regresi logistik adalah nilai duga galat baku yang *underestimates*.

Jensen (2008) dalam Rahmawati (2010) menyatakan beberapa penyebab terjadinya overdispersi :

1. Menghilangkan kovariat yang penting.
2. Kesalahan dalam penentuan *link function.*
3. Adanya korelasi antar pengamatan.
4. Kompleksitas pengamatan yang tidak dipahami secara baik.

Menurut Hajarisman (2005) munculnya masalah overdispersi dapat dijelaskan dalam dua hal yaitu karena adanya keragaman peluang respon dan adanya korelasi antar peubah respon. Apabila terdapat keragaman peluang dalam respon maka terdapat korelasi antar peubah respon, begitupun sebaliknya. McCullagh dan Nelder (1989) dalam Hajarisman (2005) menyatakan bahwa kedua kejadian tersebut dapat terjadi karena adanya pengelompokan (*clustering*) sedangkan Collet (1990) menyebutkan bahwa kejadian-kejadian tersebut muncul karena adanya pengamatan berulang sehingga diperoleh suatu peluang respon yang berbeda dari satu percobaan ke percobaan lain.

2.5. Pemodelan Dispersi

Munculnya overdispersi dapat terjadi karena adanya korelasi antar peubah respon yang dapat dimodelkan sebagai berikut (Hajarisman, 2005):

$$Var(Y_i) = n_i \pi_i (1 - \pi_i)\left[1 + (n_i - 1)\delta\right] \tag{2.22}$$

di mana:

δ : koefisien korelasi

Apabila tidak terdapat korelasi antar pengamatan biner, $\delta = 0$ maka ragam pengamatan sesuai dengan ragam distribusi binomial. Namun, jika terjadi overdispersi maka $\delta \neq 0$ yakni ragam pengamatan lebih besar daripada ragam distribusi binomial.

Sedangkan pemodelan keragaman dalam peluang respon menurut Hajarisman (2005) dinyatakan sebagai:

$$Var(Y_i) = n_i \pi_i (1 - \pi_i)[1 + (n_i - 1)\varphi] \tag{2.23}$$

di mana:

φ : keragaman di dalam peluang respon

Jika φ bernilai 0 maka tidak terjadi dispersi dan respon mengikuti sebaran binomial selebihnya akan terjadi overdispersi saat $\varphi > 0$.

Pemeriksaan terjadinya overdispersi dapat dideteksi dengan menggunakan nilai statistik χ^2 *Pearson* yang diperoleh dari hasil analisis regresi logistik. Statistik χ^2 *Pearson* merupakan fungsi dari sisaan dinyatakan sebagai berikut (Agresti, 2002):

$$\chi^2 = \sum_{i=1}^{i=N} e_i^2 \tag{2.24}$$

dengan:

$$e_i = \frac{(y_i - \hat{\pi}_i n_i)}{\sqrt{\hat{\pi}_i n_i (1 - \hat{\pi}_i)}}$$

di mana:

χ^2 : statistik uji *Pearson*

e_i : sisaan *Pearson* pada pengamatan ke-*i*

y_i : banyak sukses pada pengamtan ke-*i*

n_i : total amatan ke-i

$\hat{\pi}_i$: peluang prediksi Y pada pengamatan ke-*i*

Overdispersi terjadi jika rasio antara χ^2 *Pearson* dan derajat bebas lebih dari 1, dinyatakan sebagai:

$$\frac{\chi^2 Pearson}{db} > 1 \quad , \; db = m - k \tag{2.25}$$

di mana:

m : banyaknya pengamatan

k : banyaknya parameter yang diduga

Menurut William (1982) dalam Rahmawati (2010) rasio antara χ^2 *Pearson* dan derajat bebas menunjukkan tingkat dispersi yang terjadi pada data binomial. Semakin besar tingkat dispersi maka semakin besar keragaman yang berlebih pada data binom (*extra binomial variation*).

2.6 Generalized Linear Mixed Models (GLMMs)

Generalized Linear Mixed Models (GLMMs) merupakan perluasan dari *Generalized Linear Models* (GLMs) dan model linier campuran yang dapat digunakan untuk memodelkan overdispersi dengan melibatkan efek acak (McCulloch, 1997). Jika $\boldsymbol{x}_{it}$ dinotasikan sebagai vektor kovariat untuk efek tetap subyek ke-*i* pengamatan ke-*j*, $\boldsymbol{\beta}$ sebagai parameter efek tetap, $\boldsymbol{z}_{it}$ adalah vektor kovariat efek acak subyek ke-*i* pengamatan ke-*j*, dan $\boldsymbol{u}_i$ sebagai efek acak pada kelompok ke- i maka model GLMMs dinotasikan sebagai:

$$g\left(\boldsymbol{\mu}_{it}\right) = \boldsymbol{x}'_{it}\boldsymbol{\beta} + \boldsymbol{z}'_{it}\boldsymbol{b}_i \tag{2.26}$$

Menurut Saavedra dalam Gunawan (2009), asumsi dalam GLMMs adalah:

1. Nilai ekspektasi dari peubah respon berhubungan dengan kovariat dan efek acak seperti berikut:

$$g\left(\boldsymbol{\mu}_{it}\right) = g\left[E\left(Y_{ij} \mid b_i\right)\right] = \boldsymbol{x}'_{it}\boldsymbol{\beta} + \boldsymbol{z}'_{it}\boldsymbol{b}_i$$

2. Pemberian b_i untuk masing-masing subyek sehingga $y_1, \cdots, y_N$ diasumsikan saling bebas, dan mengikuti GLMs di mana y memiliki fungsi kepekatan keluarga eksponensial $f\left(y \mid \beta, D, \sigma^2\right)$ dengan D merupakan matriks peragam.
3. Efek acak b_i saling bebas dan menyebar normal, $b_i \sim N\left(0, D\right)$.

2.7 Regresi *Logistik-Normal* (*William Type III*)

Menurut Nia (2006), GLMMs sering digunakan untuk memodelkan data yang mengalami overdispersi. Overdispersi dapat terjadi pada data pengamatan berkorelasi yang berkaitan dengan pengamatan berulang dan beberapa bentuk pengelompokan. Salah satu tipe khusus dari GLMMs adalah regresi *logistic-normal* di mana terdapat penambahan efek acak kelompok (*cluster*) pada model. Model *logistic-normal* dapat dituliskan sebagai berikut (Agresti, 2002):

$$Y_i \sim Bin\left(n_i, \pi_i\right)$$

$$Logit(\pi_{ij}) = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + u_i \qquad (2.27)$$

di mana:

x_{ij} = vektor kovariat yang berhubungan dengan subyek ke-*j* dalam kelompok (*cluster*) *i*

β = parameter efek tetap

u_i = efek acak intersep untuk kelompok (*cluster*) ke-*i* $u_i \sim N(0, \sigma^2)$

Menurut Hinde dan Demetrio (2007) fungsi ragam untuk model *Logistik-Normal* dituliskan sebagai:

$$Var(Y_i) \approx n_i \pi_i (1-\pi_i)\left[1+\sigma^2 (n_i - 1)\pi_i (1-\pi_i)\right] \qquad (2.28)$$

Saat $\sigma^2 = 0$ ragam akan tereduksi menjadi ragam binomial sehingga model *logistic-normal* mampu mengakomodasi adanya overdispersi. Fungsi ragam pada persamaan (2.28) tersebut dikatakan juga sebagai *William Tipe III*.

Fungsi kepekatan peluang bagi peubah acak $\boldsymbol{y}_i$ adalah sebagai berikut:

$$f\left(\boldsymbol{y}_i | \mu_i; \boldsymbol{X}_i, \beta\right) = \prod_{j=1}^{n_i} \frac{y_{ij} \exp(\boldsymbol{x}_{ij}'\boldsymbol{\beta} + \mu_i)}{1 + \exp(\boldsymbol{x}_{ij}'\boldsymbol{\beta} + \mu_i)} \qquad (2.29)$$

Adapun fungsi kepekatan peluang bersama bagi $(\boldsymbol{y}_i, \mu_i)$ adalah:

$$f\left(\boldsymbol{y}_i,\mu_i;\boldsymbol{X}_i,\boldsymbol{\beta},\sigma^2\right)=\prod_{j=1}^{n_i}\frac{y_{ij}\exp(\boldsymbol{x}_{ij}^{'}\boldsymbol{\beta}+\mu_i)}{1+\exp(\boldsymbol{x}_{ij}^{'}\boldsymbol{\beta}+\mu_i)}\frac{1}{\sqrt{2\pi\sigma}}\exp(-\frac{\mu_i^2}{2\sigma^2}) \quad (2.30)$$

2.8 Estimasi Parameter *Logistik-Normal*

Menurut Jorge, dkk (2006) pendugaan *Marginal Maximum Likelihood* umum digunakan untuk menduga model *Logistic-Normal* namun sulit untuk mendapatkan solusi analitik. Fungsi *likelihood* untuk model *Logistic-Normal* dapat dituliskan sebagai berikut (Agresti, 2002):

$$l(\boldsymbol{\beta},\sigma^2\,|\,\boldsymbol{y})=\prod_{i=1}^{n}\int_{-\infty}^{\infty}\prod_{j=1}^{q}\frac{\exp\left\{y_{ij}(\boldsymbol{x}_{it}^{'}\boldsymbol{\beta}+u_i)\right\}}{1+\exp\left\{y_{ij}(\boldsymbol{x}_{it}^{'}\boldsymbol{\beta}+u_i)\right\}}\times\frac{e^{-u_i^2/2\sigma^2}}{\left(2\pi\sigma^2\right)^{1/2}}du_i \quad (2.31)$$

Karena fungsi pada persamaan (2.31) tidak memiliki solusi secara analitik maka dibutuhkan pendekatan secara numerik. Integral dapat diselesaikan dengan mengunakan metode *Gauss-Hermite Quadrature* yang berguna untuk memaksimalkan fungsi *marginal likelihood* (Pickes, 2002).

Menurut Agresti (2002) *Gauss- Hermite Quadrature* adalah metode untuk menduga integral dari perkalian fungsi $f(.)$ dengan fungsi lain yang memiliki bentuk kepadatan normal. Pada kasus dengan efek acak univariat, bentuk pendugaan dapat dituliskan :

$$\int_{-\infty}^{\infty}\exp(-x^2)f(x)dx\approx\sum_{i=1}^{Q}w_i f(x_i) \quad (2.32)$$

di mana:

w_i : bobot

x_i : titik *quadrature*

Nia (2006) menyatakan persamaan untuk nilai bobot dalam *Gaussian Hermitte Quadrature* adalah:

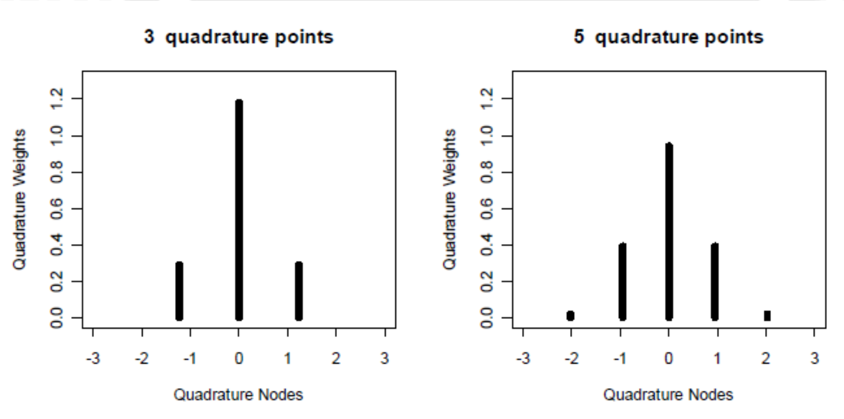
$$w_i=-\frac{2^{(n+1)}n!\sqrt{\pi}}{w_n(x_i)w_{n+1}(x_i)} \quad (2.33)$$

Dengan x_i merupakan akar-akar dari polinomial Hermitte dengan persamaan:

$$w_i(x)=(2x)w_{n-1}(x)-2(n-1)w_{n-2}(x) \quad (2.34)$$

Gunawan (2009) menyajikan nilai-nilai akar polinomial *Hermitte* dan bobotnya yang akan disajikan dalam Lampiran 2.

Jika dalam *Gaussian Quadrature* titik-titik *quadrature*nya tidak memberikan kontribusi terhadap bobot maka harus dilakukan pemusatan yang disebut *Adaptive Gaussian Quadrature* (Gunawan, 2009). Berikut adalah contoh grafik *Adaptive Gaussian Quadrature* :



Gambar 2.1 Grafik *Adaptive Gaussian Quadrature*
(Sumber : Nia, 2006)

Pada grafik *Adaptive Gaussian Quadrature* dengan tiga titik *quadrature* maka nilai bobot yang digunakan pada sisi kiri dan kanan dari titik pusat akan bernilai sama yakni 0.295409 sehingga grafik akan setangkup.

2.9 Regresi *Beta-Binomial (William Type II)*

Salah satu metode yang dapat digunakan untuk mengatasi masalah overdispersi adalah regresi *beta-binomial*. Menurut Agresti (2002) sebaran beta binomial adalah gabungan sebaran beta dengan binomial, yaitu Y diasumsikan mengikuti sebaran binomial, bin(n,π) dan π~Beta (α,β) $\alpha > 0$ $\beta > 0$. Adapun fungsi kepadatan peluang bagi sebaran beta adalah

$$f(\pi;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\pi^{\alpha-1}(1-\pi)^{\beta-1} \qquad 0 \le \pi \le 1 \tag{2.35}$$

dan fungsi kepadatan sebaran *Beta-Binomial* adalah :

$$f(y;\alpha,\beta) = f(y|\pi)\,f(\pi)$$
$$= \binom{n}{y}\pi^{y}(1-\pi)^{n-y}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\pi^{\alpha-1}(1-\pi)^{\beta-1} \qquad (2.36)$$

Persamaan (2.36) diintegralkan untuk mendapatkan fungsi kepadatan marginal sebagai berikut:

$$\begin{aligned} f(y;\alpha,\beta) &= \int_0^1 f(y|\pi)\,f(\pi) \\ &= \int_0^1 \binom{n}{y}\pi^{y}(1-\pi)^{n-y}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\pi^{\alpha-1}(1-\pi)^{\beta-1}\,d\pi \\ &= \int_0^1 \binom{n}{y}\frac{1}{B(\alpha,\beta)}\pi^{\alpha+y-1}(1-\pi)^{n+\beta-y-1}\,d\pi \\ &= \binom{n}{y}\frac{1}{B(\alpha,\beta)}B(\alpha+y,n+\beta-y) \end{aligned} \qquad (2.37)$$

di mana:

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

Apabila dimisalkan $\mu = \dfrac{\alpha}{\alpha+\beta}$ dan $\theta = \dfrac{1}{\alpha+\beta}$, maka persamaan (2.37) dapat ditulis menjadi:

$$f(y;\mu,\theta) = \binom{n}{y}\frac{\left[\prod_{k=0}^{y-1}(\mu+k\theta)\right]\left[\prod_{k=0}^{n-y-1}(1-\mu+k\theta)\right]}{\prod_{k=0}^{n-1}(1+k\theta)} \qquad (2.38)$$

di mana: $0 \le \mu \le 1$ dan $\theta \ge 0$

Rata - rata dan ragam dari persamaan (2.38) adalah :

$$E(Y) = n\mu$$
$$V(Y) = n\mu(1-\mu)\frac{1+n\theta}{1+\theta} \qquad (2.39)$$

Ragam dari persamaan (2.39) dapat juga dituliskan sebagai :

$$Var(Y) = n\mu(1-\mu)\left[1+(n-1)\phi\right] \quad (2.40)$$

Ketika $\phi = 0$ model tereduksi menjadi model binomial.

Menurut Hinde dan Demetrio (2007), fungsi ragam dari model *Beta-Binomial* disebut juga *Williams Type II*. Crowder (1978) dalam Li (2006) mereparameterisasi rata-rata dari distribusi binomial untuk *cluster* ke-i dengan memasukkan konsep regresi linear pada model *Beta-Binomial*. Rata-rata ke-i pada distribusi binomial dapat dituliskan sebagai:

$$\mu_i = \frac{\exp(\boldsymbol{x}_i'\boldsymbol{\beta})}{1+\exp(\boldsymbol{x}_i'\boldsymbol{\beta})} \quad (2.41)$$

di mana:

$\boldsymbol{x}_i$ adalah vektor kovariat yang berhubungan dengan *cluster* ke-i

$\boldsymbol{\beta}$ adalah vektor koefisien regresi yang bersesuaian dengan $\boldsymbol{x}_i$

Menurut Agresti (2002), sebaran *Beta-Binomial* memiliki parameter (μ_i, θ) sehingga model logit yang menghubungkan nilai tengah peubah respon dengan *p* peubah bebas dituliskan sebagai:

$$Logit(\mu_i) = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} \quad (2.42)$$

2.10 Estimasi Regresi Beta-Binomial

Pendugaan parameter untuk regresi *beta-binomial* menggunakan metode *kemungkinan maksimum* dengan prosedur iterasi *Newton Raphson*. Berdasarkan asumsi nilai $\theta = \dfrac{1}{\alpha+\beta}$ dan $\theta = \dfrac{1}{\alpha+\beta}$

maka fungsi *likelihood* dari persamaan (2.37) dapat dituliskan sebagai:

$$L(\mu,\theta) = \prod_i f(y_i) \quad (2.43)$$

$$L(\mu,\theta) = \prod_i \binom{n_i}{y_i} \frac{\left[\prod_{k=0}^{y_i-1}(\mu+k\theta)\right]\left[\prod_{k=0}^{n_i-y_i-1}(1-\mu+k\theta)\right]}{\prod_{k=0}^{n_i-1}(1+k\theta)}$$

Jika $l(\mu,\theta)$ merupakan fungsi logaritma dari $L(\mu,\theta)$ maka nilai turunan dari fungsi logaritma tersebut dapat didefinisikan sebagai:

$$\mathbf{S} = \begin{bmatrix} \frac{dl}{d\mu} & \frac{dl}{d\theta} \end{bmatrix}^T$$

di mana:

$$\frac{dl}{d\mu} = \sum_i \sum_{k=0}^{y_i-1} \left(\frac{1}{\mu + k\theta} \right) - \sum_i \sum_{k=0}^{n_i-y_i-1} \left(\frac{1}{1-\mu+k\theta} \right)$$

$$\frac{dl}{d\theta} = \sum_i \sum_{k=0}^{y_i-1} \left(\frac{k}{\mu + k\theta} \right) + \sum_i \sum_{k=0}^{n_i-y_i-1} \left(\frac{k}{1-\mu+k\theta} \right) - \sum_i \sum_{k=0}^{n_i-1} \left(\frac{k}{1+k\theta} \right) \tag{2.44}$$

serta nilai turunan keduanya didefinisikan sebagai:

$$\mathbf{O} = -\begin{bmatrix} \frac{d^2l}{d\mu^2} & \frac{d^2l}{d\mu d\theta} \\ \frac{d^2l}{d\mu d\theta} & \frac{d^2l}{d\theta^2} \end{bmatrix}$$

di mana :

$$\frac{d^2l}{d\mu^2} = \sum_i \sum_{k=0}^{y_i-1} \frac{1}{(\mu+k\theta)^2} - \sum_i \sum_{k=0}^{n_i-y_i-1} \frac{1}{(1-\mu+k\theta)^2}$$

$$\frac{d^2l}{d\theta^2} = \sum_i \sum_{k=0}^{y_i-1} \frac{-k^2}{(\mu+k\theta)^2} - \sum_i \sum_{k=0}^{n_i-y_i-1} \frac{k}{(1-\mu+k\theta)^2} + \sum_i \sum_{k=0}^{n_i-1} \frac{k^2}{(1+k\theta)^2}$$

$$\frac{d^2l}{d\theta d\mu} = \sum_i \sum_{k=0}^{y_i-1} \frac{-r}{(\mu+k\theta)^2} + \sum_i \sum_{k=0}^{y_i-1} \frac{-r}{(1-\mu+k\theta)^2} \tag{2.45}$$

Stokes dan Hill (1985) dalam Lau (2013) menggunakan Iterasi *Newton-Raphson* untuk menemukan penduga maksimum dari parameter μ dan θ. Secara umum, suatu titik x pada fungsi $f(x)$ yang nonlinear didekati dengan meggunakan metode *Newton Raphson* adalah:

$$x^{(t+1)} = x^{(t)} + \left[f\,''(x^{(t)}) \right]^{-1} f\,'(x^{(t)}) \tag{2.46}$$

di mana:

$x^{(t+1)}$: titik hasil iterasi ke t+1

$x^{(t)}$: titik awal atau titik hasil iterasi ke-t

$f'(x)$: turuna pertama dari $f(x)$

$f''(x)$: turunan kedua dari fungsi $f(x)$

Analog dengan persamaan (2.39) parameter μ dan θ diduga dengan menyelesaikan iterasi:

$$\begin{bmatrix} \mu \\ \theta \end{bmatrix}_m = \begin{bmatrix} \mu \\ \theta \end{bmatrix}_{m-1} + \mathbf{O}_{m-1}^{-1}\mathbf{S}_{m-1} \tag{2.47}$$

di mana m menujukkan iterasi ke-m. Nilai awal untuk μ dan θ adalah $\frac{\sum\sum y_{ij}}{n}$ dan 0.

2.11 Multikolinearitas

Multikolinearitas dapat diartikan sebagai adanya hubungan linear antar peubah X. Menurut Gujarati (2003) terdapat dua macam hubungan linear yaitu hubungan linear yang bersifat sempurna (multikolinearitas sempurna) dan hubungan linear tetapi bersifat tidak sempurna (multikolinearitas tidak sempurna).

Pada kasus multikolinearitas sempurna, nilai koefisien regresi tidak dapat ditentukan karena efek determinan matriks kovariat yang bernilai 0 (matriks singular). Sebaliknya, pada kasus multikolinearitas tidak sempurna nilai determinan matriks kovariat hampir mendekati 0 sehingga nilai invers pada matriks tersebut bernilai tak hingga, akibatnya nilai ragam menjadi besar. Nilai ragam yang relatif besar akan berpengaruh pada selang kepercayaan yang semakin lebar pula. Hal ini menyebabkan peluang untuk menerima hipotesis nol lebih besar yang artinya bahwa tidak ada satupun atau sangat sedikit sekali peubah bebas yang berpengaruh terhadap respon, meskipun nilai koefisien determinasi tinggi (Gujarati, 2003).

Menurut Li (2000) multikolinearitas dapat diperiksa dengan menggunakan nilai VIF (*Variance Inflation Factor*). Untuk regresi dengan lebih dari dua peubah, definisi VIF adalah:

$$VIF_j = \frac{1}{1-R_j^2} \tag{2.48}$$

Dengan R_j^2 merupakan koefisien determinasi *(auxiliary regression*) dari peubah bebas X_j sebagai respon yang diregresikan dengan peubah X lain sebagai prediktor. Jika nilai VIF kurang dari 10 maka antar peubah bebas tidak memiliki korelasi, namun nilai VIF yang lebih dari 10 mengidentifikasikan adanya multikolinearitas serius pada model.

2.12 Pengujian Hipotesis Parameter

Pengujian terhadap parameter dilakukan secara parsial maupun serentak. Pengujian secara parsial menggunakan statistik t dan Z sedangkan uji serentak dilakukan dengan uji *likelihood ratio test*.

a. Uji Serentak

Uji serentak dilakukan untuk menguji signifikansi dari koefisien regresi secara bersama-sama. Hipotesis yang digunakan adalah:

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

$H_1 :$ paling sedikit terdapat satu j dimana $\beta_j \neq 0$

$j = (1, \ldots, p)$

Menurut Agresti (2002), pendekatan uji statistik diperoleh melalui pendekatan *likelihood ratio test* antara model tanpa menyertakan β_i dan model yang menyertakan β_i, dimana $i = 1, 2, \cdots, p$. Menurut Hosmer dan Lemeshow (2000), model yang menyertakan β_i disebut *full model* sedangkan untuk model yang tanpa menyertakan β_i disebut *reduced model*.

$$G = -2\ln\left(\frac{L_0(\boldsymbol{\beta})}{L_p(\boldsymbol{\beta})}\right) = -2\left[\ln L_0(\boldsymbol{\beta}) - \ln L_p(\boldsymbol{\beta})\right] \tag{2.49}$$

di mana:

L_0 : *loglikelihood* tanpa β_i, dimana $i = 1, 2, \cdots, p$

L_p : *loglikelihood* dengan β_i dimana $i = 1, 2, \cdots, p$

Nilai G dibandingkan dengan statistik $\chi_{(v)}{}^2$ dengan derajat bebas v sesuai dengan jumlah parameter yang diduga. H_0 akan diterima jika niai p lebih besar dibandingkan peluang berbuat salah sebesar α dan sebaliknya jika nilai p lebih kecil dari α maka hipotesis nol ditolak.

b. Uji Parsial

Uji parsial digunakan untuk melihat signifikansi parameter secara individual pada masing-masing peubah bebas (Kutner dkk.,2004). Hipotesis yang digunakan pada uji ini adalah:

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0 \quad j = (1, \ldots, p)$$

Statistik uji yang digunakan adalah statistik *Wald*, yaitu:

$$W_j = \frac{\hat{\beta}_j}{SE\left(\hat{\beta}_j\right)} \tag{2.50}$$

dimana:

$$SE\left(\hat{\beta}_j\right) = \sqrt{Var(\hat{\beta}_j)}$$

Statistik uji W mengikuti sebaran normal. Hipotesis nol ditolak jika $\left|W_j\right| \geq Z_{\alpha/2}$.

2.13 Uji Kesesuaian Model (*Goodness of Fit*)

Untuk menguji apakah model yang dihasilkan sesuai atau tidak, maka perlu dilakukan pengujian kesesuaian model atau *goodness of fit*. Adapun statistik uji yang digunakan adalah χ^2 *Pearson* seperti persamaan (2.26) yaitu :

$$\chi^2 = \sum_{i=1}^{i=N} e_i^2$$

di mana:

$$e_i = \frac{(y_i - \hat{\pi}_i n_i)}{\sqrt{\hat{\pi}_i n_i (1 - \hat{\pi}_i)}} \tag{2.51}$$

Berdasarkan hipotesis untuk menguji kesesuaian model:

H_0: model sesuai

H_1: model tidak sesuai

Statistik uji *Pearson* χ^2 akan mengikuti sebaran χ^2 dengan derajat bebas $N-p$. H_0 ditolak jika $\chi^2 > \chi^2_{(\alpha;N-p)}$.

2.14 Pemilihan Model Terbaik

Menurut Kutner dkk (2004) pemilihan model terbaik dilakukan dengan melihat nilai *AIC (Akaike Information Criterion)*. Adapun persamaan yang digunakan untuk menghitung nilai *AIC* adalah sebagai berikut :

$$AIC = -2\ln L(\beta) + 2p \tag{2.52}$$

di mana p menunjukkan banyak parameter dalam model. Model terbaik adalah model yang memiliki nilai *AIC* terkecil.

2.15 Interpretasi Model Logit

Menurut Hosmer dan Lemeshow (2000), odds rasio adalah parameter yang lebih mudah dipahami maknanya dari pada parameter β pada model logit regresi logistik. Odds adalah perbandingan peluang suatu kejadian sukses dengan peluang kejadian gagal.Menurut Kutner dkk (2004) apabila terdapat satu buah peubah bebas, maka penduga model logit regresi logistik pada saat $X = X_j$ dapat dituliskan sebagai berikut:

$$\text{logit}\left[\hat{\pi}\left(X_j\right)\right] = \beta_0 + \beta_1 X_j \tag{2.53}$$

Penduga model logit regresi logistik pada saat $X = X_j + 1$ dapat dituliskan sebagai berikut:

$$\text{logit}\left[\hat{\pi}\left(X_j + 1\right)\right] = \beta_0 + \beta_1\left(X_j + 1\right) \tag{2.54}$$

Persamaan (2.52) adalah log dari penduga odds ketika $X = X_j$, sedangkan persamaan (2.53) adalah log dari penduga odds ketika $X = X_j + 1$. Perbedaan antara dua nilai tersebut adalah sebagai berikut:

$$\text{logit}\left[\hat{\pi}\left(X_j + 1\right)\right] - \text{logit}\left[\hat{\pi}\left(X\right)\right] = \beta_1$$

$$\log\left(odds_2\right) - \log\left(odds_1\right) = \beta_1$$

$$\log\left(\frac{odds_2}{odds_1}\right) = \beta_1$$

$$\log\text{OR} = \beta_1$$

$$\text{OR} = \exp\left(\beta_1\right) \qquad (2.55)$$

BAB III
METODE PENELITIAN

3.1 Sumber Data

Pada penelitian ini dipakai 3 data sekunder dengan uraian sebagai berikut:

3.1.1 Data 1

Data 1 menunjukkan hasil belajar mahasiswa yang berada di Amerika Serikat. Pada penelitian ini ingin diketahui bagaimana pengaruh status kependudukan mahasiswa yang berasal dari dalam dan luar negeri terhadap kelulusan mahasiswa. Selain itu, tahun kelulusan juga dipertimbangkan dalam penelitian tersebut. Rancangan yang digunakan pada penelitian tersebut adalah rancangan dengan pengelompokan. Kelompok dipilih secara acak yaitu 3 universitas di Amerika Serikat. Data diambil dari http://data.princenton.edu/www509/datasets/phd.dat.

Keterangan peubah:

X_1 : Universitas (1 = Berkeley, 2 = Columbia, 3 = Princeton)

X_2 : Status kependudukan mahasiswa (1 = asli, 2 = sementara)

X_3 : tahun kelulusan $(1,\cdots,8)$

y : jumlah mahasiswa lulus

n : jumlah seluruh mahasiswa

3.1.2 Data 2

Sebuah rancangan tanpa pengelompokan digunakan dalam penelitian yang bertujuan untuk mengetahui pengaruh pemberian obat pembasmi jenis *pyrethrins* dan *piperonyl* terhadap banyaknya kumbang mati. Setiap kenaikan konsentrasi *piperonyl* diulang sebanyak 3 kali. Sumber : http://www.stat.ufl.edu/~winner/data/
tribol_tox.dat

Keterangan peubah:

X_1 : konsentrasi *pyrethrins*

X_2 : konsentrasi *piperonyl* (1 = 0 cc, 2 = 0.25 cc, 3 = 2.5 cc, 4 = 10 cc)

y : jumlah kumbang mati

n : jumlah kumbang

3.1.3 Data 3

Penelitian dilakukan untuk mengetahui pengaruh pemberian obat pembasmi jenis *pyrethrins* terhadap banyak lalat mati (Morel dan Nagaraj, 2011). Rancangan yang digunakan dalam penelitian adalah rancangan tanpa pengelompokan dengan perlakuan yaitu 11 macam konsentrasi *pyrethrins*.

Keterangan peubah:

X_1 : konsentrasi *pyrethrins*

y : jumlah lalat mati

n : jumlah lalat

3.1 Metode Analisis *Logistic-Normal* dan *Beta-Binomial*

Langkah-langkah analisis data yang dilakukan pada penelitian ini adalah sebagai berikut:

1. Memeriksa multikolinearitas antar peubah penjelas dengan menggunakan nilai VIF sesuai dengan persamaan (2.48). Apabila diperoleh nilai VIF kurang dari 10 maka dilanjutkan ke tahap selanjutnya.
2. Menduga parameter logistik dengan menggunakan metode *maksimum likelihood* berdasarkan persamaan (2.15). Dari hasil pendugaan parameter logistik tersebut didapatkan nilai $\chi^2 Pearson$.
3. Mendeteksi overdispersi dengan menggunakan nilai statistik uji *Pearson* sesuai persamaan (2.24). Apabila $\frac{\chi^2 Pearson}{db} = 1$, tidak

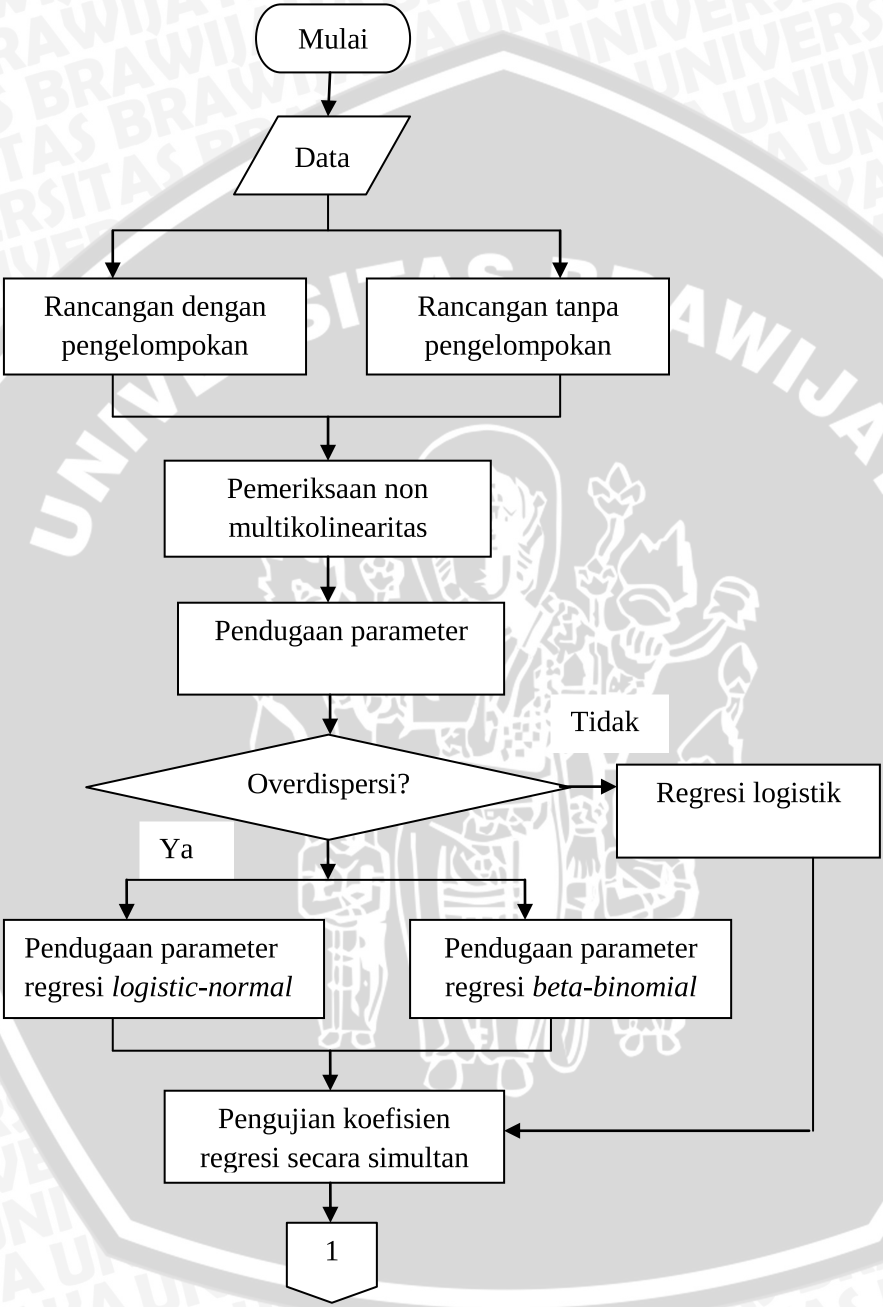
terjadi overdispersi. Apabila $\frac{\chi^2 Pearson}{db} > 1$ diindikasikan terjadi overdispersi.

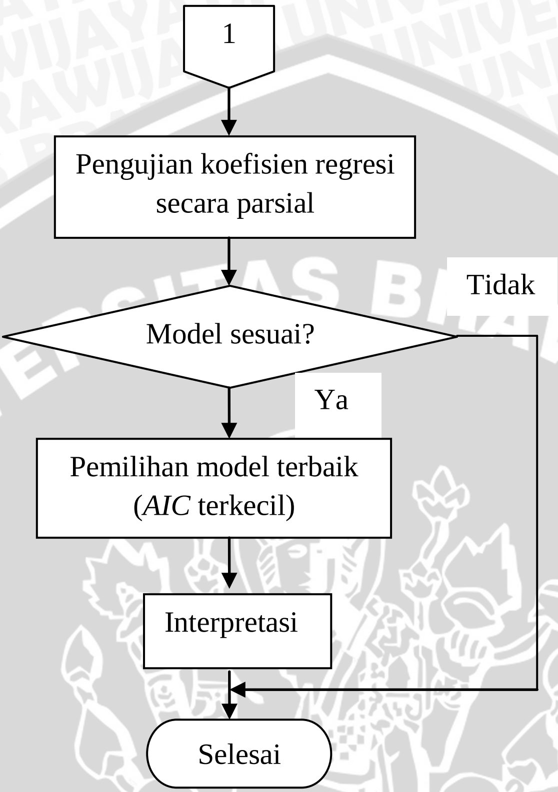
4. 1). Apabila tidak terjadi overdispersi, maka hasil pendugaan parameter regresi logistik telah sesuai.

   2). Apabila terjadi overdispersi, maka membentuk model regresi *logistic*-*normal* dan *beta*-*binomial*. Pendugaan parameter model regresi *logistic*-*normal* menggunakan metode *Gauss Hermitte Quadrature* dan pendugaan parameter regresi *beta- binomial* dengan menggunakan metode kemungkinan maksimum.

5. Menguji signifikansi parameter regresi *logistic*-*normal* dan *beta-binomial* secara simultan menggunakan *likelihood ratio test* (G) berdasarkan persamaan (2.51) dan mengambil kesimpulan dengan cara membandingkan nilai G dengan nilai χ^2 pada taraf nyata $\alpha = 5\%$. Jika hasil pengujian ini tidak signifikan maka model regresi yang digunakan tidak sesuai.
6. Menguji signifikansi parameter regresi *logistic*-*normal* secara parsial menggunakan statistik u*ji Wald* berdasarkan persamaan (2.50). Parameter yang tidak signifikan tidak dikeluarkan melainkan tetap digunakan dalam model karena model yang ingin dibentuk adalah model penuh.
7. Menguji kelayakan model menggunakan uji *Pearson* berdasarkan persamaan (2.24).
8. Melakukan pemilihan model terbaik adalah dengan melihat nilai *AIC* berdasarkan persamaan (2.52) untuk tiap model. Model terbaik yaitu model yang memiliki nilai *AIC* terkecil.
9. Interpretasi model.

Diagram alir dari metode penelitian ditujukkan pada Gambar 3.1. Analisis data dilakukan dengan menggunakan bantuan *software* Minitab dan SPSS untuk uji multikolinearitas dan *syntax* SAS 9.3.15 untuk pengujian overdispersi serta pembentukan model.


Mulai
Data
Rancangan dengan pengelompokan
Rancangan tanpa pengelompokan
Pemeriksaan non multikolinearitas
Pendugaan parameter
Overdispersi?
Tidak
Regresi logistik
Ya
Pendugaan parameter regresi *logistic-normal*
Pendugaan parameter regresi *beta-binomial*
Pengujian koefisien regresi secara simultan
1



Gambar 3.1. Diagram Alir Regresi *Logistic-Normal* dan Regresi *Beta-Binomial*

BAB IV
HASIL DAN PEMBAHASAN

4.1 Pemeriksaan Multikolinearitas

Pemeriksaan multikolinearitas terhadap data bertujuan untuk mengetahui ada tidaknya hubungan linear di antara beberapa atau semua peubah penjelas. Nilai VIF dari ketiga data sekunder yang digunakan dalam penelitian ini disajikan dalam Tabel 4.1. Secara lengkap, hasil uji multikolinieritas tertera pada Lampiran 3.

Tabel 4.1. Hasil Pemeriksaan Multikolinieritas

Data	Peubah	Nilai VIF
1	X_1 (tahun kelulusan)	1.0
	X_2 (status kependudukan mahasiswa)	1.03
	X_3 (universitas)	1.07
2	X_1 (konsentrasi *pyrethrins*)	1.513
	X_2 (konsentrasi *piperonyl*)	1.513
3	X_1 (konsentrasi *pyrethrins*)	-

Berdasarkan Tabel 4.1, dapat diketahui peubah prediktor pada data 1, dan 2 memiliki nilai VIF kurang dari 10 sedangkan pada data 3 tidak terdapat nilai VIF karena peubah prediktor yang digunakan hanya 1. Hal ini menunjukkan bahwa data sekunder yang digunakan tidak terdapat multikolinieritas sehingga bisa dilakukan analisis selanjutnya.

4.2 Pendugaan Parameter Regresi Logistik

Hasil pendugaan parameter regresi logistik digunakan untuk mendapatkan nilai χ^2 *Pearson* pada setiap data. Data yang memiliki kategori lebih dari 2 diubah menjadi peubah *dummy*.

Berikut adalah pendugaan parameter regresi logistik dengan pendekatan kemungkinan maksimum (*Maximum Likelihood Estimator*).

Tabel 4.2 Hasil Pendugaan Parameter Regresi Logistik

Data	Peubah	Parameter Duga	Statistik Uji *Wald*	Nilai p	χ^2 *Pearson*	db
1	Konstanta	-2.6557	1155.20	<.0001	333.5494	37
	tahun_1	-1.6657	283.46	<.0001		
	tahun_2	-0.3033	17.27	<.0001		
	tahun_3	0.4970	54.92	<.0001		
	tahun_4	0.6738	100.59	<.0001		
	tahun_5	0.6432	87.34	<.0001		
	tahun_6	0.4904	46.68	<.0001		
	tahun_7	0.2837	13.93	0.0002		
	status_kependudukan	0.5315	223.77	<.0001		
	univ_1	-0.9254	615.96	<.0001		
	univ_2	-0.8250	1072.96	<.0001		
2	Konstanta	-2.3085	147.25	<.0001	50.3686	7
	pyrethrins	6.9845	372.02	<.0001		
	piperonyl_1	-4.8931	320.63	<.0001		
	piperonyl_2	-3.4766	267.03	<.0001		
	piperonyl_3	-0.9344	32.80	<.0001		
3	Konstanta	-1.5762	331.31	<.0001	111.5718	9
	pyrethrins	0.0228	873.59	<.0001		

Berdasarkan hasil pendugaan parameter regresi logistik pada Tabel 4.2., diperoleh nilai χ^2 *Pearson* masing-masing data dengan nilai yang berbeda-beda. Hasil selengkapnya dapat dilihat pada Lampiran 4.

4.3 Pendeteksian Kasus Overdispersi

Pendeteksian kasus overdispersi pada regresi logistik dilakukan dengan melihat rasio antara χ^2 *Pearson* dengan derajat bebas data. Jika diperoleh nilai lebih dari 1 maka dinyatakan terjadi kasus overdispersi.

Secara ringkas hasil pengujian kasus overdispersi pada keempat data disajikan dalam Tabel 4.3. Hasil selengkapnya tertera pada Lampiran 4.

Tabel 4.3. Hasil Pengujian Overdispersi

Data	χ^2 *Pearson* / db
1	9.0148
2	7.1955
3	12.3969

Hasil pengujian overdispersi pada Tabel 4.3 diperoleh hasil bahwa rasio antara χ^2 *Pearson* dengan derajat bebas masing-masing data yaitu secara berturut-turut adalah 9.0148, 7.1955 dan 12.3969. Ketiga data tersebut mengalami kasus overdispersi di mana semua data memiliki nilai dispersi lebih dari 4.

4.4 Data 1

4.4.1 Pemodelan Regresi *Logistic-Normal*

Data 1 merupakan data pengamatan kelulusan mahasiswa di Amerika Serikat yang memiliki rancangan dengan pengelompokan. Dalam penelitian ini, universitas yang berada di Amerika Serikat antara lain Berkeley, Columbia dan Princeton dianggap sebagai kelompok. Berikut akan disajikan nilai duga parameter regresi *logistic-normal* pada Tabel 4.4. Secara lengkap, hasil pendugaan parameter regresi *logistic-normal* terdapat pada Lampiran 5.

Tabel 4.4. Nilai Parameter Duga Model Regresi *Logistic-Normal*

Peubah	Nilai parameter Duga
Konstanta	-3.4097
tahun_1	-1.6638
tahun_2	-0.3021
tahun_3	0.4980
tahun_4	0.6749
tahun_5	0.6443
tahun_6	0.4912
tahun_7	0.2848
status_kependudukan	0.4459

Model regresi *logistic-normal* yang terbentuk adalah sebagai berikut:

$$\text{Logit}(\mu_i) = (-3.4097 + u_i) - 1.6638 * \text{tahun\_1} - 0.3021 * \text{tahun\_2}$$
$$+0.4980 * \text{tahun\_3} + 0.6749 * \text{tahun\_4} + 0.6444 * \text{tahun\_5}$$
$$+0.4912 * \text{tahun\_6} + 0.2848 * \text{tahun\_7} + 0.4459 * \text{status\_kependudukan}$$

Nilai odds ratio untuk peubah tahun_1 adalah sebesar 0.1894, memiliki arti bahwa peluang mahasiswa lulus pada tahun pertama adalah 0.1894 kali dibanding tahun ke-delapan.

4.4.2 Pengujian Signifikansi Parameter Regresi *Logistic-Normal*

Hasil pendugaan koefisien regresi, statistik uji G dan nilai statistik uji *Wald* regresi *logistic-normal* untuk data 1 disajikan dalam Tabel 4.5. Secara lengkap, hasil pengujian signifikansi parameter regresi *logistic-normal* data 1 terdapat pada lampiran 5.

Tabel 4.5. Hasil Pendugaan Koefisien Regresi, Statistik uji G dan Nilai Statistik Uji *Wald* Regresi *Logistic-Normal*

Parameter Duga	Efek	Parameter Duga	Galat Baku	Statistik Uji Wald	Nilai p
β_0	-	-3.4097	0.9044	-3.77	0.0130
β_1	Tetap	-1.6638	0.09896	-16.81	<.0001
β_2	Tetap	-0.3021	0.07302	-4.14	0.0090
β_3	Tetap	0.4980	0.06708	7.42	0.0007
β_4	Tetap	0.6749	0.06720	10.04	0.0002
β_5	Tetap	0.6443	0.06883	9.36	0.0002
β_6	Tetap	0.4912	0.07179	6.84	0.0010
β_7	Tetap	0.2848	0.07602	3.75	0.0133
β_8	Tetap	0.4459	0.5715	0.78	0.4706
σ_u^2	Acak	0.4870	0.2835	1.72	0.1465
Statistik Uji G					658.8

Berdasarkan hasil pengujian parameter secara simultan untuk data 2 yang terdapat pada Tabel 4.5., didapatkan hasil nilai statistik uji G adalah 658.8. Derajat bebas pada data tersebut bernilai 9 sehingga jika statistik uji G dibandingkan dengan $\chi^2_{0.05,9}$ = 16.91898 maka dapat dikatakan bahwa hasil dari pendugaan parameter model penuh regresi *logistic-normal* untuk data 1 secara simultan berpengaruh terhadap banyaknya mahasiswa yang lulus.

Selain itu, hasil pengujian signifikansi parameter duga regresi secara parsial yang didasarkan pada nilai statistik uji *Wald* menyatakan bahwa semua efek tetap memiliki nilai p kurang dari 0.05, kecuali untuk efek tetap β_8 dan efek acak yang memiliki nilai p lebih dari 0.05. Hal ini menunjukkan bahwa tahun kelulusan ke-1 sampai dengan ke-7 berpengaruh secara parsial terhadap banyaknya mahasiswa yang lulus sedangkan status kependudukan mahasiswa dianggap tidak mempengaruhi banyaknya mahasiswa yang lulus secara parsial. Adanya pengelompokan (universitas) juga tidak menunjukkan adanya perbedaan antara universitas 1 dengan yang lainnya terhadap jumlah kelulusan mahasiswa. Dengan kata lain, pemilihan 3 universitas sebagai kelompok dianggap homogen sehingga pengelompokan tidak memiliki arti.

4.4.3 Uji Kesesuaian dan Pemilihan Model Terbaik

Model yang telah terbentuk selanjutnya diperiksa untuk mengetahui apakah model sesuai atau tidak menggunakan statistik uji *Pearson* sedangkan pemilihan model terbaik didasarkan pada nilai *AIC*. Adapun nilai *Pearson* model regresi *logistic-normal* adalah sebesar 45.47350. Jika dibandingkan dengan $\chi^2_{(0.05,38)} = 53.38354$ maka model regresi *logistic-normal* yang digunakan sesuai (nilai *Pearson* < $\chi^2_{(0.05,38)}$). Nilai AIC yang dihasil dari model regresi *logistic-normal* adalah sebesar 678.8.

4.4.4 Pemodelan Regresi *Beta-Binomial*

Data 1 memiliki rancangan dengan pengelompokan. Peubah universitas yang pada mulanya dianggap efek acak kelompok, maka pada pemodelan regresi *beta-binomial* diasumsikan sebagai efek tetap. Hasil pendugaan parameter regresi *beta-binomial* akan disajikan dalam Tabel 4.6. Secara lengap tertera pada Lampiran 6.

Tabel 4.6. Nilai Parameter Duga Model Regresi *Beta-Binomial*

Peubah	Nilai parameter Duga
Konstanta	-2.7239
tahun_1	-1.5060
tahun_2	-0.1763
tahun_3	0.6136
tahun_4	0.7640
tahun_5	0.6761
tahun_6	0.5037
tahun_7	0.2782
status_kependudukan	0.5075
univ_1	-0.8827
univ_2	-0.8130

Model *beta-binomial* yang terbentuk adalah sebagai berikut:

$$\begin{aligned}\text{Logit}(\mu_i) = & -2.7239 - 1.5060 * \text{tahun\_1} - 0.1763 * \text{tahun\_2} \\ & + 0.6136 * \text{tahun\_3} + 0.7640 * \text{tahun\_4} + 0.6761 * \text{tahun\_5} \\ & + 0.5037 * \text{tahun\_6} + 0.2782 * \text{tahun\_7} + 0.5075 * \text{status\_kependudukan} \\ & - 0.8827 * \text{univ\_1} - 0.8130 * \text{univ\_2}\end{aligned}$$

Nilai odds ratio peubah tahun_1 sebesar 0.2217 mempunyai arti bahwa peluang mahasiswa lulus pada tahun pertama adalah 0.2217 kali dibanding tahun ke-delapan.

4.4.5 Pengujian Signifikansi Parameter Regresi *Beta-Binomial*

Hasil pendugaan koefisien regresi, statistik uji G dan nilai statistik uji *Wald* regresi *beta-binomial* disajikan dalam Tabel 4.7. dan secara lengkap terdapat pada Lampiran 6.

Tabel 4.7. Hasil Pendugaan Koefisien Regresi, Statistik Uji G dan Nilai Statistik Uji *Wald* Regresi *Beta-Binomial*

Peubah Bebas	Parameter Duga	Galat Baku	Statistik Uji Wald	Nilai p
Konstanta	-2.7239	0.09052	-30.09	<.0001
tahun_1	-1.5060	0.1144	-13.17	<.0001

Tabel 4.7. (Lanjutan)

tahun_2	-0.1763	0.08701	-2.03	0.0428
tahun_3	0.6136	0.08051	7.62	<.0001
tahun_4	0.7640	0.08086	9.45	<.0001
tahun_5	0.6761	0.08346	8.10	<.0001
tahun_6	0.5037	0.08711	5.78	<.0001
tahun_7		0.09241	3.01	0.0026
status_kependudukan	0.5075	0.03921	12.94	<.0001
univ_1	-0.8827	0.04166	-21.19	<.0001
univ_2	-0.8130	0.02780	-29.25	<.0001
Statistik Uji G				566.1

Pada Tabel 4.7., didapatkan hasil nilai statistik uji G adalah 566.1. Derajat bebas pada data tersebut bernilai 10 sehingga jika statistik uji G dibandingkan dengan $\chi^2_{0.05,10} = 18.30704$ maka dapat dikatakan bahwa hasil dari pendugaan parameter model penuh regresi *beta-binomial* untuk data 1 secara simultan berpengaruh terhadap banyaknya mahasiswa yang lulus.

Selain itu, hasil pengujian signifikansi parameter duga regresi secara parsial yang didasarkan pada nilai statistik uji *Wald* menyatakan bahwa semua efek tetap memiliki nilai p kurang dari 0.05. Hal ini menunjukkan bahwa tahun kelulusan ke-1, sampai dengan ke-7, status kependudukan, univ_1 dan univ_2 berpengaruh secara parsial terhadap banyaknya mahasiswa yang lulus.

4.4.6 Uji Kesesuaian dan Pemilihan Model Terbaik

Model yang telah terbentuk selanjutnya diperiksa untuk mengetahui apakah model telah sesuai atau tidak menggunakan statistik uji *Pearson* sedangkan pemilihan model terbaik didasarkan pada nilai *AIC*. Adapun nilai *Pearson* yang dihasilkan sebesar 260. Jika dibandingkan dengan $\chi^2_{(0.05,37)} = 52.19232$ maka model regresi *beta-binomial* tidak sesuai (nilai *Pearson* > $\chi^2_{(0.05,37)}$). Hal ini juga disebabkan karna pada data 1 sebenarnya memiliki rancangan dengan pengelompokan sehingga ketika pemodelan yang digunakan adalah *beta-binomial* sehingga model yang diperoleh tidak sesuai. Nilai AIC yang diperoleh dari model regresi *beta-binomial* adalah sebesar 590.1.

4.5 Data 2

4.5.1 Pemodelan Regresi *Logistic-Normal*

Data 2 merupakan data pengamatan kumbang yang memiliki rancangan tanpa pengelompokan. Metode *logistic-normal* seharusnya digunakan untuk memodelkan data yang memiliki rancangan dengan pengelompokan. Namun, untuk memperkuat teori jenis data yang sesuai dimodelkan menggunakan model regresi *logistic-normal*, maka data berikut akan tetap dimodelkan menggunakan metode tersebut. Peubah yang digunakan adalah konsentrasi *pyrethrins* dan *piperonyl*. Pada pemodelan ini, *piperonyl* diasumsikan sebagai efek acak kelompok sedangkan *pyrethrins* sebagai efek tetap. Hasil pendugaan parameter regresi *logistic-normal* untuk data pengamatan kumbang akan disajikan pada Tabel 4.8. Secara lengkap, hasil pendugaan parameter regresi *logistic-normal* data 2 terdapat pada Lampiran 5.

Tabel 4.8. Nilai Parameter Duga Model Regresi *Logistic-Normal*

Peubah	Nilai parameter Duga
Konstanta	-4.6034
pyrethrins	6.9416

Model regresi *logistic-normal* yang terbentuk adalah sebagai berikut:

$$\mathrm{Logit}(\mu_i) = (-4.6034 + \mathrm{u}_i) + 6.9416 * \mathrm{pyrethrins}$$

Nilai odds ratio untuk peubah *pyrethrins* dengan kadar 1.5 cc adalah 0.997, mempunyai arti bahwa peluang kumbang mati dibanding kumbang hidup jika diberi obat pembasmi jenis *pyrethrins* dengan kadar 1.5cc adalah 0.997 kali.

4.5.2 Pengujian Signifikansi Parameter Regresi *Logistic-Normal*

Hasil pendugaan koefisien regresi, statistik uji G dan nilai statistik uji *Wald* regresi *logistic-normal* untuk data 2 disajikan dalam Tabel 4.9. Secara lengkap, hasil pengujian signifikansi parameter regresi *logistic-normal* data 2 terdapat pada lampiran 5.

Tabel 4.9. Hasil Pendugaan Koefisien Regresi, Statistik uji G dan Nilai Statistik Uji *Wald* Regresi *Logistic-Normal*

Parameter Duga	Efek	Parameter Duga	Galat Baku	Statistik Uji Wald	Nilai p
β_0	-	-4.6034	1.0040	-4.59	0.0195
β_1	Tetap	6.9416	0.3616	19.20	0.0003
σ_u^2	Acak	3.7655	2.6924	1.40	0.2564
Statistik Uji G					133.3

Berdasarkan hasil pengujian parameter secara simultan untuk data 2 yang terdapat pada Tabel 4.9., didapatkan hasil nilai statistik uji G adalah 133.3. Derajat bebas pada data tersebut bernilai 2 sehingga jika statistik uji G dibandingkan dengan $\chi^2_{0.05,2}$= 5.991 maka dapat dikatakan bahwa hasil dari pendugaan parameter model penuh regresi *logistic-normal* untuk data 2 secara simultan berpengaruh terhadap banyak kumbang mati.

Berdasarkan pengujian signifikansi parameter secara parsial yang dapat dilihat dari statistik uji *Wald* dipeoleh hasil bahwa efek tetap (*pyrethrins*)memiliki nilai p kurang dari 0.05 sedangkan efek acak kelompok (*piperonyl*) memiliki nilai p lebih dari 0.05. Ini berarti hanya penggunaan konsentrasi *pyrethrins* yang berpengaruh secara parsial terhadap banyak kumbang mati.

4.5.3 Uji Kesesuaian dan Pemilihan Model Terbaik

Model yang telah terbentuk selanjutnya diperiksa untuk mengetahui apakah model telah sesuai atau tidak menggunakan statistik uji *Pearson.* Adapun nilai *Pearson* model regresi *logistic-normal* yang dihasilkan adalah 26.19008. Jika dibandingkan dengan $\chi^2_{(0.05,38)} = 15.50731$ maka model *logistic-normal* tidak sesuai (nilai *Pearson* $> \chi^2_{(0.05,38)} = 15.50731$). Hal ini juga berhubungan dengan penggunaan model yang tidak sesuaikan dengan rancangan data. Memang telah disebutkan sebelumnya bahwa data yang memiliki rancangan tanpa pengelompokan pemodelan yang tepat adalah regresi *beta-binomial*. Ini terbukti pada hasil analisis yang menyatakan bahwa ketika data 2 yang memiliki rancangan tanpa pengelompokan dimodelkan menggunakan regresi *logistic-normal* maka model menunjukan ketidaksesuaiannya. Nilai AIC yang dihasil dari model regresi *logistic-normal* adalah sebesar 139.3.

4.5.4 Pemodelan Regresi *Beta-Binomial*

Pada data 2, rancangan yang digunakan adalah rancangan tanpa pengelompokan sehingga pemodelan regresi *beta-binomial* dapat langsung dilakukan. Adapun hasil pendugaan parameter regresi beta-binomial akan disajikan pada Tabel 4.10. dan secara lengkap dapat dilihat pada Lamipran 6.

Tabel 4.10. Nilai Parameter Duga Model Regresi *Beta-Binomial*

Peubah	Nilai parameter Duga
Konstanta	-2.3203
pyrethrins	6.9995
piperonyl_1	-4.9206
piperonyl_2	-3.4721
piperonyl_3	-0.9405

Model regresi *beta-binomial* yang terbentuk adalah sebagai berikut:

$$\text{Logit}(\mu_i)=-2.3203+6.9995*\textit{pyrethrins}-4.9206*\textit{piperonyl\_1}$$
$$-3.4721X*\textit{piperonyl\_2}-0.9405*\textit{piperonyl\_3}$$

Nilai odds ratio untuk peubah *pyrethrins* dengan kadar 1.5 cc adalah 0.999, mempunyai arti bahwa peluang kumbang mati dibanding kumbang hidup jika diberi obat pembasmi jenis *pyrethrins* dengan kadar 1.5cc adalah 0.999 kali.

4.5.5 Pengujian Signifikansi Parameter Regresi *Beta-Binomial*

Hasil pendugaan koefisien regresi, statistik uji G dan nilai statistik uji *Wald* regresi *beta-binomial* disajikan dalam Tabel 4.11. dan secara lengkap terdapat pada Lampiran 6.

Tabel 4.11. Hasil Pendugaan Koefisien Regresi, Statistik Uji G dan Nilai Statistik Uji *Wald* Regresi *Beta-Binomial*

Peubah Bebas	Parameter Duga	Galat Baku	Statistik Uji Wald	Nilai p
Konstanta	-2.3203	0.3940	-5.89	<.0001
pyrethrins	6.9995	0.7654	9.14	<.0001
piperonyl_1	-4.9206	0.5708	-8.62	<.0001
piperonyl_2	-3.4721	0.4231	-8.21	<.0001
piperonyl_3	-0.9405	0.3071	-3.06	0.0022
Statistik uji G				87.1149

Pada Tabel 4.11., tampak bahwa nilai statistik uji G pada data 2 adalah 87.1149 dengan $\chi^2_{(0.05,4)}$ = 9.487729. Statistik uji G > $\chi^2_{(0.05,4)}$, sehingga dapat dikatakan bahwa hasil pendugaan parameter model *beta-binomial* secara keseluruhan memiliki pengaruh terhadap banyaknya kumbang mati.

Pengujian signifikansi parameter duga secara parsial model *beta-binomial* untuk data 2 juga dapat dilihat pada Tabel 4.11. Pada metode *beta-binomial*, parameter duga yang digunakan dalam membentuk model regresi *beta-binomial* adalah efek tetap. Berdasarkan hasil pengujian signifikansi parameter duga diketahui bahwa nilai p untuk statistik uji *Wald* semua efek tetap bernilai kurang dari 0.05. Hal ini menunjukkan bahwa konsentrasi *pyrethrins*, konsentrasi *piperonyl* dengan kadar 0 cc, 0.25 cc, dan 2.5cc secara parsial terhadap banyaknya kumbang yang mati.

4.5.6 Uji Kesesuaian dan Pemilihan Model Terbaik

Model yang telah terbentuk selanjutnya diperiksa untuk mengetahui apakah model telah sesuai atau tidak menggunakan statistik uji *Pearson* sedangkan pemilihan model terbaik didasarkan pada nilai *AIC*. Adapun nilai *Pearson* model regresi *beta-binomial* yang diperoleh adalah sebesar 12.9351. Jika dibandingkan dengan $\chi^2_{(0.05,7)} = 15.50731$ maka model *beta-binomial* sesuai (nilai *Pearson* < $\chi^2_{(0.05,7)}$). Nilai AIC yang diperoleh dari model regresi *beta-binomial* adalah sebesar 99.1149.

4.6 Data 3

4.6.1 Pemodelan Regresi *Logistic-Normal*

Data 3 merupakan data pengamatan lalat yang terdiri dari 1 peubah yang sebenarnya memiliki rancangan tanpa pengelompokan. Data yang akan dimodelkan mengguankaan model *logistic-normal* harus memiliki dua efek yaitu efek tetap dan efek acak kelompok. Ketika data 4 akan dimodelkan menggunakan model regresi *logistic-normal* tentu akan mengalami kesulitan sebab peubah yang digunakan diasumsikan efek acak kelompok sesangakn efek tetap tidak ada. Pada akhirnya, data 3 tidak dapat dimodelkan menggunakan model regresi *logistic-normal*.

4.6.2 Pengujian Signifikansi Parameter Regresi *Logistic-Normal*

Telah dijelaskan sebelumnya bahwa data 3 tidak dapat dimodelkan menggunakan regresi *logistic-normal* sehingga pendugaan maupun pengujian parameter pun tidak dapat dilakukan.

4.6.3 Uji Kesesuaian dan Pemilihan Model Terbaik

Pengujian kesesuain dan pemilihan model terbaik tidak dapat dilakukan karena tidak ada pemodelan regresi *logistic-normal*.

4.6.4 Pemodelan Regresi *Beta-Binomial*

Data 3 merupakan hasil rancangan yang sama seperti data 2 yaitu rancangan tanpa pengelompokan. Hasil pendugaan regresi *beta-binomial* akan disajikan pada Tabel 4.12. Secara lengkap hasil pendugaan parameter regresi *beta-binomial* untuk data 4 tertera pada Lampiran 6.

Tabel 4.12. Nilai Parameter Duga Model Regresi Logistik dan Regresi *Beta-Binomial*

Peubah	Nilai Parameter Duga
Konstanta	-1.5753
pyrethrins	0.02262

Adapun model *beta-binomial* yang terbentuk adalah sebagai berikut:

$$Logit(\mu_i) = -1.5753 + 0.02262 * pyrethrins$$

Nilai odds ratio untuk peubah *pyrethrins* dengan kadar 40 cc adalah 0.338, mempunyai arti bahwa peluang lalat mati dibanding lalat hidup jika diberi obat pembasmi jenis *pyrethrins* dengan kadar 40cc adalah 0.338 kali.

4.6.5 Pengujian Signifikansi Parameter Regresi *Beta-Binomial*

Hasil pendugaan koefisien regresi, statistik uji G dan nilai statistik uji *Wald* regresi *beta-binomial* disajikan dalam Tabel 4.13. dan secara lengkap terdapat pada Lampiran 6.

Tabel 4.13. Hasil Pendugaan Koefisien Regresi, Statistik Uji G dan Nilai Statistik Uji *Wald* Regresi *Beta-Binomial*

Peubah Bebas	Parameter Duga	Galat Baku	Statistik Uji Wald	Nilai p
Konstanta	-1.5753	0.2440	-6.46	<.0001
pyrethrins	0.02262	0.00232	9.73	<.0001
Statistik uji G				99.5724

Pada Tabel 4.13., tampak bahwa nilai statistik uji G pada data 3 adalah 99.5724 dengan $\chi^2_{(0.05,1)}$ = 3.841459. Statistik uji G > $\chi^2_{(0.05,1)}$ sehingga dapat dikatakan bahwa hasil pendugaan parameter model *beta-binomial* secara keseluruhan memiliki pengaruh terhadap banyaknya lalat yang mati.

Pengujian signifikansi parameter duga regresi secara parsial dapat dilihat pula pada Tabel 4.23. Berdasarkan statistik uji *wald*, nilai p efek tetap bernilai kurang dari 0.001. Nilai tersebut lebih kecil dari 0.05 sehingga dapat dikatakan konsentrasi *pyrethrins* berpengaruh secara parsial terhadap banyaknya lalat yang mati.

4.6.6 Uji Kesesuaian dan Pemilihan Model Terbaik

Model yang telah terbentuk selanjutnya diperiksa untuk mengetahui apakah model telah sesuai atau tidak menggunakan statistik uji *Pearson* sedangkan pemilihan model terbaik didasarkan pada nilai *AIC*. Adapun nilai *Pearson* model regersi *beta-binomial* yang dihasilkan adalah 15.3368. Jika dibandingkan dengan akan $\chi^2_{(0.05,7)} = 18.30704$ maka model regresi *beta-binomial* sesuai (nilai *Pearson* < $\chi^2_{(0.05,7)}$). Nilai AIC yang diperoleh dari model regresi *beta-binomial* adalah sebesar 105.6.

4.7 Perbandingan Regresi *Logistic-Normal* dan *Regresi Beta-Binomial* Secara Keseluruhan

Meskipun data mengalami overdispersi, namun hasil pendugaan regresi logistik sebenarnya sudah menghasilkan nilai parameter duga yang bersifat tak bias. Hal ini terlihat pada perubahan nilai duga parameter dari regresi *logistic-normal* dan regresi *beta-binomial* yang tidak berbeda jauh.

Selain dari nilai parameter duga, untuk menentukan data yang sesuai dimodelkan menggunakan regresi *logistic-normal* ataupun regresi *beta-binomial* dapat juga dibandingkan melalui uji kesesuain model serta nilai AIC yang akan disajikan pada Tabel 4.14.

Tabel 4.14. Perbandingan Statistik Uji *Pearson* dan AIC

Data	Uji Kesesuaian Model		AIC	
	LN	*BB*	*LN*	*BB*
1	Sesuai	Tidak Sesuai	678.8	590.1
2	Tidak Sesuai	Sesuai	139.3	99.1149
3	-	Sesuai	-	105.6

Berdasarkan Tabel 4.14. dapat dilihat bahwa data 1 ketika dimodelkan menggunakan regresi *logistic-normal* menghasilkan model yang sesuai sedangkan ketika dimodelkan menggunakan regresi *beta-binomial* menghasilkan model yang tidak sesuai. Hal ini dikarenakan pada data 1 sebenarnya merupakan data dari hasil rancangan dengan pengelompokan dan berdasarkan teori yang ada bahwa data yang mengandung unsur pengelompokan maka seharusnya dimodelkan menggunakan regresi *logistic-normal*. Meskipun nilai AIC pada model regresi *beta-binomial* lebih kecil namun tetap saja model tersebut tidak layak digunakan. Oleh karena itu, model yang sesuai digunakan untuk data dari hasil rancangan dengan pengelompokan adalah *logistic-normal*.

Pada data 2 dan 3 ketika dimodelkan menggunakan regresi *beta-binomial* dihasilkan model yang sesuai sedangkan ketika dimodelkan menggunakan regresi *logistic-normal* tidak sesuai. Hal ini juga dikarenakan data yang digunakan adalah data hasil rancangan tanpa pengelompokan sehingga memang benar bahwa data dari hasil rancangan tanpa pengelompokan lebih sesuai dimodelkan menggunakan regresi *beta-binomial*. Jika dilihat nilai AIC untuk data 2, model *beta-binomial* memiliki nilai yang lebih kecil daripada regresi *logistic-normal*. Kecuali untuk data 3, karena tidak dapat dimodelkan menggunakan model regresi *logistic-normal* yang telah dibahas sebelumnya maka tanpa perlu membandingkan dapat dikatakan bahwa model *beta-binomial* lebih sesuai untuk data yang memiliki rancangan tanpa pengelompokan.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil analisis pada bab sebelumnya didapatkan model dari hasil penanganan overdispersi sebagai berikut:

1. Model *logistic-normal* yang terbentuk adalah:

   Data 1 :

$$
\begin{aligned}
Logit(\mu_i) = (-3.4097 + u_i) - 1.6638 * \text{tahun\_1} - 0.3021 * \text{tahun\_2} \\
+0.4980 * \text{tahun\_3} + 0.6749 * \text{tahun\_4} + 0.6444 * \text{tahun\_5} \\
+0.4912 * \text{tahun\_6} + 0.2848 * \text{tahun\_7} + 0.4459 * \text{status\_kependudukan}
\end{aligned}
$$

2. Model *beta-binomial* yang terbentuk adalah:

   Data 2 :

$$
\begin{aligned}
\text{Logit}(\mu_i) = -0.3203 + 6.9995 * \textit{pyrethrins} - 4.9206 * \textit{piperonyl\_1} \\
-3.4721X * \textit{piperonyl\_2} - 0.9405 * \textit{piperonyl\_3}
\end{aligned}
$$

   Data 3 :

$$
Logit(\mu_i) = -1.5753 + 0.02262 * \textit{pyrethrins}
$$

3. Berdasarkan uji kesesuaian model dan nilai AIC diperoleh hasil bahwa data dari hasil rancangan dengan pengelompokan lebih sesuai dimodelkan dengan regresi *logistic-normal* sedangkan data dari rancangan tanpa pengelompokan lebih sesuai dimodelkan dengan regresi *beta-binomial*.

5.2 Saran

Perlu diperhatikan rancangan data terlebih dahulu sebelum memilih metode yang akan digunakan. Sebaiknya model *logistic-normal* dan *beta-binomial* juga perlu diterapkan pada data yang terindikasi underdispersi.

DAFTAR PUSTAKA

Agresti, A. 2002. *Categorical Data Analysis*. John Wiley and Sons, Inc: New York.

Anonymous. 1997. http://data.princenton.edu/www509/datasets /phd.dat, tanggal akses : 27 Oktober 2013

Anonymous. 1994. 2011. http://www.stat.ufl.edu/~winner/data/tribol_ tox.data, tanggal akses : 25 Maret 2014

Gujarati, D. 2003. *Basic Econoetrics*. Mc Graw-Hill : New York.

Gunawan, B. 2009. *Analisis Data Longitudinal Diskrit Menggunakan Generalized Linear Mixed Model (GLMMs) pada Data Binom.* Jurusan Matematika, Fakultas MIPA. Universitas Brawijaya. Malang. Skripsi. Tidak dipublikasikan.

Hajarisman, N. 2005. *Pemodelan Overdispersi dalam Analisis Data Biner Melalui Model Regresi William*. Statistika. Vol. 5 No.1. Halaman 23-30.

Hinde, J. and C.G.B. Demetrio. 2007. *Overdispersi: Models and Estimation*. http://www.lce.esalq.usp.br/arquivos/aulas/2011/ /LCE5868/OverdispersionBook.pdf.

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

Jorge, G., F.Tuerlinckx, P.D e Boeck dan R. Cools. 2006. *Numerical Integration Logistic-Normal Model*. Computational Statistics and Data Analysis hal 1539.

Kurnia, A., A. Saefuddin dan E. Sutisna. 2002. *Overdispersi dalam Regresi Logistik*. Forum Statistika dan Komputasi hal. 11 –17.

Kutner, M.H., C.J. Nachtsheim., J. Neter., and W. Li. 2004. *Applied Linier Statistical Models. Fifth Edition*. McGraw-Hill, Inc: New York

Lau, A. 2013. *Using Maximum Likelihood Estimator for Identifying Interviewer Effect with the Beta-binomial Model*. http://www.stat.fi/isi99/ proceedings/arkisto/varasto/lau_0717. pdf.

Li, F. 2000. *Multicollinearity Department of Statistics*. Stockholm University. http://people.su.se/~fli/teachin/geconometrics/slides /slides/F7-Melticollinearity.pdf.

Li, Yinmei. 2006. *Power Analysis For A Mixed Effect Logistic Regression Model*. http://www.sciquest.org.nz/elibrary/ download/63542/Power_analysis_for_mixed effects_logistic _regressi.pdf.

Morel, J. dan Nagaraj Neerchal. 2011. *Overdispersion Models in SAS*. SAS Institute Inc: USA. http://books.google.co.id. Tanggal akses : 26 Maret 2014.

Nia, V. P. 2006. *Gauss-Hermitte Quadrature: Numerical of Statistcal Method*. http://vahid.probstat.ch/paper/ghd.pdf.

Rahmawati, D. 2010. *Pengaruh Tingkat Dispersi pada Regresi Logistik*. Jurusan Matematika, Fakultas MIPA. Universitas Brawijaya. Malang. Skripsi. Tidak dipublikasikan.

Saefuddin, A dan N. A. Setiabudi. 2011. *The Effect of Overdispersion on Logistic Regression Analysis of Poverty in Indonesia*. International Journal for Statistician. Vol. 2(1) Page 1 – 4. http://asaefuddin.com/ dokumen/The-Effect-of-Overdispersion-on-Logistic-Regression-Analysis-of-Poverty-in-Indonesia.pdf.

Senja, V. R. D. 2013. *Penanganan Overdispersi Menggunakan Regresi Beta Binomial Pada Regresi Logistik*. Jurusan Matematika, Fakultas MIPA. Universitas Brawijaya. Malang. Skripsi. Tidak dipublikasikan.

Lampiran 1. Data Sekunder

Data 1. Data Pengamatan Kelulusan Mahasiswa di Amerika Serikat

Tahun kelulusan	Status	Universitas	Banyak mahasiswa lulus	Jumlah mahasiswa
1	1	1	31	7422
2	1	1	177	7166
3	1	1	393	6759
4	1	1	484	6138
5	1	1	500	5506
6	1	1	399	4824
7	1	1	309	4277
8	1	1	210	3836
1	1	2	14	3247
2	1	2	43	3038
3	1	2	95	2801
4	1	2	99	2502
5	1	2	106	2216
6	1	2	88	1952
7	1	2	64	1683
8	1	2	45	1429
1	1	3	52	1628
2	1	3	166	1509
3	1	3	269	1278
4	1	3	189	961
5	1	3	117	728
6	1	3	62	582
7	1	3	34	485
8	1	3	13	426
1	2	1	26	1558
2	2	1	103	1485

Lampiran 1. (Lanjutan)

3	2	1	193	1334
4	2	1	191	1097
5	2	1	98	876
6	2	1	74	746
7	2	1	46	641
8	2	1	33	564
1	2	2	12	805
2	2	2	14	722
3	2	2	36	657
4	2	2	39	563
5	2	2	25	460
6	2	2	12	384
7	2	2	8	330
8	2	2	6	255
1	2	3	26	427
2	2	3	65	376
3	2	3	80	280
4	2	3	51	178
5	2	3	17	114
6	2	3	8	85
7	2	3	2	72
8	2	3	2	37

Data 2. Data Pengamatan Kumbang

Konsentrasi *pyrethrins*	Konsentrasi *piperonyl*	Banyak kumbang mati	Jumlah kumbang
1.5	0	138	150
1.06	0	75	149
0.75	0	32	150
1.1	0.25	129	151
0.78	0.25	65	151

Lampiran 1. (Lanjutan)

0.55	0.25	19	150
0.8	2.5	143	149
0.57	2.5	112	150
0.4	2.5	37	140
0.65	10	141	150
0.46	10	117	150
0.32	10	56	149

Data 3. Data Pengamatan Lalat

Konsentrasi *pyrethrins*	Banyak lalat mati	Jumlah lalat
40	109	462
60	199	500
80	296	467
100	370	515
120	459	561
140	400	469
160	495	550
180	499	542
200	450	479
250	476	497
300	442	453

Lampiran 2. Nilai-Nilai Akar Polinomial *Hermitte*

Jumlah titik	X	w_i
2	$x_{\pm 1}=0.70710678$	$w_i=0.5$
3	$x_0=0$ $x_{\pm 1}=1.224745$	$w_i=1.181636$ $w_i=0.295409$
4	$x_{\pm 1}=0.52464762$ $x_{\pm 2}=1.65068012$	$w_i=0.80491409$ $w_i=0.08131284$
5	$x_0=0$ $x_{\pm 1}=0.958572$ $x_{\pm 2}=2.020183$	$w_i=0.945309$ $w_i=0.393619$ $w_i=0.019953$
6	$x_{\pm 1}=0.43607741$ $x_{\pm 2}=1.33584907$ $x_{\pm 3}=2.35060497$	$w_i=0.72462960$ $w_i=0.15706732$ $w_i=0.00453001$
7	$x_0=0$ $x_{\pm 1}=0.2029255775$ $x_{\pm 2}=0.3707655928$ $x_{\pm 3}=0.4745539562$	$w_i=0.2089795918$ $w_i=0.1909150253$ $w_i=0.1398526057$ $w_i=0.0647424831$
8	$x_{\pm 1}=0.38118699$ $x_{\pm 2}=1.15719371$ $x_{\pm 3}=1.98165676$ $x_{\pm 4}=2.93063742$	$w_i=0.66114701$ $w_i=0.20780233$ $w_i=0.01707798$ $w_i=0.00019960$
9	$x_0=0$ $x_{\pm 1}=0.1621267117$ $x_{\pm 2}=0.3066857164$ $x_{\pm 3}=0.4180155537$ $x_{\pm 4}=0.4840801198$	$w_i=0.1651196775$ $w_i=0.1561735385$ $w_i=0.1303053482$ $w_i=0.09032408035$ $w_i=0.04063719418$

Lampiran 2. (Lanjutan)

Jumlah titik	X	w_i
10	$x_{\pm 1}$ =0.34290133 $x_{\pm 2}$ =0.103661083 $x_{\pm 3}$ =1.75668365 $x_{\pm 4}$ =2.53273167 $x_{\pm 5}$ =3.43615912	w_i =0.61086263 w_i =0.24013861 w_i =0.03387439 w_i =0.00134365 w_i =0.00000746

Lampiran 3. Hasil Uji Non Multikolinearitas

Data 1. Data Pengamatan Kelulusan Mahasiswa di Amerika Serikat

a. Perhitungan nilai *VIF* untuk peubah X_1 (tahun kelulusan)

Pseudo R-Square	
Cox and Snell	.000
Nagelkerke	.000
McFadden	.000

Berdasarkan output regresi multinomial antara peubah X_1 (tahun kelulusan) sebagai peubah respon dengan X_2 (tahun kelulusan) dan X_3 (universitas) sebagai peubah bebas menggunakan SPSS, diperoleh nilai *Cox and Snell* sebesar 0. Berdasarkan persamaan (2.48), nilai *VIF* yang dihasilkan adalah sebagai berikut:

$$VIF_{X_3} = (1-R^2)^{-1} = (1-0)^{-1} = 1$$

b. Perhitungan nilai *VIF* untuk peubah X_2 (status kependudukan mahasiswa)

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	88.465[a]	.031	.041
a. Estimation terminated at iteration number 3 because parameter estimates changed by less than .001.			

Nilai *Cox dan Snell* yang diperoleh dari regresi logistik antara peubah X_2 (status kependudukan mahasiswa) sebagai peubah respon dengan peubah X_1 (tahun kelulusan) dan X_3 (universitas) sebagai

peubah bebas menggunakan SPSS adalah 0.031. Berdasarkan persamaan (2.48), maka nilai *VIF* yang diasilkan adalah sebagai berikut:

$$VIF_{X_2} = (1-R^2)^{-1} = (1-0.031)^{-1} = 1.03$$

Lampiran 3. (Lanjutan)

c. Perhitungan nilai *VIF* untuk peubah X_3 (universitas)

Pseudo R-Square	
Cox and Snell	.067
Nagelkerke	.076
McFadden	.032

Nilai *Cox and Snell* yang diperoleh dari regresi multinomial antara peubah X_3 (universitas) sebagai peubah respon dengan peubah X_2 (status kependudukan mahasiswa) dan X_1 (tahun kelulusan) sebagai peubah bebas sebesar 0.067. Nilai *VIF* yang diperoleh berdasarkan persamaan (2.48) adalah :

$$VIF_{X_1} = (1-R^2)^{-1} = (1-0.019)^{-1} = 1.07$$

Data 2. Data Pengamatan Kumbang

a. Perhitungan nilai *VIF* untuk peubah X_1 (konsentrasi *pyrethrin)*

Regression Analysis: X1 versus X2

```
The regression equation is
X1 = 0.893 - 0.0465 X2

Predictor        Coef   SE Coef      T      P
Constant       0.8933    0.1059   8.43  0.000
X2           -0.04651   0.02055  -2.26  0.047

S = 0.288421   R-Sq = 33.9%   R-Sq(adj) = 27.3%
```

Nilai R^2 yang diperoleh dari regresi linier antara peubah X_1 (konsentrasi *pyrethrin*) sebagai peubah respon dengan peubah X_2 (konsentrasi *piperonyl*) sebagai peubah bebas sebesar 33.9% = 0.339.

Nilai VIF yang diperoleh berdasarkan persamaan (2.48) untuk peubah X_1 (konsentrasi *pyrethrin*) adalah :

$$VIF_{X_1} = (1-R^2)^{-1} = (1-0.339)^{-1} = 1.513$$

Lampiran 3. (Lanjutan)

b. Pethitungan nilai *VIF* untuk peubah X_2 (konsentrasi *piperonyl*)

Regression Analysis: X2 versus X1

```
The regression equation is
X2 = 8.61 - 7.28 X1

Predictor     Coef  SE Coef      T      P
Constant     8.614    2.614   3.30  0.008
X1          -7.284    3.218  -2.26  0.047

S = 3.60928   R-Sq = 33.9%   R-Sq(adj) = 27.3%
```

Nilai R^2 yang diperoleh dari regresi linier antara peubah X_2 (konsentrasi *piperonyl*) sebagai peubah respon dengan peubah X_1 (konsentrasi *pyrethrin*) sebagai peubah bebas sebesar 33.9% = 0.339. Nilai VIF yang diperoleh berdasarkan persamaan (2.48) untuk peubah X_2 (konsentrasi *piperonyl*) adalah :

$$VIF_{X_1} = (1 - R^2)^{-1} = (1 - 0.339)^{-1} = 1.513$$

Lampiran 4. *Syntax* dan Output SAS 9.3.15 Untuk Hasil Pendugaan Parameter Model Regresi Logistik

Data 1. Data Pengamatan Kelulusan Mahasiswa di Amerika Serikat

Syntax SAS

```
data kelulusan;
input tahun_1 tahun_2 tahun_3 tahun_4 tahun_5 tahun_6
tahun_7 status_kependudukan univ_1 univ_2 y n;
datalines;
1   0   0   0   0   0   0   1   1   0   31  7422
0   1   0   0   0   0   0   1   1   0   177 7166
0   0   1   0   0   0   0   1   1   0   393 6759
0   0   0   1   0   0   0   1   1   0   484 6138
0   0   0   0   1   0   0   1   1   0   500 5506
0   0   0   0   0   1   0   1   1   0   399 4824
0   0   0   0   0   0   1   1   1   0   309 4277
0   0   0   0   0   0   0   1   1   0   210 3836
⋮   ⋮   ⋮   ⋮   ⋮   ⋮   ⋮   ⋮   ⋮   ⋮   ⋮   ⋮
⋮   ⋮   ⋮   ⋮   ⋮   ⋮   ⋮   ⋮   ⋮   ⋮   ⋮   ⋮
1   0   0   0   0   0   0   2   0   0   26  427
0   1   0   0   0   0   0   2   0   0   65  376
0   0   1   0   0   0   0   2   0   0   80  280
0   0   0   1   0   0   0   2   0   0   51  178
0   0   0   0   1   0   0   2   0   0   17  114
0   0   0   0   0   1   0   2   0   0   8   85
0   0   0   0   0   0   1   2   0   0   2   72
0   0   0   0   0   0   0   2   0   0   2   37
run;
proc genmod;
model y/n =tahun_1 tahun_2 tahun_3 tahun_4 tahun_5
tahun_6 tahun_7  status_kependudukan univ_1
univ_2/dist=bin link=logit;
run;
```

Lampiran 4. (Lanjtan)

Output SAS

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Deviance	37	358.8089	9.6975
Scaled Deviance	37	358.8089	9.6975
Pearson Chi-Square	37	333.5494	9.0148
Scaled Pearson X2	37	333.5494	9.0148
Log Likelihood		-18040.6176	
Full Log Likelihood		-316.7638	
AIC (smaller is better)		655.5276	
AICC (smaller is better)		662.8609	
BIC (smaller is better)		676.1108	

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	-2.6557	0.0781	-2.8088	-2.5026	1155.20	<.0001
tahun_1	1	-1.6657	0.0989	-1.8596	-1.4718	283.46	<.0001
tahun_2	1	-0.3033	0.0730	-0.4464	-0.1603	17.27	<.0001
tahun_3	1	0.4970	0.0671	0.3656	0.6284	54.92	<.0001
tahun_4	1	0.6738	0.0672	0.5421	0.8055	100.59	<.0001
tahun_5	1	0.6432	0.0688	0.5083	0.7780	87.34	<.0001
tahun_6	1	0.4904	0.0718	0.3497	0.6310	46.68	<.0001

Lampiran 4. (Lanjutan)

tahun_7	1	0.2837	0.0760	0.1347	0.4327	13.93	0.0002
status_kependudukan	1	0.5315	0.0355	0.4619	0.6012	223.77	<.0001
univ_1	1	-0.9254	0.0373	-0.9984	-0.8523	615.96	<.0001
univ_2	1	-0.8250	0.0252	-0.8744	-0.7756	1072.96	<.0001
Scale	0	1.0000	0.0000	1.0000	1.0000		

Data 2. Data Pengamatan Kumbang

Syntax SAS

```
data beetles;
input pyrethrins piperonyl_1 piperonyl_2 piperonyl_3
y n;
datalines;
1.5    1    0    0    138   150
1.06   1    0    0    75    149
0.75   1    0    0    32    150
⋮      ⋮    ⋮    ⋮    ⋮     ⋮
0.65   0    0    0    141   150
0.46   0    0    0    117   150
0.32   0    0    0    56    149
run;
proc genmod;
model y/n = pyrethrins piperonyl_1 piperonyl_2
piperonyl_3 / dist=bin link=logit;
run;
```

Lampiran 4. (Lanjutan)

Output SAS

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Deviance	7	48.8149	6.9736
Scaled Deviance	7	48.8149	6.9736
Pearson Chi-Square	7	50.3686	7.1955
Scaled Pearson X2	7	50.3686	7.1955
Log Likelihood		-872.7528	
Full Log Likelihood		-53.4570	
AIC (smaller is better)		116.9140	
AICC (smaller is better)		126.9140	
BIC (smaller is better)		119.3385	

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	-2.3085	0.1902	-2.6813	-1.9356	147.25	<.0001
pyrethrins	1	6.9845	0.3621	6.2748	7.6942	372.02	<.0001
piperonyl_1	1	-4.8931	0.2733	-5.4287	-4.3576	320.63	<.0001
piperonyl_2	1	-3.4766	0.2128	-3.8935	-3.0596	267.03	<.0001
piperonyl_3	1	-0.9344	0.1632	-1.2542	-0.6146	32.80	<.0001
Scale	0	1.0000	0.0000	1.0000	1.0000		

Lampiran 4. (Lanjutan)

Data 3. Data Pengamatan Lalat

Syntax SAS

```
data flies;
input pyrethrins y n;
datalines;
40      109     462
60      199     500
80      296     467
⋮       ⋮       ⋮
200     450     479
250     476     497
300     442     453
run;
proc genmod;
model y/n = pyrethrins / dist=bin link=logit;
run;
```

Output SAS

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Deviance	9	91.1026	10.1225
Scaled Deviance	9	91.1026	10.1225
Pearson Chi-Square	9	111.5718	12.3969
Scaled Pearson X2	9	111.5718	12.3969
Log Likelihood		-2286.0304	
Full Log Likelihood		-77.3318	
AIC (smaller is better)		158.6636	
AICC (smaller is better)		160.1636	
BIC (smaller is better)		159.4594	

Lampiran 4. (Lanjutan)

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	-1.5762	0.0866	-1.7459	-1.4065	331.31	<.0001
pyrethrins	1	0.0228	0.0008	0.0213	0.0243	873.59	<.0001
Scale	0	1.0000	0.0000	1.0000	1.0000		

Lampiran 5. Syntax dan Output SAS 9.3.15 Untuk Membentuk Model Regresi *Logistic-Normal*

Data 1. Data Pengamatan Kelulusan Mahasiswa di Amerika Serikat

Syntax SAS

```
data kelulusan;
input tahun_1 tahun_2 tahun_3 tahun_4 tahun_5 tahun_6
tahun_7 status_kependudukan universitas y n;
datalines;
1   0   0   0   0   0   0   1   1   31  7422
0   1   0   0   0   0   0   1   1   177 7166
0   0   1   0   0   0   0   1   1   393 6759
0   0   0   1   0   0   0   1   1   484 6138
0   0   0   0   1   0   0   1   1   500 5506
0   0   0   0   0   1   0   1   1   399 4824
0   0   0   0   0   0   1   1   1   309 4277
0   0   0   0   0   0   0   1   1   210 3836
1   0   0   0   0   0   0   1   2   14  3247
:   :   :   :   :   :   :   :   :   :   :
:   :   :   :   :   :   :   :   :   :   :
1   0   0   0   0   0   0   2   3   26  427
0   1   0   0   0   0   0   2   3   65  376
0   0   1   0   0   0   0   2   3   80  280
0   0   0   1   0   0   0   3   3   51  178
0   0   0   0   1   0   0   2   3   17  114
0   0   0   0   0   1   0   2   3   8   85
0   0   0   0   0   0   1   2   3   2   72
0   0   0   0   0   0   0   2   3   2   37
run;
proc nlmixed data=kelulusan;
parms beta0=-1.5 beta1=-1 beta2=-0.5 beta3=-0.3
beta4=0 beta5=0.3 beta6=0.5 beta7=1 beta8=1.5 s2u=2;
eta=beta0+beta1*tahun_1+beta2*tahun_2+beta3*tahun_3+b
eta4*tahun_4+beta5*tahun_5+beta6*tahun_6+beta7*tahun_
7+beta8*status_kependudukan+u;
expeta=exp(eta);
p=expeta/(1+expeta);
model y~binomial(n,p);
random u~normal (0,s2u)
subject=universitas;
run;
```

Lampiran 5. (lanjutan)

Outpus SAS

Fit Statistics	
-2 Log Likelihood	658.8
AIC (smaller is better)	678.8
AICC (smaller is better)	684.8
BIC (smaller is better)	676.8

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Gradient
beta0	-3.4097	0.9044	5	-3.77	0.0130	0.05	-5.7345	-1.0848	-0.00003
beta1	-1.6638	0.09896	5	-16.81	<.0001	0.05	-1.9182	-1.4095	9.973E-6
beta2	-0.3021	0.07302	5	-4.14	0.0090	0.05	-0.4898	-0.1144	-0.00003
beta3	0.4980	0.06708	5	7.42	0.0007	0.05	0.3256	0.6704	0.000032
beta4	0.6749	0.06720	5	10.04	0.0002	0.05	0.5022	0.8476	0.000041
beta5	0.6443	0.06883	5	9.36	0.0002	0.05	0.4674	0.8213	-0.00003
beta6	0.4912	0.07179	5	6.84	0.0010	0.05	0.3067	0.6758	-0.00015
beta7	0.2848	0.07602	5	3.75	0.0133	0.05	0.08938	0.4802	0.000012
beta8	0.4459	0.5715	5	0.78	0.4706	0.05	-1.0232	1.9149	-0.00005
s2u	0.4870	0.2835	5	1.72	0.1465	0.05	-0.2418	1.2159	-0.00009

Lampiran 5. (Lanjutan)

Data 3. Data Pengamatan Lalat

Syntax SAS

```
data beetles;
input pyrethrins piperonyl y n;
datalines;
1.5   0     138   150
1.06  0     75    149
0.75  0     32    150
1.1   0.25  129   151
0.78  0.25  65    151
0.55  0.25  19    150
0.8   2.5   143   149
0.57  2.5   112   150
0.4   2.5   37    140
0.65  10    141   150
0.46  10    117   150
0.32  10    56    149
run;
proc nlmixed data=beetles;
parms beta0=-1 beta1=1 s2u=2;
eta=beta0+beta1*pyrethrins+u;
expeta=exp(eta);
p=expeta/(1+expeta);
model y~ binomial (n,p);
random u~ normal (0,s2u)
subject=piperonyl;
run;
```

Output SAS

Fit Statistics	
-2 Log Likelihood	133.3
AIC (smaller is better)	139.3
AICC (smaller is better)	142.3
BIC (smaller is better)	137.5

Lampiran 5. (Lanjutan)

Parameter Estimates									
Parameter	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**	**Alpha**	**Lower**	**Upper**	**Gradient**
beta0	-4.6034	1.0040	3	-4.59	0.0195	0.05	-7.7985	-1.4083	-8.79E-7
beta1	6.9416	0.3616	3	19.20	0.0003	0.05	5.7909	8.0923	0.000078
s2u	3.7655	2.6924	3	1.40	0.2564	0.05	-4.8028	12.3338	6.986E-6

Lampiran 6. Syntax dan Output SAS 9.3.15 Untuk Hasil Pendugaan Parameter Model Regresi *Beta-Binomial*

Data 1. Data Pengamatan Kelulusan Mahasiswa di Amerika Serikat

Syntax SAS

```
data kelulusan;
input tahun_1 tahun_2 tahun_3 tahun_4 tahun_5 tahun_6
tahun_7 status_kependudukan univ_1 univ_2 y n;
datalines;
1   0   0   0   0   0   0   1   1   31  7422
0   1   0   0   0   0   0   1   1   177 7166
0   0   1   0   0   0   0   1   1   393 6759
0   0   0   1   0   0   0   1   1   484 6138
0   0   0   0   1   0   0   1   1   500 5506
0   0   0   0   0   1   0   1   1   399 4824
0   0   0   0   0   0   1   1   1   309 4277
0   0   0   0   0   0   0   1   1   210 3836
1   0   0   0   0   0   0   1   2   14  3247
:   :   :   :   :   :   :   :   :   :   :
:   :   :   :   :   :   :   :   :   :   :
1   0   0   0   0   0   0   2   3   26  427
0   1   0   0   0   0   0   2   3   65  376
0   0   1   0   0   0   0   2   3   80  280
0   0   0   1   0   0   0   3   3   51  178
0   0   0   0   1   0   0   2   3   17  114
0   0   0   0   0   1   0   2   3   8   85
0   0   0   0   0   0   1   2   3   2   72
0   0   0   0   0   0   0   2   3   2   37
run;
proc fmm data=kelulusan;
model y/n =tahun_1 tahun_2 tahun_3 tahun_4 tahun_5
tahun_6 tahun_7  status_kependudukan univ_1
univ_2/dist=betabinomial;
run;
```

Lampiran 6. (Lanjutan)

Output SAS

Fit Statistics	
-2 Log Likelihood	566.1
AIC (smaller is better)	590.1
AICC (smaller is better)	599.0
BIC (smaller is better)	612.6
Pearson Statistic	260.2

Parameter Estimates for 'Beta-Binomial' Model				
Effect	Estimate	Standard Error	z Value	Pr > |z|
Intercept	-2.7239	0.09052	-30.09	<.0001
tahun_1	-1.5060	0.1144	-13.17	<.0001
tahun_2	-0.1763	0.08701	-2.03	0.0428
tahun_3	0.6136	0.08051	7.62	<.0001
tahun_4	0.7640	0.08086	9.45	<.0001
tahun_5	0.6761	0.08346	8.10	<.0001
tahun_6	0.5037	0.08711	5.78	<.0001
tahun_7	0.2782	0.09241	3.01	0.0026
status_kependudukan	0.5075	0.03921	12.94	<.0001
univ_1	-0.8827	0.04166	-21.19	<.0001
univ_2	-0.8130	0.02780	-29.25	<.0001
Scale Parameter	7055.80	0		

Data 2. Data Pengamatan Kumbang

Syntax SAS

```
data beetles;
input pyrethrins piperonyl_1 piperonyl_2 piperonyl_3
y n;
datalines;
1.5    1     0     0     138   150
1.06   1     0     0     75    149
0.75   1     0     0     32    150
:      :     :     :     :     :
0.65   0     0     0     141   150
0.46   0     0     0     117   150
0.32   0     0     0     56    149
run;
proc fmm data=beetles;
model y/n = pyrethrins piperonyl_1 piperonyl_2
piperonyl_3 / dist=betabinomial;
run;
```

Output SAS

Fit Statistics	
-2 Log Likelihood	87.1149
AIC (smaller is better)	99.1149
AICC (smaller is better)	115.9
BIC (smaller is better)	102.0
Pearson Statistic	12.9351

Lampiran 6. (Lanjutan)

Parameter Estimates for 'Beta-Binomial' Model				
Effect	**Estimate**	**Standard Error**	**z Value**	**Pr > |z|**
Intercept	-2.3203	0.3940	-5.89	<.0001
pyrethrins	6.9995	0.7654	9.14	<.0001
piperonyl_1	-4.9206	0.5708	-8.62	<.0001
piperonyl_2	-3.4721	0.4231	-8.21	<.0001
piperonyl_3	-0.9405	0.3071	-3.06	0.0022
Scale Parameter	49.5675	26.9466		

Data 3. Data Pengamatan Lalat

Syntax SAS

```
data flies;
input pyrethrins y n;
datalines;
40     109    462
60     199    500
80     296    467
⋮      ⋮      ⋮
200    450    479
250    476    497
300    442    453
run;
proc fmm data=flies;
model y/n = pyrethrins / dist=betabinomial;
run;
```

Lampiran 6. (Lanjutan)

Output SAS

Fit Statistics	
-2 Log Likelihood	99.5724
AIC (smaller is better)	105.6
AICC (smaller is better)	109.0
BIC (smaller is better)	106.8
Pearson Statistic	15.3368

Parameter Estimates for 'Beta-Binomial' Model				
Effect	**Estimate**	**Standard Error**	**z Value**	**Pr > |z|**
Intercept	-1.5753	0.2440	-6.46	<.0001
pyrethrins	0.02262	0.002325	9.73	<.0001
Scale Parameter	77.8567	38.9229		