OVERDISPERSI PADA REGRESI LOGISTIK BINER

MENGGUNAKAN METODE BETA BINOMIAL

Heru Wibowo, Suyono, Widyanti Rahayu
Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Negeri Jakarta
Jl. Rawamangun Muka, Jakarta Timur 13220, Indonesia
heruwibowo29@gmail.com

Abstrak Analisis regresi logistik merupakan metode yang digunakan untuk mengetahui hubungan antara peubah prediktor (X) dan dengan peubah respon (Y) yang bersifat kategorik. Pada pemodelan data biner, masalah overdispersi sering terjadi yang disebabkan oleh keragaman antar peluang respon serta korelasi antar respon biner. Overdispersi merupakan suatu keadaan dimana ragam amatan lebih besar dibandingkan ragam dugaan sedangkan pada regresi logistik peubah respon (Y) diasumsikan berdistribusi binomial dengan ragam amatan sama dengan ragam dugaan. Metode Beta Binomial yang merupakan gabungan antara distribusi beta dan distribusi binomial cukup baik mengkoreksi keragaman pada data yang terindikasi terdapat overdispersi.

Kata kunci: Overdispersi, Regresi Logistik Biner, Beta Binomial..

1 Pendahuluan

Persoalan yang melibatkan dua variabel atau lebih yang diduga memiliki hubungan tertentu disebut regresi dari satu variabel atas variabel lain. Analisis Regresi merupakan suatu metode untuk menentukan suatu hubungan sebab akibat antara satu variabel dengan variabel lainnya. Variabel dalam analisis regresi dibedakan menjadi dua, yaitu variabel respon (*dependent variable*), dan prediktor (*independent variable*). Variabel respon biasanya disimbolkan dengan Y dan variabel bebas biasanya disimbolkan dengan X. Hubungan ini dapat dinyatakan dalam persamaan matematis yang dinamakan persamaan regresi. Persamaan regresi dapat berbentuk linear maupun non-linier. Analisis regresi logistik dapat digunakan mengetahui pengaruh beberapa variabel bebas yang dapat bersifat kategorik maupun numerik terhadap variabel respon yang mempunyai sifat kategorik. Variabel respon pada regresi logistik berdata biner berdistribusi binomial (Hosmer dan Lemeshow, 2000:7). Menurut Agresti (2002:123) regresi logistik adalah Generalized Linear Models (GLM) atau model linear umum dengan komponen acak binomial dan fungsi penghubung logit (invers dari fungsi logistik). Pendugaan parameter regresi logistik dilakukan dengan pendekatan kemungkinan maksimum dengan menggunakan metode iterasi Newton-Raphson. Menurut Hosmer dan Lemeshow (2000:140-141) regresi logistik sensitif terhadap kolinieritas (hubungan linierantar variabel) di antara peubah bebas yang terdapat di dalam model. Multikolinieritas yang terdapat didalam model dapat diperiksa dengan melihat nilai *Variance Ination Factors* (VIF). Asumsi dalam regresi logistik adalah ragam amatan sama dengan ragam dugaan. Apabila asumsi ini tidak terpenuhi akan menyebabkan overdispersi yaitu ragam amatan lebih besar daripada ragam dugaan (ragam distribusi binomial). Hal-hal penyebab overdispersi antara lain, kesalahan dalam penentuan fungsi penghubung, terjadinya korelasi antar pengamatan, terdapat pengelompokan serta menghilangkan kovariat yang penting (Hinde & Demetrio, 2007:5). Adanya overdispersi menyebabkan galat baku menjadi underestimate, sehingga diperoleh model regresi yang kurang tepat dan berpengaruh terhadap penarikan kesimpulan yang juga tidak tepat. Overdispersi pada regresi logistik dapat diselesaikan dengan beberapa cara yaitu memasukkan efek acak pada model lalu menggunakan metode William, serta menggunakan metode beta binomial. Berdasarkan latar belakang dalam penulisan kali ini akan digunakan metode beta binomial yang berupa gabungan antara sebaran beta dengan binomial untuk menyelesaikan masalah overdispersi pada regresi logistik.

2 Landasan Teori

2.1 Distribusi Binomial

Jika peubah acak bernoulli diamati pada *n* percobaan dimana antara setiap percobaan bersifat saling bebas dan peluang munculnya kejadian sukses pada setiap percobaan bersifat konstan yaitu *p*, maka banyaknya kejadian sukses *X* diantara *n* percobaan akan berdistribusi binomial dengan parameter *n* dan *p* serta dapat dinotasikan dengan X ~ Binomial (*n*,*p*). Fungsi kepadatan peluang untuk *X* adalah :

$$f(x) = \begin{cases} \binom{n}{x} p^x q^{n-x} & ; \quad x = 0,1,\ldots,n \\ 0 & ; \quad \text{selainnya} \end{cases}$$

Dengan mean-nya : E(X) = *np* dan varians-nya : Var(X) = np(1-p)

2.2 Analisis Regresi Linier Berganda

Analisis regresi linier berganda dipergunakan untuk mengetahui pengaruh antara satu peubah respon dengan banyak atau lebih dari satu peubah penjelas. Menurut Suyono (2015:26) untuk pengujian hipotesis dan untuk membuat interval konfidensi parameter galat-galat acak diasumsikan sebagai berikut :

1. Nilai rata-rata dari galat adalah nol
2. Variansi dari galat adalah konstan (homoskedastik)
3. Tidak terjadi autokorelasi pada galat
4. Galat berdistribusi normal

Persamaan umum dari analisis regresi linier berganda sebagai berikut :

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + \varepsilon_i$$

dimana Y=Variabel respon ; X_1 ,X_2 ,X_k =Variabel penjelas ; β_0=konstanta atau Intersep ; β_1 ,β_2 ,β_k = Koefsien dengan ε_i = sisa(error) untuk pengamatan ke-i yang diasumsikan berdistribusi normal yang saling bebas dan identik dengan rata-rata 0 (nol) dan variansi σ^2.

Model regresi linier berganda dapat diperoleh dengan melakukan estimasi terhadap parameter-parameternya menggunakan metode kuadrat terkecil (MKT) (ordinary least square/OLS) dan metode kemungkinan maksimum (MKM) (maximum likelihood estimation/MLE) dengan penduga β yaitu $\beta = (X^T X)^{-1} X^T Y$

2.3 Model Linier Umum

Model linear umum ialah pengembangan dari model linear klasik yang variabel responnya saling bebas. Model linier umum diasumsikan mengikuti distribusi keluarga eksponensial yang merupakan distribusi dengan sifat yang lebih umum dan adapat dituliskan sebagai berikut :

$$(g)(\mu_i) = (\eta_i)$$

Dengan $\eta_i = \beta_0 + \beta_1 X_{i1} + \ldots + \beta_k X_k$

Menurut (Gill J, 2001:31) Model Linear umum memiliki tiga komponen,
yaitu :

1. Komponen acak, yaitu komponen X_i , i=1,2,...n merupakan variabel acak yang yang distribusinya masuk ke dalam distribusi keluarga eksponensial dangan rata-rata μ_i , misal μ_i=E(Y_i), i=1,2,..., N
2. Komponen sistematik yang menunjukkan fungsi linier dari variabel penjelas. Misal X_{ij} menyatakan nilai penduga j dimana (j=1,2,...,p) untuk subjek i.
$$\eta_i = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_p X_{pi}$$
3. Fungsi penghubung, yaitu suatu fungsi *g* yang menghubungkan komponen acak dan komponen sistematik sehingga E(Y_i) = μ_i=$g^{-1}(\eta)$.

2.4 Analisis Regresi Logistik Biner

Analisis regresi logistik adalah analisis yang digunakan untuk menganalisis hubungan antara peubah respon (Y) yang bersifat kategorik atau numerik. Apabila untuk setiap objek amatan ke-i, i = 1, 2,..., n kategori dari peubah respon hanya memiliki dua kemungkinan nilai (misalnya $Y_i=1$ untuk kejadian sukses dan $Y_i=0$ untuk kejadian gagal) dan Y_i diasumsikan berdistribusi binomial dengan paramtere n_i dan p_i maka analisis regresi logistik tersebut dinamakan juga sebagai analisis regresi logistik biner. Model regresi logistik biner yang menunjukkan hubungan linier antara nilai-nilai dari transformasi logit dengan peubah penjelasnya diberikan oleh

$$\log\left(\frac{p_i}{1-p_i}\right)=\beta_0+\beta_1X_{1i}+\beta_2X_{2i}+...+\beta_kX_{ki}$$

2.4.1 Pendugaan Parameter Model Regresi Logistik Biner

Metode kuadrat terkecil (MKT) tidak dapat digunakan untuk menduga parameter pada regresi logistik biner karena dapat mengakibatkan diperolehnya taksiran yang terlalu rendah (underestimate) atau taksiran yang terlalu tinggi (overestimate) yang akan mempengaruhi perhitungan interval kepercayaan. Maka digunakan metode kemungkinan maksimum (MKM) dengan persamaannya

$$\frac{\partial \ln L(\beta)}{\partial\beta_1}=\sum_{i=1}^{n}X_{1i}\left\{Y_i-n_ip_i\right\}$$

Dalam mencari nilai dugaan β menggunakan metode newton raphson diperlukan beebrapa langkah yaitu

1. Memilih taksiran awal β
2. Menentukan vektor gradien (U') yang beranggotakan turunan pertama dari fungsi likelihood terhadap β
3. membentuk matriks Hessian yang beranggotakan turunan kedua fungsi likelihood terhadap β
4. Pada setiap iterasi ke-(t+1), dengan t=0,1,.. dihitung dengan nilai dugaan baru yaitu $\beta^{(t+1)}$ secara iteratif dengan rumus

$$\beta^{(t+1)}=\beta^{(t)}\left[H^{(t)}\right]^{-1}-U^{(t)}$$

Iterasi berlanjut hingga $\beta^{(t+1)}\approx\beta^{(t)}$,atau iterasi dihentikan jika selisih Antara $\beta^{(t+1)}$ Dengan $\beta^{(t)}$ sangatlah kecil.

2.4.2 Uji Signifikansi Parameter

a. Uji Serempak

Uji Serempak dilakukan untuk melihat pengaruh variabel penjelas secara bersama-sama terhadap variabel respon. Uji serempak melalui perbandingan rasio kemungkinan atau uji G memiliki hipotesis

$H_0:\beta_1=\beta_2=...=\beta_k=0$ dan H_1: Tidak Semua $\beta_i=0$

dimana

i = 1,2,....,k dengan k = banyaknya variabel bebas X dan

β_i = Parameter ke - i model regresi linier dengan statistik uji G

$$G=-2\ln\left[\frac{L(p_i)}{L(p_i)}\right]\sim\chi_k^2$$

dimana

p_i: Fungsi likelihood model dugaan

p_i: Fungsi likelihood model penuh

Jika $G\geq\chi_k^2$ maka H_0 ditolak yang berarti peubah penjelas X berpengaruh nyata terhadap peubah respon Y .

b. Uji Parsial

Uji Parsial (Uji Wald) dilakukan untuk melihat peranan masing-masing variabel

penjelas terhadap variabel respon. Hipotesis yang akan diuji
H_0 : $\beta_i = 0$ (koefsien β_i tidak signikan secara statistik)
H_1 : $\beta_i \neq 0$ (koefisien β_i signikan secara statistik)
dimana
$i = 1,2,....,k$ dengan k = banyaknya variabel bebas X dan
β_i = Parameter ke - i model regresi linier
Statistik Uji Wald :

$$W = \frac{(\beta_i)}{se(\beta_i)}$$

dengan :
$se(\beta_i)$: Galat baku penduga β_i , β_i : koefsien dugaan peubah prediktor
i : 0,1,2,...,n
statistik uji Wald mengikuti sebaran normal dengan kriteria keputusan menerima H_0 jika $W \leq Z$ berarti β_i tidak signifikan dan dapat disimpulkan bahwa peubah penjelas β_i tidak berpengaruh terhadap peubah respon.

2.5 Kesesuaian Model Logistik Biner

Pada analisis regresi logistik biner, pengujian kesesuaian model dugaan didasarkan pada statistik uji chi kuadrat *Pearson*. Statistik uji Pearson merupakan fungsi dari sisaan yaitu selisih dari nilai aktual dengan nilai dugaan dan didefinisikan sebagai berikut :

$$\chi^2 = \sum_{i=1}^{n} r(y_i, \hat{p}_i)^2$$

Nilai harapan dari sebaran χ^2 sama dengan derajat bebasnya jika nilai chi kuadrat Pearson jauh lebih besar dari derajat bebasnya, maka asumsi dari keragaman binom tidak terpenuhi.

Akaike's Information Criterion (AIC) atau Kriteria Informasi Akaike ialah ukuran relatif kebaikan untuk mengepas dari model statistik dengan membahas antara akurasi dan kompleksitas model. AIC didefiniskan dengan :

$$AIC(k) = -2\ln(\theta) + 2k$$

dimana k = Jumlah parameter dalam model statistik dan θ = nilai maksimal dari fungsi kemungkinan maksimum untuk model dugaan.

2.6 Rasio Odds

Hosmer dan Lemeshow (2000:53) menyatakan bahwa log odds untuk model regresi logistik dengan nilai peubah bebas lebih dari satu didefnisikan sebagai selisih antara penduga logit untuk suatu peubah bebas X_{ji} bernilai a dengan penduga logit untuk peubah bebas X_{ji} bernilai b dimana peubah bebas lainnya diasumsikan konstan, yaitu

$$\begin{aligned}
\ln\left[\widehat{\psi}(a,b)\right] &= \widehat{g}(X_{ji}=a) - \widehat{g}(X_{ji}=b) \\
&= \left(\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \ldots + \hat{\beta}_j \times a + \ldots + \hat{\beta}_k X_{ki}\right) \\
&\quad - \left(\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \ldots + \hat{\beta}_j \times b + \ldots + \hat{\beta}_k X_{ki}\right) \\
&= \widehat{\beta}_j (a-b)
\end{aligned}$$

3 Pembahasan

3.1 Overdispersi pada Regresi Logistik Biner

Saat model logistik linier dianggap sudah cukup sesuai tetapi nilai mean Pearson lebih dari satu maka asumsi keragaman berdasarkan distribusi binomial menjadi tidak valid. Masalah tersebut dikatakan sebagai Overdispersi. Overdispersi mengakibatkan nilai penduga ragam lebih besar daripada ragam di bawah asumsi binomial. Apabila nilai penduga ragam tersebut digunakan untuk menarik kesimpulan (pengujian hipotesis) maka cenderung akan menolak H_0. Menurut (Hinde dan Demetrio,2007:5) terdapat sejumlah keadaan yang mengindikasikan overdispersi, yaitu komponen sistematis dari model yang tidak tepat (pengelompokan, penghilangan kovariat penting), terjadinya korelasi antar amatan, adanya pencilan serta penggunaan fungsi penghubung (link function) yang tidak tepat. Apabila hal-hal yang dapat menyebabkan overdispersi tersebut dapat diatasi maka overdispersi dapat dijelaskan oleh keragaman antara peluang respon serta adanya korelasi antara peubah respon.

3.1.1 Keragaman Antar Peluang Respon

Pada kelompok objek percobaan yang kondisinya homogen, peluang respon atau peluang terjadi-nya kejadian sukses (p_i) dapat berbeda antara satu kelompok dengan lainnya. Hal ini disebabkan oleh suatu pengaruh yang tidak terobservasi. Diberikan $\text{var}(Y_i) = n_i p_i (1 - p_i)\{1 + (n_i - 1)\phi\}$, Jika tidak terdapat keragaman acak dalam peluang respon, (Y_i) akan menyebar Binomial ($n_i p_i$), artinya $\text{var}(Y_i) = n_i p_i (1 - p_i)$ sama dengan ragam dibawah asumsi binomial. Namun jika terdapat keragaman maka ragam dari *Yi* akan berubaha sebesar $\{1 + (n_i - 1)\phi\}$.

3.1.2 Korelasi Antar Respon Biner

Jika tidak ada korelasi antar amatan maka $\text{var}(Y_i) = n_i p_i (1 - p_i)$ sama dengan ragam di bawah asumsi binomial. Jika terdapat korelasi positif antar amatan, maka $\text{var}(Y_i) > n_i p_i (1 - p_i)$ sehingga dapat disimpulkan bahwa adanya korelasi antar amatan dalam kelompok yang sama dapat mengakibatkan keragaman yang lebih besar dibandingkan dengan keragamaan yang dihasilkan jika antar amatan diasumsikan saling bebas.

3.2 Metode Beta Binomial

Distribusi binomial mengasumsikan bahwa nilai parameternya telah ditetapkan, distribusi marjinal mengizinkan terjadinya variasi dibandingkan hanya distribusi binomial saja. Metode beta binomial ialah salah satu cara untuk mengatasi overdispersi yang terjadi pada model binomial biasa. Metode beta binomial ialah hasil gabungan dari distribusi beta dengan distribusi binomial dengan fungsi kepadatan peluangnya

$$P(Y = y \mid n, a, b) = \binom{n}{y} \frac{B(a + y, n + b - y)}{B(a, b)}; \text{y} = 0, 1, \ldots, n$$

Dengan Mean-nya $E(Y \mid n, a, b) = \dfrac{na}{a + b}$ dan varians-nya $Var(Y \mid n, a, b) = \dfrac{nab(a + b + n)}{(a + b)^2 (a + b + 1)}$

3.3 Aplikasi Pemodelan Overdispersi

Pemodelan Overdispersi pada penelitian ini menggunakan data primer yang terindikasi terdapat overdispersi didalamnya dan permasalahan overdispersi akan diselesaikan menggunakan metode beta binomial. Data primer dengan judul "Data nilai bahasa Indonesia mahasiswa matematika angkatan 2012-2015" diperoleh dari Pusat Teknologi Informasi dan Komputer Universitas Negeri Jakarta (PUSTIKOM UNJ) serta masih menerapkan kategori nilai A,B,C,D, dan E tanpa tanda (+) ataupun (-). Penggunaan variabel bebas dalam data mengacu pada jurnal acuan "Overdispersi dalam Regresi Logistik" karya Anang Kurnia yang telah dipublikasikan dalam Forum statistika dan komputasi tahun 2002.

3.3.1 Deskripsi Data

Data primer ini menjelaskan mengenai data diri atau faktor-faktor yang mungkin mempengaruhi mata kuliah Bahasa Indonesia yang bernilai A atau B. Peubah respon yang diamati dalam data ini adalah mata kuliah Bahasa Indonesia yang bernilai A atau B di semester pertama setiap angkatan dan bukan merupakan hasil pebaikan di semester selanjutnya. Pemilihan peubah respon ini didasarkan bahwa tidak ada mahasiswa yang telah memperoleh nilai A atau B pada mata kuliah bahasa indonesia untuk mengulang mata kuliah tersebut di semester selanjutnya.

Hipotesis selanjutnya yaitu peluang mahasiswa untuk mendapatkan nilai A atau B pada mata kuliah bahasa indonesia dapat berbeda antar satu mahasiswa dengan mahasiswa yang lainnya, hal tersebut terjadi karena disebabkan oleh adanya suatu pengaruh yang tidak terobservasi sehingga diharapkan data ini akan menunjukkan overdispersi. Data yang diperolah dalam penelitian ini sejumlah 206 data yang terbagi ke dalam 4 angkatan yaitu angkatan 2012 (45 orang), angkatan 2013 (51 orang), angkatan 2014 (34 orang), dan angkatan 2015 (76 orang). Populasi dari penelitian ini ialah mahasiswa matematika angkatan 2012-2015. Tahapan pengolahan data sebagai berikut

1. Deskripsi Data Nilai Bahasa Indonesia Profil dan Latar Belakang Mahasiswa Prodi Matematika 2012-2015
   Dijelaskan mengenai informasi umum dari data yang akan diolah beserta dengan hipotesisnya.

2. Penentuan dan Pengklasifikasian Variabel dependen dan Independen
   Data yang dijelaskan secara umum kemudian dijabarkan lebih lanjut tentang pengkategorian (beserta alasan) , jumlah persentase, dan info lainnya.
3. Uji Multikolinearitas
   Uji multikolinearitas dengan melihat nilai VIF, jika kurang dari 10 maka tidak terdapat multikolinearitas didalamnya

Tabel 3.1: Nilai VIF

Predictor	VIF
Constant	0,048
X_1(angkatan)	1,088
X_2(jenis kelamin)	1,052
X_3(jalur masuk)	1,083
X_4(asal sekolah)	1,165
X_5(nilai ijazah)	1,246
X_6(nilai Uan)	1,276
X_7(lama sekolah)	1,143
X_8(status tempat tingal)	1,084
X_9(Indeks prestasi semester)	1,156
X_{10}(Pendidikan Ayah)	2,341
X_{11}(Pekerjaan Ayah)	1,145
X_{12}(Penghasilan Ayah)	1,487
X_{13}(Pendidikan Ibu)	2,524
X_{14}(Pekerjaan Ibu)	1,715
X_{15}(Penghasilan Ibu)	1,933

4. Pendugaan Parameter Regresi Logistik

$$\begin{aligned} \text{logit}(\mu_i) = & - 1,810 + 0,674X_1 - 0,355X_2 + 0,089X_3 + 0,132X_4 + 0,493X_5 \\ & - 0,071X_6 + 1,436X_7 - 0,183X_8 + 1,066X_9 + 0,017X_{10} + 0,071X_{11} \\ & + 0,327X_{12} + 0,029X_{13} + 0,033X_{14} - 0,219X_{15} \end{aligned} \quad (3.32)$$

5. Pengecekan Overdispersi
   Model regresi logistik yang telah diduga kemudian di cek menggunakan statistik uji pearson

Tabel 3.2: Uji Pearson

Statistik Uji	Chi-Square	DF	Mean
Pearson	191,960	189	1,016

6. Uji Signifikansi Parameter
   1. Serempak
      Diperoleh nilai G=70,119 dan χ^2= 24,99579 kareana G> χ^2 maka terima H_0
   2. Parsial
      Parameter yang tidak signifikan (nilai parameter diatas= 0,05) pada model penuh regresi beta binomial dikeluarkan satu persatu dari model mulai dari nilai yang terbesar hingga ke yang terkecil secara berurutan. Parameter yang paling tidak signifikan dikeluarkan terlebih dahulu, kemudian dilakukan pendugaan parameter regresi beta binomial.

Tabel 3.3: Hasil Uji Parsial Koefisien Regresi Beta Binomial(Model Penuh)

Peubah Bebas	Koefisien	Galat Baku	Uji Wald	Nilai P
Konstanta	-1,810	1,510	-1,199	0,231
(X_1) Angkatan	**0,674**	**0,167**	**4,033**	**0,000**
(X_2) Jenis Kelamin	-0,355	0,385	-0,924	0,356
(X_3) Jalur Masuk	0,089	0,188	0,473	0,636
(X_4) Asal Sekolah	0,132	0,652	0,203	0,839
(X_5) Nilai Ijazah	0,493	0,272	1,811	0,070
(X_6) Nilai UAN	-0,071	0,261	-0,270	0,787
(X_7) Lama Sekolah	1,436	1,322	1,086	0,277
(X_8) Status Tempat Tinggal	-0,183	0,172	-1,065	0,287
(X_9) Indeks Prestasi Semester	**1,066**	**0,228**	**4,682**	**0,000**
(X_{10}) Pendidikan Ayah	0,017	0,184	0,092	0,927
(X_{11}) Pekerjaan Ayah	0,071	0,056	1,271	0,204
(X_{12}) Penghasilan Ayah	**0,327**	**0,153**	**2,136**	**0,033**
(X_{13}) Pendidikan Ibu	0,029	0,190	0,151	0,880
(X_{14}) Pekerjaan Ibu	0,033	0,042	0,781	0,435
(X_{15}) Penghasilan Ibu	-0,219	0,192	-1,140	0,254

7. Pendugaan Parameter Regresi Beta Binomial

Eliminasi Peubah		Model Regresi
X_{10}	$\text{logit}(\mu_i) =$	$-1,800 + 0,676X_1 - 0,356X_2 + 0,089X_3$ $+0,140X_4 + 0,494X_5 - 0,070X_6 + 1,447X_7$ $-0,181X_8 + 1,067X_9 + 0,071X_{11} + 0,323X_{12}$ $+0,0388X_{13} + 0,03X_{14} - 0,218X_{15}$
X_{13}	$\text{logit}(\mu_i) =$	$-1,732 + 0,674X_1 - 0,352X_2 + 0,083X_3$ $+0,129X_4 + 0,500X_5 - 0,084X_6 + 1,459X_7$ $-0,183X_8 + 1,068X_9 + 0,071X_{11} + 0,309X_{12}$ $+0,035X_{14} - 0,192X_{15}$
X_4	$\text{logit}(\mu_i) =$	$-1,720 + 0,672X_1 - 0,351X_2 + 0,081X_3$ $+0,506X_5 - 0,082X_6 + 1,424X_7 - 0,185X_8$ $+1,076X_9 + 0,072X_{11} + 0,306X_{12} + 0,036X_{14}$ $-0,191X_{15}$
X_6	$\text{logit}(\mu_i) =$	$-1,781 + 0,679X_1 - 0,342X_2 + 0,087X_3$ $+0,478X_5 + 1,407X_7 - 0,186X_8 + 1,053X_9$ $+0,073X_{11} + 0,310X_{12} + 0,035X_{14} - 0,196X_{15}$
X_3	$\text{logit}(\mu_i) =$	$-1,758 + 0,683X_1 - 0,351X_2 + 0,485X_5$ $+1,423X_7 - 0,190X_8 + 1,056X_9 + 0,072X_{11}$ $+0,304X_{12} + 0,037X_{14} - 0,191X_{15}$
X_{14}	$\text{logit}(\mu_i) =$	$-1,677 - 0,668X_1 - 0,342X_2 + 0,513X_5$ $+1,548X_7 - 0,215X_8 + 1,063X_9 + 0,082X_{11}$ $+0,268X_{12} - 0,268X_{15}$
X_2	$\text{logit}(\mu_i) =$	$-1,887 + 0,655X_1 + 0,522X_5 + 1,595X_7$ $-0,215X_8 + 1,063X_9 + 0,076X_{11} + 0,287X_{12}$ $-0,273X_{15}$
X_8	$\text{logit}(\mu_i) =$	$-2,041 + 0,671X_1 + 0,522X_5 + 1,562X_7$ $+1,040X_9 + 0,080X_{11} + 0,262X_{12} - 0,248X_{15}$
X_7	$\text{logit}(\mu_i) =$	$-0,518 + 0,682X_1 + 0,590X_5 + 1,985X_9$ $+0,084X_{11} + 0,273X_{12} - 0,252X_{15}$
X_{15}	$\text{logit}(\mu_i) =$	$-0,700 + 0,676X_1 + 0,533X_5 + 0,920X_9$ $+0,083X_{11} + 0,236X_{12}$
X_{11}	$\text{logit}(\mu_i) =$	$-0,106 + 0,682X_1 + 0,547X_5 + 0,890X_9$ $+0,282X_{12}$
X_5	$\text{logit}(\mu_i) =$	$0,402 + 0,611X_1 + 0,900X_9 + 0,273X_{12}$

8. Uji Kesesuain Model
   Uji Kesesuaian model dilakukan dengan melihat nilai statistik Uji Pearson dan membandingkannya dengan χ^2. Apabila nilai statistik uji Pearson lebih kecil dari nilai chi-square tabelnya maka model yang terbentuk telah sesuai.

Tabel 3.5: Uji Kesesuaian Model

Model	Pearson	DF	χ^2	Keterangan
Model Penuh	191,96	189	222,076	sesuai
Tanpa X_{10}	191,866	190	223,160	sesuai
Tanpa X_{10}, X_{13}	191,727	188	220,991	sesuai
Tanpa X_{10}, X_{13}, X_4	189,611	188	220,991	sesuai
Tanpa X_{10}, X_{13}, X_4, X_6	187,784	185	217,735	sesuai
Tanpa $X_{10}, X_{13}, X_4, X_6, X_3$	176,504	181	213,391	sesuai
Tanpa $X_{10}, X_{13}, X_4, X_6, X_3$ X_{14}	163,276	178	210,130	sesuai
Tanpa $X_{10}, X_{13}, X_4, X_6, X_3$ X_{14}, X_2	159,597	171	202,513	sesuai
Tanpa $X_{10}, X_{13}, X_4, X_6, X_3$ X_{14}, X_2, X_8	161,587	167	198,154	sesuai
Tanpa $X_{10}, X_{13}, X_4, X_6, X_3$ X_{14}, X_2, X_8, X_7	161,237	167	198,154	sesuai
Tanpa $X_{10}, X_{13}, X_4, X_6, X_3$ $X_{14}, X_2, X_8, X_7, X_{15}$	131,544	149	178,485	sesuai
Tanpa $X_{10}, X_{13}, X_4, X_6, X_3$ $X_{14}, X_2, X_8, X_7, X_{15}, X_{11}$	94,622	94	117,632	sesuai
Tanpa $X_{10}, X_{13}, X_4, X_6, X_3$ $X_{14}, X_2, X_8, X_7, X_{15}, X_{11}, X_5$	65,9985	63	82,529	sesuai

9. Pemilihan Model Terbaik
   Model terbaik dengan melihat nilai AIC terkecil dari model model yang tidak lagi mengalami overdispersi

Tabel 3.6: Nilai AIC dari Model Regresi Beta Binomial

Model	n	SSE	$\widehat{\sigma^2}$	K	AIC Value
Model Penuh	206	32,615	0,158	15	-347,682
Tanpa X_{10}	206	32,622	0,158	14	-349,637
Tanpa X_{10}, X_{13}	206	32,626	0,158	13	-351,610
Tanpa X_{10}, X_{13}, X_4	206	32,636	0,158	12	-353,547
Tanpa X_{10}, X_{13}, X_4, X_6	206	32,646	0,158	11	-355,486
Tanpa $X_{10}, X_{13}, X_4, X_6, X_3$	206	32,664	0,159	10	-357,370
Tanpa $X_{10}, X_{13}, X_4, X_6, X_3$ X_{14}	206	32,850	0,159	9	-358,204
Tanpa $X_{10}, X_{13}, X_4, X_6, X_3$ X_{14}, X_2	206	32,938	0,160	8	-359,652
Tanpa $\mathbf{X_{10}, X_{13}, X_4, X_6, X_3}$ $\mathbf{X_{14}, X_2, X_8}$	**206**	**33,188**	**0,161**	**7**	**-360,095**
Tanpa $X_{10}, X_{13}, X_4, X_6, X_3$ X_{14}, X_2, X_8, X_7	206	33,562	0,163	6	-359,781
Tanpa $X_{10}, X_{13}, X_4, X_6, X_3$ $X_{14}, X_2, X_8, X_7, X_{15}$	206	33,933	0,165	5	-359,522
Tanpa $X_{10}, X_{13}, X_4, X_6, X_3$ $X_{14}, X_2, X_8, X_7, X_{15}, X_{11}$	206	34,487	0,167	4	-358,182
Tanpa $X_{10}, X_{13}, X_4, X_6, X_3$ $X_{14}, X_2, X_8, X_7, X_{15}, X_{11}, X_5$	206	35,383	0,172	3	-354,901

10. Interpretasi dengan Rasio Odds

Tabel 3.8: Odds Ratio

Parameter Model	Odds Ratio	Lower	Upper
(X_1) Angkatan	0,51	0,37	0,70
(X_5) Nilai Ijazah	1,68	1,02	2,77
(X_7) Lama Sekolah	4,77	0,40	6,41
(X_9) Indeks Prestasi Semester	2,83	1,88	4,25
(X_{11}) Pekerjaan Ayah	1,08	0,97	1,20
(X_{12}) Penghasilan Ayah	0,77	0,60	0,98
(X_{15}) Penghasilan Ibu	0,78	0,60	1,02

1. Parameter X_1, Angkatan, memiliki nilai odds Ratio terkecil dibanding parameter lainnya. Niai odds ratio variabel X_1 sebesar 0,51 dalam melihat pengaruh terhadap Nilai Mata Kuliah Bahasa Indonesia dengan batas bawah selang sebesar 0,37 dan batas atasnya sebesar 0,7.
2. Parameter Lama Sekolah, X_7 memiliki nilai odds Ratio terbesar dibanding parameter lainnya. Nilai odds ratio variabel X_7 sebesar 4,77 dalam melihat pengaruh terhadap nilai Mata Kuliah Bahasa Indonesia dengan batas bawah selang sebesar 0,4 dan batas atasnya sebesar 6,41.
3. Parameter X_5 yaitu parameter nilai ijazah dan parameter X_9 yaitu parameter indeks prestasi semester merupakan kedua parameter yang tidak memuat angka nol pada batas nilai *lower* dan *upper*. Hal ini berarti bahwa parameter X_5 dan X_9 bersifat signifikan atau bermakna apabila dibandingkan oleh parameter lain di dalam model terbaik yang menyatakan pengaruh terhadap nilai Mata Kuliah Bahasa Indonesia

4 Kesimpulan

1. Model regresi yang variabel responnya berdistribusi binomial disebut model regresi Logistik Biner. Model regresi Logistik Biner yang menunjukkan hubungan linier antara nilai-nilai dari transformasi logit dengan peubah penjelas diberikan oleh

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki}$$

Distribusi binomial memiliki asumsi $Var(Y_i) = n_i p_i (1 - p_i)$.

2. Overdispersi pada regresi Logistik Biner terjadi ketika nilai (δ> 0) sehingga mengakibatkan nilai ragam yang dihasilkan lebih besar daripada nilai ragam di bawah asumsi binomial yaitu $\text{var}(Y_i) > n_i p_i (1-p_i)\delta$ Model regresi Beta Binomial merupakan gabungan dari distribusi beta dengan distribusi binomial. Regresi Beta Binomial digunakan untuk menduga data binom dengan memodelkan keragaman peluang respon melalui sebaran beta, dimana fungsi kepadatan peluang beta binomial yaitu

$$P(Y = y \mid n, a, b) = \binom{n}{y} \frac{B(a+y, n+b-y)}{B(a,b)}; \text{y} = 0, 1, ..., n$$

3. Model Beta Binomial mampu mengatsi masalah overdispersi, hal ini terbukti dari statistik uji pearson dari semula 1,016 menjadi 0,968.

Daftar Pustaka

Collet, D.2003. "*Modelling Binary Data 2nd edition*". Chapman Hall : London.

Ennies, D.M, Bi, Jian,1997, "The Beta-Binomial Model: Accounting for Intertrial Variation in Replicated Diference and Preference Tests", Journal of Sensory Studies vol.13 , pp.389-412

Gill,Jeff.2001."Generalized Linier Models : A Unified Approach.". Sage Publications Inc: London.

Gujarati, D.N dan Porter, C.D.2003. "*Basic Econometrics 5thedition*.". McGraw-Hill: New York.

Hajarisman, Nusar,1998, "Kajian Perbandingan Model Regresi Beta Binomial dengan Model Regresi Logistik dan Penerepannya untuk Menduga Pola Kelulusan Mahasiswa TPB-IPB", Thesis pada Program Pascasarjana, Program studi Statistika, IPB Bogor , Thesis tidak dipublikasikan.

Halim, Izzudin,2012, "Penyelesaian Masalah Overdispersi pada Regresi Poisson Menggunakan Regresi Binomial Negatif", Skripsi pada FMIPA Universitas Negeri Jakarta , Tidak Diterbitkan

Hosmer, D.W dan Lemeshow, S.2000. "*Applied Logistic Regression 2 edition*."Wiley: New York.

Kurnia, Anang,2002, "Overdispersi dalam Regresi Logistik", Forum Statistika dan Komputasi,pp. 11-17

Kutner, M.H dkk.2005. "*Applied Linear Statistical Models 5 edition*.". McGraw-Hill: New York.

Lee, Peter M.2012. "Bayesian Statistics: An Introduction, 4th edition.". Wiley:New York.

Ningsih, Retno,2011, "Overdispersi pada Regresi Logistik Biner", Skripsi padaFMIPA Universitas Negeri jakarta , Tidak Diterbitkan

Senja, D.R.V, dkk,2013, "Penanganan Overdispersi Menggunakan Regresi Beta Binomial pada Regresi Logistik",Jurnal FMIPA Universitas Brawijaya Malang , Tidak Diterbitkan

Yuniana, D.R,2014 "Pemilihan Metode Beta-Binomial dan Logistic-Normal ", Jurnal FMIPA Universitas Brawijaya , Tidak Diterbitkan

William, D.A,1982, "Extra-Binomial Variation in Logistic Linear Models", Journal Royal Statistical Society , pp.31, 144-148