

Perbandingan Metode Regresi Multilevel dan Beta *Generalized Linear Mixed Models* pada Data Longitudinal Capaian IPK Mahasiswa

Gusti Tasya Meilania¹, Utami Dyah Syafitri^{2*}, I Made Sumertajaya³

^{1,2,3}Departemen Statistika, IPB University, Indonesia
e-mail: utamids@apps.ipb.ac.id

Diajukan: 1 Oktober 2024, Diperbaiki: 31 Oktober 2024, Diterima: 13 Nopember 2024

Abstrak

Penelitian ini membandingkan kinerja model Beta *Generalized Linear Mixed Model* (Beta GLMM) dengan Regresi Multilevel pada data Indeks Prestasi Kumulatif (IPK) mahasiswa. Data IPK yang digunakan dalam penelitian ini tampak asimetris ke sisi kiri atau memiliki ekor kiri yang lebih panjang yang mencerminkan kecenderungan mahasiswa memperoleh nilai yang lebih besar daripada rata-rata IPK keseluruhan. Hal ini mengindikasikan bahwa data tidak berdistribusi normal, melainkan diduga berdistribusi Beta. Tujuan dari penelitian ini adalah melakukan perbandingan terhadap metode regresi multilevel dan Beta GLMM untuk melihat faktor-faktor yang memengaruhi IPK mahasiswa setiap semester. Data yang digunakan adalah data longitudinal karena IPK dan beberapa peubah lainnya pada setiap mahasiswa diamati per semester. Pendekatan Beta GLMM digunakan karena Beta GLMM menggabungkan antara pendekatan *Linear Mixed Model* (LMM) dengan *Generalized Linear Model* (GLM). Berdasarkan analisis yang dilakukan, diperoleh hasil bahwa metode Beta GLMM memiliki nilai *Akaike Information Criterion* (AIC) yang lebih rendah dibandingkan metode regresi multilevel. Adapun faktor-faktor yang mempengaruhi capaian IPK mahasiswa berdasarkan analisis Beta GLMM diantaranya semester mahasiswa, SKS mahasiswa setiap semester, status perkawinan, jalur masuk kuliah, sumber biaya pendidikan (beasiswa), interaksi semester dengan status perkawinan, dan interaksi antara semester dengan jalur masuk kuliah. Selain itu, diketahui bahwa proporsi keragaman IPK yang dapat dijelaskan oleh perbedaan antar mahasiswa adalah sebesar 0.837. Hal ini menunjukkan bahwa 83.7% dari total variasi IPK dapat dijelaskan oleh perbedaan antar mahasiswa (Level 2), sedangkan sisanya 16.3% dijelaskan oleh variasi pada setiap mahasiswa setiap semester (Level 1).

Kata Kunci: Beta GLMM, data longitudinal, IPK, regresi multilevel, struktur bersarang

Abstract

This study compares the performance of the Beta Generalized Linear Mixed Model (Beta GLMM) and Multilevel Regression on students' Grade Point Averages (GPAs). The GPA data used in this study appears to be skewed to the left or has a longer left tail, indicating that the data is not normally distributed but rather suspected to be Beta-distributed. It led to a comparison between Multilevel Regression and Beta Generalized Linear Mixed model (GLMM) methods to identify the effects of GPA per semester. Type of data in this research is longitudinal data, since data consist of GPA and other explanatory variables of each student. Beta GLMM approach considered used because it combines Linear Mixed Model (LMM) with Generalized Linear Model (GLM). According to the analysis, Beta GLMM was found to have a lower Akaike Information Criterion (AIC) value than the multilevel regression. The factors influencing students' GPA, based on the Beta GLMM analysis, include the students' semester, the number of credit units taken each semester, marital status, admission pathway, education funding source (scholarship/not), interaction between semester and marital status, and interaction between semester and admission pathway. Additionally, the proportion of GPA variance explained by differences between students was 0.837. It

indicates that 83.7% of the total IPK variation can be explained by differences between students (Level 2), while the remaining 16.3% is explained by variation within each student per semester (Level 1).

Keywords: *Beta GLMM, longitudinal data, GPA, multilevel regression, nested structure*

1 Pendahuluan

Data longitudinal merupakan data dengan desain pengukuran berulang karena dikumpulkan dari lebih dari satu individu serta diamati dan dicatat secara berurutan selama periode waktu tertentu. Ketika data longitudinal memiliki struktur bersarang, pengukuran yang dilakukan pada level yang lebih rendah berada dalam unit yang lebih besar. Dalam data dengan struktur bersarang, individu yang berada dalam kelompok yang sama cenderung memiliki karakteristik yang serupa, sehingga pengukuran pada level yang lebih rendah tidak saling bebas [1]. Keragaman antara individu dan kelompok dalam peubah respon terhadap peubah penjelas dari waktu ke waktu mengindikasikan adanya heterogenitas yang dapat terjadi pada setiap tingkatan atau level [2].

Model regresi multilevel merupakan salah satu bagian dari *Linear Mixed Models* (LMM) yang mempertimbangkan adanya struktur data bersarang, dimana beberapa peubah menunjukkan [3]. Regresi multilevel memungkinkan untuk mengidentifikasi dan menganalisis heterogenitas antar individu dan antar kelompok, sehingga memungkinkan untuk menentukan komponen acak dalam setiap level analisis. Model ini memiliki asumsi kenormalan dan linieritas terhadap residual. Akan tetapi, data longitudinal tidak selalu berdistribusi normal. Dalam beberapa kasus, data longitudinal justru memiliki peubah respon yang tidak berdistribusi normal yang menunjukkan bahwa asumsi kenormalan telah terlanggar.

Generalized Linear Models (GLM) adalah perluasan dari regresi linier yang dapat menangani peubah respon dengan berbagai jenis distribusi tidak normal. Untuk menganalisis data longitudinal yang memungkinkan model untuk menangani variasi antar kelompok atau unit pengamatan pada data yang tidak mengikuti distribusi normal digunakan *Generalized Linear Mixed Models* (GLMM). GLMM menggabungkan dua metode statistik dasar yaitu LMM dan GLM sehingga dapat digunakan untuk menganalisis data yang bersifat hirarki dengan mempertimbangkan korelasi di dalam kelompok tersebut. Dalam GLMM, penggabungan LMM dan GLM memungkinkan untuk memodelkan peubah respon yang tidak berdistribusi normal dengan menggunakan fungsi hubung yang sesuai [4].

Untuk mengatasi data longitudinal yang melibatkan pengamatan berulang pada subjek yang sama dari waktu ke waktu, telah dikembangkan GLM dengan peubah respon berdistribusi Beta (Beta GLM). Pengamatan berulang pada subjek yang sama cenderung berkorelasi, GLMM

memungkinkan penambahan efek acak yang memperhitungkan variabilitas antar individu. Sehingga penambahan efek acak pada Beta GLM akan menghasilkan Beta GLMM [5].

Pada bidang pendidikan, data berstruktur tersarang sangat umum untuk ditemukan, salah satunya data capaian Indeks Prestasi Kumulatif (IPK) mahasiswa. IPK merupakan data dengan desain pengukuran berulang karena IPK akan diamati selama beberapa semester. Data IPK seringkali menunjukkan distribusi dengan ekor yang memanjang ke arah nilai IPK rendah dan puncak di dekat nilai IPK yang tinggi. Ini berarti bahwa sebagian besar mahasiswa memiliki IPK yang tinggi dan mendekati nilai maksimum. Akibatnya, distribusi data IPK tidak mengikuti distribusi normal [6]. Karakteristik sebaran data IPK ini dianggap lebih sesuai dengan distribusi Beta yang lebih efektif dalam menangani masalah tepi interval, yang sering muncul dalam data dengan batasan nilai, seperti IPK.

Penelitian terkait perbandingan antara LMM dan Beta GLMM terhadap analisis data longitudinal sudah pernah dilakukan sebelumnya dan diperoleh hasil bahwa GLMM lebih baik dalam memodelkan data dengan peubah respon yang tidak berdistribusi normal serta berada pada suatu rentang nilai tertentu [7]. Penelitian terhadap analisis data IPK sebelumnya pernah dilakukan oleh Tampil *et al.* menggunakan metode regresi logistik. Dalam penelitian tersebut diperoleh hasil bahwa program studi dan tempat tinggal mahasiswa teruji signifikan terhadap IPK [8]. Penelitian terbaru dilakukan oleh Riswan *et al.* dengan membandingkan model regresi multilevel, model regresi multilevel dengan transformasi logaritma natural, dan model regresi logistik biner multilevel terhadap analisis data IPK. Metode terbaik yang diperoleh adalah model regresi logistik biner multilevel dengan peubah yang teruji signifikan diantaranya semester, SKS, jalur masuk, jenis kelamin, sumber biaya pendidikan, usia, dan status perkawinan [6].

Berdasarkan penjabaran yang telah dijelaskan, penelitian ini memiliki kontribusi baru dibandingkan penelitian sebelumnya dengan melakukan perbandingan antara metode regresi multilevel dan Beta GLMM, yang belum diterapkan dalam konteks analisis data longitudinal capaian IPK mahasiswa. Dari hasil analisis yang dilakukan, akan diidentifikasi faktor-faktor yang mempengaruhi capaian IPK mahasiswa.

2 Metode Penelitian

2.1 Data Penelitian

Data yang digunakan dalam penelitian ini berupa data sekunder 619 mahasiswa FMIPA program pascasarjana tahun masuk 2019 yang diperoleh dari Direktorat Administrasi dan Pendidikan (DAP) IPB. Data penelitian ini berstruktur data longitudinal karena dilakukan pengamatan secara berulang tiap semester terhadap setiap mahasiswa.

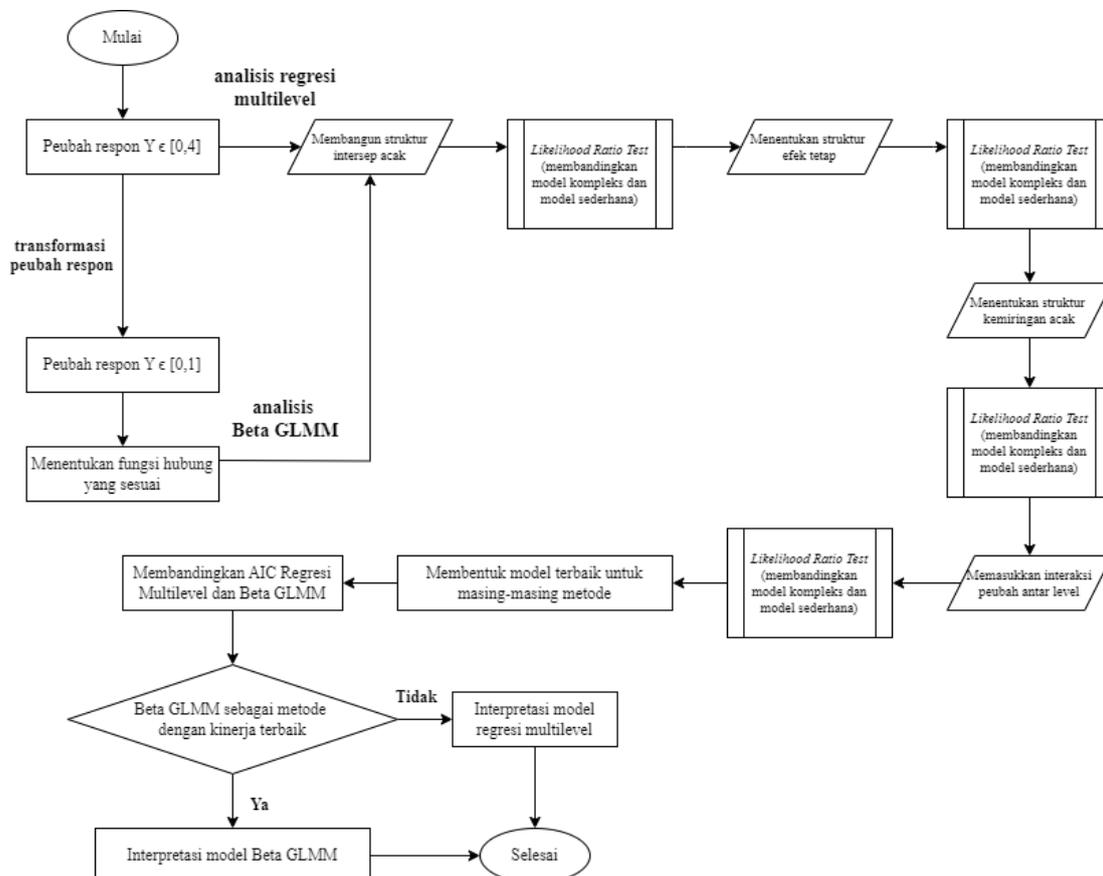
Tabel 1. Data Penelitian

| | Keterangan | Tipe Data |
|-------------------------|---------------------------------|-----------|
| Peubah Respon | | |
| Y | Indeks Prestasi Kumulatif (IPK) | Rasio |
| Peubah Penjelas Level-1 | | |
| X ₁ | Semester | Rasio |
| X ₂ | Satuan Kredit Semester (SKS) | Rasio |
| Peubah Penjelas Level-2 | | |
| V ₁ | Usia | Rasio |
| V ₂ | Status perkawinan | Nominal |
| V ₃ | Jalur masuk kuliah | Nominal |
| V ₄ | Sumber biaya pendidikan | Nominal |
| V ₅ | Jenis kelamin | Nominal |

Untuk melakukan analisis dengan metode Beta-GLMM, akan dilakukan transformasi terhadap peubah respon sebagai berikut [9].

$$Y^* = \frac{(Y - a)}{(b - a)} \tag{1}$$

dengan a adalah nilai minimum dari peubah respon (Y) dan b adalah nilai maksimum dari peubah respon (Y). Analisis dalam penelitian ini dilakukan menggunakan software R dengan kode sintaks yang disimpan dalam repositori Github [10]. Berikut diagram alir terkait tahapan analisis yang dilakukan dalam penelitian ini.



Gambar 1. Diagram Alir Pembentukan Model

2.2 Regresi Multilevel

Analisis multilevel merupakan salah satu bagian dari model linier campuran. Model multilevel digunakan untuk mengetahui hubungan antar peubah pada tingkat yang berbeda. Model regresi multilevel secara umum dapat diformulasikan melalui matriks dan vektor yang berbentuk model linier campuran (*Linear Mixed Model*) sebagai berikut [11].

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon} \quad (2)$$

dengan $\mathbf{u} \sim \text{MVN}(\mathbf{0}, \mathbf{G})$ dan $\boldsymbol{\varepsilon} \sim \text{MVN}(\mathbf{0}, \mathbf{R})$

notasi \mathbf{Y} adalah peubah respon, \mathbf{X} merupakan matriks desain untuk efek tetap, $\boldsymbol{\beta}$ merupakan vektor koefisien untuk efek tetap, \mathbf{Z} merupakan matriks desain untuk efek acak, \mathbf{u} merupakan vektor efek acak, dan $\boldsymbol{\varepsilon}$ adalah vektor parameter residual. Sementara itu, \mathbf{G} adalah matriks kovarian dari efek acak dalam \mathbf{u} dan \mathbf{R} adalah matriks kovarian berdefinit positif untuk semua residual dalam $\boldsymbol{\varepsilon}$.

2.3 Distribusi Beta

Dalam distribusi Beta, terdapat dua parameter yang dapat memberikan dugaan parameter terkait perubahan dalam rata-rata dan dispersi dari suatu peubah respon. Berikut fungsi densitas dari peubah acak berdistribusi Beta [12].

$$f(y; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} y^{a-1} (1-y)^{b-1} \quad , 0 < y < 1 \quad (3)$$

dengan $a > 0, b > 0, \Gamma(\cdot)$ merupakan fungsi gamma.

Pada nilai tengah dan ragam dilakukan reparameterisasi dimana parameter α dan β dapat didefinisikan menjadi nilai tengah $\mu = \frac{\alpha}{\alpha+\beta}$ dan parameter presisi $\phi = \alpha + \beta$, sehingga fungsi densitas, nilai tengah, ragam distribusi Beta menjadi sebagai berikut [10].

$$f(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1} (1-y)^{(1-\mu)\phi-1} \quad (4)$$

$$E(y) = \mu \quad (5)$$

$$\text{Var}(y) = \frac{\mu(1-\mu)}{1+\phi} \quad (6)$$

2.4 Beta Generalized Linear Mixed Models

GLMM merupakan salah satu pendekatan yang dirancang untuk menangani data dengan sifat distribusi tidak normal serta mempertimbangkan struktur pengulangan atau pengelompokan dalam data melalui keragaman antar subjek yang tidak teramati (*unobserved heterogeneity*) seperti efek acak [14]. Berikut model regresi Beta ($\mu\phi, (1-\mu)\phi$) dengan fungsi hubung yang digunakan adalah fungsi logit [15].

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} \quad (7)$$

dengan fungsi hubung yang digunakan adalah fungsi logit.

$$g(\mu) = \log\left(\frac{\mu}{1-\mu}\right) = \boldsymbol{\eta} \quad (8)$$

Adanya penambahan efek acak pada model regresi Beta akan menghasilkan model Beta-GLMM sebagai berikut [16].

$$\log\left(\frac{\mu}{1-\mu}\right) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} \quad \text{dengan } \mathbf{u} \sim N(\mathbf{0}, \mathbf{G}) \quad (9)$$

$$\boldsymbol{\mu} = \frac{\exp(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u})}{1 + \exp(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u})} \quad (10)$$

dengan \mathbf{X} merupakan matriks desain untuk efek tetap, $\boldsymbol{\beta}$ merupakan vektor koefisien untuk efek tetap, \mathbf{Z} merupakan matriks desain untuk efek acak, \mathbf{u} merupakan vektor efek acak, \mathbf{G} merupakan matriks kovarian berdefinit positif dari efek acak dengan σ_0^2 merupakan keragaman dari intersep acak dan σ_1^2 merupakan keragaman dari kemiringan acak.

2.5 Evaluasi Model

Pemilihan model regresi multilevel dan Beta-GLMM dapat dilakukan dengan membandingkan nilai *Akaike Information Criterion* (AIC) pada masing-masing model. Semakin kecil nilai AIC, maka model dikatakan semakin cocok digunakan pada data. AIC dapat diformulasikan sebagai berikut.

$$AIC = -2 \log\left(\frac{L_{\text{tersarang}}}{L_{\text{penuh}}}\right) + 2k \quad (11)$$

dengan $L_{\text{tersarang}}$ adalah nilai fungsi likelihood pada model tersarang, L_{penuh} adalah nilai fungsi likelihood pada model penuh dan k adalah jumlah parameter yang di estimasi [17].

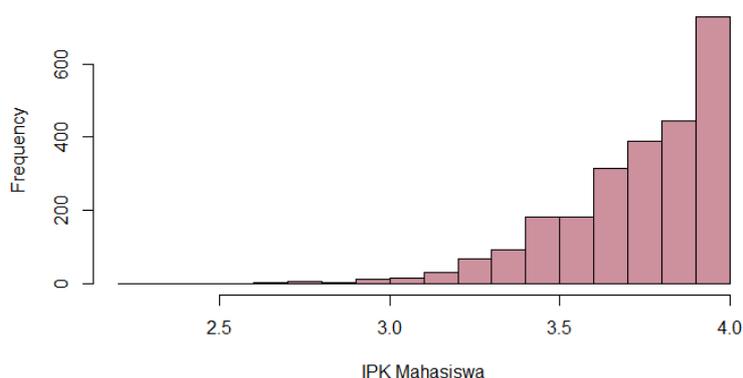
2.6 *Intraclass Correlation Coefficient* (ICC)

Korelasi intrakelas mengukur seberapa besar proporsi keragaman total yang dapat dijelaskan oleh perbedaan antar kelompok dalam data. ICC penting untuk menentukan apakah struktur kelompok dalam data perlu diperhatikan dalam analisis [17]. Korelasi intrakelas dapat diperoleh untuk setiap level kelompok pada data bersarang. Jika keragaman dalam kelompok dilambangkan dengan σ_{eij}^2 dan keragaman antar kelompok dilambangkan dengan σ_{ui}^2 , maka keragaman peubah respon akibat perbedaan karakteristik level 2 dinyatakan sebagai berikut [18].

$$\rho = \frac{\sigma_{u_{0j}}^2}{\sigma_{u_{0j}}^2 + \sigma_{eij}^2} \quad (12)$$

3 Hasil dan Pembahasan

Data IPK yang akan digunakan dalam penelitian ini ditemukan terlihat tidak menyebar mengikuti distribusi normal, melainkan mengikuti distribusi beta. Hal ini dapat dijelaskan dengan adanya *skewness* negatif dalam data, yang berarti distribusi data condong ke sisi kiri atau memiliki ekor kiri yang lebih panjang yang mencerminkan kecenderungan mahasiswa memperoleh nilai yang lebih besar daripada rata-rata IPK keseluruhan (Gambar 2).



Gambar 2. Sebaran IPK Mahasiswa

3.1 Pembentukan Model Regresi Multilevel dan Beta GLMM

Untuk memperoleh model terbaik dari metode Regresi Multilevel dan Beta GLMM, dilakukan tahapan analisis sebagai berikut.

1. Penentuan struktur intersep acak

Berdasarkan hasil analisis yang dilakukan, diperoleh nilai koefisien korelasi intraklas sebesar 0.837. Hal ini menunjukkan bahwa proporsi keragaman IPK yang dapat dijelaskan oleh mahasiswa tanpa dipengaruhi oleh faktor lainnya adalah sebesar 83.7 %.

2. Pemilihan struktur efek tetap

Berdasarkan hasil pengujian *Likelihood Ratio* pada pemilihan struktur efek tetap, diketahui bahwa peubah jenis kelamin (V_5) menunjukkan hasil tidak signifikan pada taraf nyata 5% maupun 10%. Oleh karena itu, struktur efek tetap yang akan digunakan pada analisis selanjutnya adalah peubah penjelas pada level-1 yaitu semester (X_1) dan SKS (X_2), serta peubah penjelas pada level-2 yaitu usia (V_1), status perkawinan (V_2), jalur masuk (V_3), dan beasiswa (V_4).

3. Pemilihan struktur kemiringan acak

Berdasarkan hasil pengujian *Likelihood Ratio* pada pemilihan struktur kemiringan acak, diketahui bahwa kemiringan acak untuk peubah X_1 dan X_2 menunjukkan hasil yang signifikan pada taraf nyata 5%. Akan tetapi, pada hasil analisis diketahui bahwa kemiringan acak SKS mahasiswa (X_2) memiliki korelasi sebesar 1.0 (korelasi sempurna) terhadap intersep acak. Hal ini

menunjukkan bahwa keragaman pada kemiringan acak (X_2) sepenuhnya dapat dijelaskan oleh keragaman dalam intersep acak. Oleh karena itu, struktur kemiringan acak yang akan digunakan pada analisis selanjutnya adalah X_1 .

4. Penambahan interaksi peubah antar level

Berdasarkan hasil pengujian *Likelihood Ratio* pada penambahan interaksi peubah antar level, diketahui bahwa penambahan interaksi X_1 dan V_2 , X_1 dan V_3 , serta X_1 dan V_4 menunjukkan hasil yang signifikan pada taraf nyata 5%. Oleh karena itu, dilakukan pemodelan dengan melibatkan interaksi peubah penjelas antar level tersebut. Berikut model regresi multilevel yang terbentuk.

$$Y_{ij} = (\beta_{0i} + u_{0i}) + (\beta_{1i} + u_{1i})X_{1ij} + \beta_{2i}X_{2ij} + \beta_{3i} V_{1i} + \beta_{4i} V_{2i} + \beta_{5i}V_{3i} + \beta_{6i}V_{4i} + \beta_{7i}X_{1ij} * V_{2i} + \beta_{8i}X_{1ij} * V_{3i} + \beta_{9i}X_{1ij} * V_{4i} + \epsilon_{ij} \tag{13}$$

Sementara itu, model Beta GLMM yang terbentuk adalah sebagai berikut.

$$\eta_{ij} = (\beta_{0i} + u_{0i}) + (\beta_{1i} + u_{1i})X_{1ij} + \beta_{2i}X_{2ij} + \beta_{3i} V_{1i} + \beta_{4i} V_{2i} + \beta_{5i}V_{3i} + \beta_{6i}V_{4i} + \beta_{7i}X_{1ij} * V_{2i} + \beta_{8i}X_{1ij} * V_{3i} + \beta_{9i}X_{1ij} * V_{4i} \tag{14}$$

3.2 Evaluasi Model dan Interpretasi Model Terbaik

Tabel 2. Perbandingan AIC terhadap Model Regresi Multilevel dan Beta GLMM

| Model | AIC |
|--------------------|---------|
| Regresi Multilevel | -4103.9 |
| Beta GLMM | -8968.9 |

Berdasarkan Tabel 2., diperoleh hasil bahwa metode Beta GLMM merupakan metode terbaik karena memiliki nilai AIC yang lebih rendah dibandingkan metode regresi multilevel. Oleh karena itu, selanjutnya akan dilakukan interpretasi pada metode Beta GLMM untuk mengidentifikasi faktor-faktor yang mempengaruhi capaian IPK mahasiswa.

Tabel 3. Uji Signifikansi Parameter Beta GLMM

| | Nilai p |
|---|---------|
| Intersep | 0.0000 |
| Semester (X_1) | 0.0066 |
| SKS (X_2) | 0.0000 |
| Usia (V_1) | 0.7827 |
| Status Perkawinan (V_2) | 0.0002 |
| Jalur Masuk (V_3) | 0.0000 |
| Beasiswa (V_4) | 0.0065 |
| Semester : Status Perkawinan ($X_1: V_2$) | 0.0002 |
| Semester : Jalur Masuk ($X_1: V_3$) | 0.0059 |
| Semester : Beasiswa ($X_1: V_4$) | 0.1184 |

Berdasarkan Tabel 3., diketahui bahwa model terbaik yang diperoleh dari serangkaian hasil pengujian dalam pembentukan model Beta GLMM adalah model dengan peubah penjelas

X_1, X_2, V_1, V_2, V_3 , dan V_4 , adanya intersep acak dan kemiringan acak X_1 disertai dengan penambahan interaksi peubah antar level yaitu X_1 dengan V_2 , X_1 dengan V_3 , dan X_1 dengan V_4 . Sebagai ilustrasi, model Beta GLMM yang terbentuk untuk mahasiswa ke-1 dan ke-2 adalah sebagai berikut.

$$\hat{\eta}_{1j} = 1.1228 + 0.0824 X_{1ij} + 0.0127 X_{2ij} + 0.0028 V_{1i} + 0.7299 V_{2i} - 0.9039 V_{3i} + 0.4587 V_{4i} - 0.0962 X_{1ij} * V_{2i} + 0.0831 X_{1ij} * V_{3i} - 0.0394 X_{1ij} * V_{4i} \quad (15)$$

$$\hat{\eta}_{2j} = 2.9081 + 0.0415 X_{1ij} + 0.0127 X_{2ij} + 0.0028 V_{1i} + 0.7299 V_{2i} - 0.9039 V_{3i} + 0.4587 V_{4i} - 0.0962 X_{1ij} * V_{2i} + 0.0831 X_{1ij} * V_{3i} - 0.0394 X_{1ij} * V_{4i} \quad (16)$$

Efek tetap menggambarkan hubungan antara peubah penjelas dengan peubah respon yang diasumsikan konstan untuk seluruh pengamatan. Peubah X_1, X_2, V_1, V_2, V_4 , dan interaksi antara $X_1:V_3$ teruji signifikan pada taraf nyata 5% dengan koefisien positif. Hal ini berarti peningkatan peubah-peubah tersebut akan meningkatkan log odds dari proporsi IPK. Sedangkan peubah V_3 dan interaksi antara $X_1:V_2$ teruji signifikan pada taraf nyata 5% dengan koefisien negatif. Hal ini mengindikasikan bahwa jika peubah-peubah tersebut meningkat, maka log odds dari proporsi IPK akan menurun. Sementara itu, peubah penjelas usia (V_1) dan interaksi peubah semester (X_1) dengan beasiswa (V_4) tidak teruji signifikan pada taraf nyata 5% maupun 10%.

Efek acak memberikan informasi tentang keragaman antar mahasiswa. Intersep acak menunjukkan bahwa mahasiswa memiliki *baseline* IPK yang beragam karena karena intersep acak menangkap perbedaan karakteristik yang tidak teramati antar mahasiswa (u_{0i}). Kemiringan acak X_1 mengindikasikan bahwa pengaruh semester terhadap IPK tidak sama untuk setiap mahasiswa. Keragaman u_{1i} menunjukkan bahwa beberapa mahasiswa mungkin mendapatkan peningkatan IPK yang lebih besar per semester, sementara mahasiswa lain mungkin mendapatkan peningkatan yang lebih kecil atau bahkan penurunan. Selain itu, diperoleh nilai parameter dispersi (ϕ) sebesar 82.4 yang mengindikasikan adanya keragaman tambahan dalam data yang tidak dijelaskan oleh model. Hal ini menunjukkan bahwa ada faktor lain yang menyebabkan variasi IPK mahasiswa yang tidak dijelaskan oleh faktor-faktor yang digunakan dalam model.

4 Simpulan

Data IPK yang digunakan dalam penelitian ini terlihat condong ke sisi kiri atau memiliki ekor kiri yang lebih panjang, dimana hal ini mengindikasikan bahwa data tidak berdistribusi normal, melainkan diduga berdistribusi Beta sehingga dilakukan perbandingan terhadap metode regresi multilevel dan Beta GLMM. Berdasarkan analisis yang dilakukan, diperoleh hasil bahwa metode Beta GLMM memiliki nilai AIC yang lebih rendah yaitu sebesar -8968.9 dibandingkan metode regresi multilevel sebesar -4103.9 . Hal ini menunjukkan bahwa jika tujuan analisis adalah

untuk menentukan model yang lebih efisien dan sesuai, terutama ketika data memiliki ukuran sampel besar dan sebaran peubah respon asimetris ke salah satu sisi, maka Beta GLMM adalah metode yang lebih baik dibandingkan dengan regresi multilevel. Adapun faktor-faktor yang mempengaruhi capaian IPK mahasiswa berdasarkan hasil model terbaik diantaranya semester mahasiswa, SKS mahasiswa setiap semester, status perkawinan, jalur masuk kuliah, sumber biaya pendidikan (beasiswa), interaksi semester dengan status perkawinan dan interaksi antara semester dengan jalur masuk kuliah. Selain itu, diketahui bahwa 83.7% dari total variasi IPK dapat dijelaskan oleh perbedaan antar mahasiswa (Level 2), sedangkan sisanya 16.3% dijelaskan oleh variasi pada setiap mahasiswa disetiap semester (Level 1).

5 Daftar Pustaka

- [1] J. J. Hox, *Multilevel Analysis - Techniques and Applications*, Second Edi. Taylor & Francis Group, 2010. doi: 10.4028/www.scientific.net/AMM.556-562.5968.
- [2] T. A. B. Snijders and R. J. Bosker, *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*, Second Edi. SAGE, 2011.
- [3] H. Goldstein, *Multilevel Statistical Models*. London: Institute of Education, 1999. doi: 10.1177/003591577106400124.
- [4] W. Cao, "Longitudinal Data Prediction in EHR: Comparison of GLMM and Machine Learning Methods," University of Rhode Island, 2019.
- [5] D. Zimprich, "Modeling change in skewed variables using mixed beta regression models," *Res. Hum. Dev.*, vol. 7, no. 1, pp. 9–26, 2010, doi: 10.1080/15427600903578136.
- [6] Riswan, U. D. Syafitri, and M. N. Aidi, "Analisis Regresi Multilevel pada Indeks Prestasi Kumulatif Mahasiswa Pascasarjana," vol. 12, no. 1, pp. 81–94, 2024.
- [7] M. Sciandra and I. C. Spera, "A model-based approach to Spotify data analysis: a Beta GLMM," *J. Appl. Stat.*, 2020, doi: 10.1080/02664763.2020.1803810.
- [8] Y. Tampil, H. Komalig, and Y. Langi, "Analisis Regresi Logistik Untuk Menentukan Faktor-Faktor Yang Mempengaruhi Indeks Prestasi Kumulatif (IPK) Mahasiswa FMIPA Universitas Sam Ratulangi Manado," *d'CARTESIAN*, vol. 6, no. 2, p. 56, 2017, doi: 10.35799/dc.6.2.2017.17023.
- [9] L. A. Gray and M. H. Alava, "A command for fitting mixture regression models for bounded dependent variables using the beta distribution," *Stata J.*, vol. 18, no. 1, pp. 51–75, 2018, doi: 10.1177/1536867x1801800105.
- [10] G. T. Meilania, "Multilevel Regression vs Beta GLMM." 2024. [Online]. Available:

- <https://github.com/tsymln/multilevel-reg-vs-beta-glmm.git>
- [11] B. T. West, K. Welch, A. Galecki, and B. Gillespie, *Linear mixed models: A practical guide using statistical software*. USA: Chapman and Hall/CRC, Taylor & Francis Group, 2007. doi: 10.1201/9781003181064.
- [12] R. V Hogg, J. W. McKean, and A. T. Craig, *Introduction to Mathematical Statistics*, Eight Edit. Pearson, 2019.
- [13] S. L. P. Ferrari and F. Cribari-Neto, “Beta regression for modelling rates and proportions,” *J. Appl. Stat.*, vol. 31, no. 7, pp. 799–815, 2004, doi: 10.1080/0266476042000214501.
- [14] G. Bonanno, F. Domma, and L. Errico, “Income Inequality and Inner Areas - A Study on the Italian Case,” 2022.
- [15] M. Smithson and J. Verkuilen, “A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables,” *Psychol. Methods*, vol. 11, no. 1, pp. 54–71, 2006, doi: 10.1037/1082-989X.11.1.54.
- [16] J. Verkuilen and M. Smithson, “Mixed and mixture regression models for continuous bounded responses using the beta distribution,” *J. Educ. Behav. Stat.*, vol. 37, no. 1, pp. 82–113, 2012, doi: 10.3102/1076998610396895.
- [17] J. J. Hox and J. K. Roberts, *Hanbook of Advanced Multilevel Analysis*. Taylor & Francis Group, 2011.
- [18] H. Goldstein, *Multilevel Statistical Models*, 4th Editio. United Kingdom: John Wiley & Sons. Inc, 2011. doi: 10.2307/1534624.