ESTIMASI: Journal of Statistics and Its Application
Vol. 7, No. 1, Januari, 2026, Hal. 159-180
DOI: 10.20956/ejsa.v7i1.44426

Pemilihan Model Regresi Logistik Ordinal Terbaik Menggunakan Metode *Stepwise* (Studi Kasus: Data Indeks Prestasi Kumulatif Lulusan Program Sarjana FMIPA Unmul)

Selsi[1], Darnah[2*], Sri Wahyuningsih[3]
[123]Program Studi S1 Statistika, Jurusan Matematika, FMIPA, Universitas Mulawarman, Samarinda, 75119, Indonesia
*** Corresponding author, email: darnah.98@gmail.com**

Abstract

Ordinal logistic regression is one of the statistical methods used to model response variables with two or more categories that have levels. This study aims to model the cumulative grade point average data of undergraduate graduates of the Faculty of Mathematics and Natural Sciences, Mulawarman University in 2023 using ordinal logistic regression. Estimation of ordinal logistic regression model parameters is done using the Maximum Likelihood Estimation (MLE) method and Newton-Raphson iteration. The best model selection was conducted using the stepwise method based on the smallest Akaike Information Criterion (AIC) value and significant predictor variables. The selection process started with six predictor variables, then gradually eliminated three predictor variables because they were not significant and did not reduce the AIC value. The stepwise stage stopped at the model with three significant predictor variables that had an AIC value of 349,22. The results showed that the factors that had a significant effect on the cumulative grade point average of undergraduate graduates of the Faculty of Mathematics and Natural Sciences, Mulawarman University based on the best ordinal logistic regression model were study program, age, and admission pathway.

Keywords*: Grade Point Average, Stepwise Method, MLE, Newton-Raphson, Ordinal Logistic Regression.*

Abstrak

Regresi logistik ordinal merupakan salah satu metode statistika yang digunakan untuk memodelkan variabel respon dengan dua atau lebih kategori yang memiliki tingkatan. Penelitian ini bertujuan untuk memodelkan data indeks prestasi kumulatif lulusan program sarjana FMIPA Unmul tahun 2023 menggunakan regresi logistik ordinal. Penaksiran parameter model regresi logistik ordinal dilakukan dengan menggunakan metode *Maximum Likelihood Estimation* (MLE) dan iterasi Newton-Raphson. Pemilihan model terbaik dilakukan menggunakan metode *stepwise* berdasarkan nilai *Akaike Information Criterion* (AIC) terkecil dan variabel prediktor yang signifikan. Proses seleksi dimulai dari enam variabel prediktor, kemudian secara bertahap dieliminasi tiga variabel prediktor karena tidak signifikan dan tidak menurunkan nilai AIC. Tahapan *stepwise* berhenti pada model dengan tiga variabel prediktor signifikan yang memiliki nilai AIC sebesar 349,22. Hasil penelitian menunjukkan bahwa faktor-faktor yang berpengaruh signifikan terhadap indeks prestasi kumulatif lulusan program sarjana FMIPA Unmul berdasarkan model regresi logistik ordinal terbaik adalah program studi, usia, dan jalur peneriman.

Kata Kunci: Indeks Prestasi Kumulatif, Metode *Stepwise*, MLE, Newton-Raphson, Regresi Logistik Ordinal.

1. Pendahuluan

Analisis regresi adalah teknik statistika yang bertujuan untuk memodelkan dan menjelaskan hubungan antara variabel respon dan variabel prediktor [1]. Model regresi terbagi berdasarkan skala pengukuran variabel responnya. Jika variabel responnya bersifat kuantitatif, maka yang digunakan adalah regresi linier. Tetapi jika variabel responnya berupa data kualitatif atau kategori maka yang digunakan adalah regresi logistik [2]. Regresi logistik dibagi menjadi dua jenis berdasarkan banyaknya kategori variabel respon, yaitu regresi logistik dikotomus untuk variabel respon dua kategori dan berdistribusi Bernoulli sedangkan regresi logistik polikotomus untuk variabel dua kategori atau lebih dan berdistribusi multinomial. Selanjutnya, regresi logistik polikotomus terdiri dari dua jenis model, yaitu model regresi logistik multinomial dan model regresi logistik ordinal, yang masing-masing digunakan berdasarkan sifat kategori pada variabel respon [3].

Salah satu pemilihan model regresi terbaik dalam penelitian ini menggunakan metode *stepwise*. Metode *stepwise* adalah kombinasi dari metode *forward* dan *backward*. Proses seleksi dilakukan beberapa tahapan dengan memasukkan semua variabel prediktor kedalam model, dan secara bertahap mengeliminasi variabel prediktor yang tidak signifikan. Proses ini dilakukan terus menerus hingga tidak ada lagi variabel untuk ditambahkan atau dieliminasi [4].

Metode *stepwise* digunakan untuk memilih variabel prediktor yang paling signifikan dalam membentuk model regresi logistik ordinal terhadap indeks prestasi kumulatif (IPK). IPK merupakan indikator utama pencapaian akademik mahasiswa dan mencerminkan keberhasilan proses pembelajaran, Nilai IPK dipengaruhi oleh berbagai faktor, baik internal maupun eksternal [5]. Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Universitas Mulawarman (UNMUL) memiliki peran penting dalam meningkatkan kualitas akademik lulusan mahasiswa melalui pencapaian IPK. Dengan demikian, penelitian ini menjadi penting untuk dilakukan guna memperoleh pemahaman yang mendalam mengenai faktor-faktor yang mempengaruhi IPK mahasiswa di FMIPA UNMUL sebagai upaya peningkatan kualitas pendidikan.

2. Material dan Metode

Data yang digunakan dalam penelitian ini adalah data sekunder, yaitu data indeks prestasi kumulatif lulusan program sarjana FMIPA Unmul yang diperoleh dari Bagian Akademik dan Kemahasiswaan FMIPA Unmul tahun 2023 dengan jumlah observasi 251 lulusan. Variabel yang digunakan dalam penelitian ini terdiri dari variabel respon dan variabel prediktor. Indeks Prestasi Kumulatif sebagai variabel respon (Y), Program Studi (X_1), Jenis Kelamin (X_2), Usia (X_3), Asal Daerah (X_4), Jalur Penerimaan (X_5), dan Pekerjaan Orang Tua (X_6).

2.1 Regresi Logistik Ordinal

Regresi logistik ordinal adalah salah satu metode analisis yang digunakan untuk mengetahui hubungan variabel respon dengan variabel prediktor, dimana variabel responnya bersifat polikotomus dengan skala ordinal [6]. Model logit ini memiliki sifat ordinal dari respon Y yang dituangkan dalam peluang kumulatif. Model logit kumulatif merupakan suatu metode yang didapatkan dengan cara membandingkan peluang kumulatif yaitu peluang kurang dari atau sama dengan kategori ke-*j* dengan peluang lebih besar dari kategori respon ke-*j* [7].

Jika variabel respon model regresi logistik ordinal dilambangkan dengan Y dan mempunyai kategori sebanyak J berskala ordinal dan $\mathbf{x}_i$ menyatakan vektor variabel prediktor sebanyak p pada pengamatan ke-*i*, yaitu $\mathbf{x}_i = \begin{bmatrix} x_{i1} & x_{i2} & \dots & x_{ip} \end{bmatrix}^T$ dengan $i = 1, 2, \dots, n.$ Jadi peluang kumulatif yang didapatkan $P(Y_i \le j \mid \mathbf{x}_i)$ merupakan persamaan umum regresi logistik ordinal dituliskan pada Persamaan (1)

$$P(Y_i \le j \mid \mathbf{x}_i) = \frac{\exp\left(\beta_{0j} + \mathbf{x}_i^T \boldsymbol{\beta}\right)}{1 + \exp\left(\beta_{0j} + \mathbf{x}_i^T \boldsymbol{\beta}\right)} \tag{1}$$

dimana $j = 1, 2, \dots, J-1, \beta_{0j}$ yaitu parameter intersep memenuhi kondisi $\beta_{01} \le \beta_{02} \le \dots \le \beta_{0J-1}$ dan $\boldsymbol{\beta} = \begin{bmatrix} \beta_1 & \beta_2 & \dots & \beta_p \end{bmatrix}^T$ termasuk vektor parameter regresi [8].

Pendugaan parameter regresi dilakukan dengan cara menguraikan menggunakan transformasi logit dari $P(Y_i \le j \mid \mathbf{x}_i)$ dituliskan pada Persamaan (2)

$$Logit\left[P(Y_i \le j \mid \mathbf{x}_i)\right] = g_j(\mathbf{x}_i) = \ln\left[\frac{P(Y_i \le j \mid \mathbf{x}_i)}{P(Y_i > j \mid \mathbf{x}_i)}\right] = \beta_{0j} + \mathbf{x}_i^T \boldsymbol{\beta} \tag{2}$$

Misalkan $\pi_j(\mathbf{x}_i) = P(Y_i = j \mid \mathbf{x}_i)$ menyatakan peluag variabel pengamatan ke-*i* mempunyai kategori ke-*j* terhadap $\mathbf{x}_i$, maka:

$$P(Y_i \le j \mid \mathbf{x}_i) = P(Y_i = 1 \mid \mathbf{x}_i) + P(Y_i = 2 \mid \mathbf{x}_i) + \dots + P(Y_i = j \mid \mathbf{x}_i) = \pi_1(\mathbf{x}_i) + \pi_2(\mathbf{x}_i) + \dots + \pi_j(\mathbf{x}_i) \tag{3}$$

Sehingga peluang untuk masing-masing kategori variabel respon dapat dituliskan pada Persamaan (4)

$$\pi_j(\mathbf{x}_i) = P(Y_i = j \mid \mathbf{x}_i) = P(Y_i \le j \mid \mathbf{x}_i) - P(Y_i \le j-1 \mid \mathbf{x}_i) \qquad j = 1, 2, \dots, J \tag{4}$$

Berdasarkan Persamaan (2) hingga persamaan (4), diperoleh pada Persamaan (5)

$$\pi_j(\mathbf{x}_i) = \frac{\exp\left(\beta_{0j} + \mathbf{x}_i^T \boldsymbol{\beta}\right)}{1 + \exp\left(\beta_{0j} + \mathbf{x}_i^T \boldsymbol{\beta}\right)} - \frac{\exp\left(\beta_{0j-1} + \mathbf{x}_i^T \boldsymbol{\beta}\right)}{1 + \exp\left(\beta_{0j-1} + \mathbf{x}_i^T \boldsymbol{\beta}\right)} \qquad j = 1, 2, \dots, J \tag{5}$$

Jika terdapat tiga kategori respon yaitu $j = 1, 2, 3$ maka model logit dari respon ke-*j* dituliskan pada Persamaan (6) dan Persamaan (7):

$$Logit\left[P(Y_i \le 1 \mid \mathbf{x}_i)\right] = g_1(\mathbf{x}_i) = \ln\left[\frac{P(Y_i \le 1 \mid \mathbf{x}_i)}{P(Y_i > 1 \mid \mathbf{x}_i)}\right] = \beta_{01} + \mathbf{x}_i^T \boldsymbol{\beta} \tag{6}$$

$$Logit\left[P\left(Y_i \le 1 \mid \mathbf{x}_i\right)\right] = g_2\left(\mathbf{x}_i\right) = \ln\left[\frac{P\left(Y_i \le 2 \mid \mathbf{x}_i\right)}{P\left(Y_i > 2 \mid \mathbf{x}_i\right)}\right] = \beta_{02} + x_i^T \boldsymbol{\beta} \tag{7}$$

Setelah mendapatkan model logit, maka peluang kumulatif dari respon ke-*j* akan dijelaskan pada Persamaan (8) dan Persamaan (9):

$$P\left(Y_i \le 1 \mid \mathbf{x}_i\right) = \frac{\exp\left(\beta_{01} + \mathbf{x}_i^T \boldsymbol{\beta}\right)}{1 + \exp\left(\beta_{01} + \mathbf{x}_i^T \boldsymbol{\beta}\right)} \tag{8}$$

$$P\left(Y_i \le 2 \mid \mathbf{x}_i\right) = \frac{\exp\left(\beta_{02} + \mathbf{x}_i^T \boldsymbol{\beta}\right)}{1 + \exp\left(\beta_{02} + \mathbf{x}_i^T \boldsymbol{\beta}\right)} \tag{9}$$

Berdasarkan peluang kumulatif pada Persamaan (8) dan (9), sehingga didapatkan peluang untuk masing-masing kategori respon dituliskan pada Persamaan (10), (11), dan (12)

- Peluang kategori pertama

$$\pi_1\left(\mathbf{x}_i\right) = P\left(Y_i = 1 \mid \mathbf{x}_i\right) = P\left(Y_i \le 1 \mid \mathbf{x}_i\right) = \frac{\exp\left(\beta_{01} + x_i^T \boldsymbol{\beta}\right)}{1 + \exp\left(\beta_{01} + x_i^T \boldsymbol{\beta}\right)} \tag{10}$$

- Peluang kategori kedua

$$\pi_2\left(\mathbf{x}_i\right) = P\left(Y_i = 2 \mid \mathbf{x}_i\right) = P\left(Y_i \le 2 \mid \mathbf{x}_i\right) - P\left(Y_i \le 1 \mid \mathbf{x}_i\right) = \frac{\exp\left(\beta_{02} + \mathbf{x}_i^T \boldsymbol{\beta}\right) - \exp\left(\beta_{01} + \mathbf{x}_i^T \boldsymbol{\beta}\right)}{1 + \exp\left(\beta_{02} + \mathbf{x}_i^T \boldsymbol{\beta}\right)\left(1 + \exp\left(\beta_{01} + \mathbf{x}_i^T \boldsymbol{\beta}\right)\right)} \tag{11}$$

- Peluang kategori ketiga

$$\pi_3\left(\mathbf{x}_i\right) = P\left(Y_i = 3 \mid \mathbf{x}_i\right) = P\left(Y_i \le 3 \mid \mathbf{x}_i\right) - P\left(Y_i \le 2 \mid \mathbf{x}_i\right) = \frac{1}{1 + \exp\left(\beta_{02} + \mathbf{x}_i^T \boldsymbol{\beta}\right)} \tag{12}$$

Nilai peluang untuk setiap kategori variabel respon dapat digunakan memprediksi ketepatan klasifikasi dari variabel respon. Suatu pengamatan akan masuk dalam variabel respon ke-*j* berdasarkan nilai peluang terbesar [9].

2.2 Estimasi Parameter

Estimasi parameter dalam regresi logistik ordinal dapat dilakukan dengan menggunakan metode *Maximum Likelihood Estimation* (MLE). Langkah awal dalam metode MLE adalah membuat fungsi *likelihood* dari regresi logistik ordinal untuk sampel dengan *n* independent observasi $\left(y_i, x_i\right)$ sebagai berikut:

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n}\prod_{j=1}^{J-1}\left(\pi_j\left(\mathbf{x}_i\right)\right)^{y_{ij}} = \prod_{i=1}^{n}(\pi_1(\mathbf{x}_i))^{y_{i1}}(\pi_2(\mathbf{x}_i)^{y_{i2}}...(\pi_{J-1}(\mathbf{x}_i))^{y_{ij}} \tag{13}$$

Ide dasar pada metode MLE adalah menentukan estimator untuk vektor parameter $\boldsymbol{\theta} = \begin{bmatrix}\beta_{01} & \beta_{02} & \cdots & \beta_{0J-1} & \beta_1 & \beta_2 & \cdots & \beta_p\end{bmatrix}^T$ dengan cara memaksimumkan fungsi *likelihood*. Untuk mempermudah perhitungan, maka dilakukan transformasi menggunakan transformasi *natural logarithm* (ln) terhadap fungsi *likelihood* sehingga didapatkan fungsi *log-likelihood* sebagai berikut [7]

$$\ell(\boldsymbol{\theta}) = \left\{y_{i1}\ln(\pi_1(\mathbf{x}_i)) + y_{i2}\ln(\pi_2(\mathbf{x}_i)) + ... + y_{iJ-1}\ln(\pi_{J-1}(\mathbf{x}_i))\right\} + y_{iJ}\ln(\pi_J(\mathbf{x}_i)) \tag{14}$$

Selanjutnya memaksimumkan fungsi *log-likelihood* dengan cara menentukan turunan parsial orde pertama fungsi *log-likelihood* terhadap parameter yang akan diestimasi dan kemudian disamakan dengan nol.

Jika variabel respon terdiri dari 3 kategori $J=3$, maka bentuk fungsi *log-likelihood* dapat dinyatakan secara khusus sesuai dengan jumlah kategori tersebut:

$$\ell(\boldsymbol{\theta})=\sum_{i=1}^{n}\Big(y_{i1}\ln\left[\pi_1(\mathbf{x}_i)\right]+y_{i2}\ln\left[\pi_2(\mathbf{x}_i)\right]+y_{i3}\ln\left[\pi_3(\mathbf{x}_i)\right]\Big)$$

$$\begin{aligned}\ell(\boldsymbol{\theta})=\sum_{i=1}^{n}\Bigg(&y_{i1}\ln\left[\frac{\exp\left(\beta_{01}+\mathbf{x}_i^T\boldsymbol{\beta}\right)}{1+\exp\left(\beta_{01}+\mathbf{x}_i^T\boldsymbol{\beta}\right)}\right]+y_{i2}\ln\left[\frac{\exp\left(\beta_{02}+\mathbf{x}_i^T\boldsymbol{\beta}\right)}{1+\exp\left(\beta_{02}+\mathbf{x}_i^T\boldsymbol{\beta}\right)}\right.\\&\left.-\frac{\exp\left(\beta_{01}+\mathbf{x}_i^T\boldsymbol{\beta}\right)}{1+\exp\left(\beta_{01}+\mathbf{x}_i^T\boldsymbol{\beta}\right)}\right]+\left(1-y_{i1}-y_{i2}\right)\ln\left[1-\frac{\exp\left(\beta_{02}+\mathbf{x}_i^T\boldsymbol{\beta}\right)}{1+\exp\left(\beta_{02}+\mathbf{x}_i^T\boldsymbol{\beta}\right)}\right]\Bigg)\end{aligned}\tag{15}$$

Fungsi *log-likelihood* pada Persamaan (15) disederhanakan menjadi:

$$\begin{aligned}\ell(\boldsymbol{\theta})=\sum_{i=1}^{n}\Big(&y_{i1}\left(\beta_{01}+\mathbf{x}_i^T\boldsymbol{\beta}\right)-\left(y_{i1}+y_{i2}\right)\ln\left[1+\exp\left(\beta_{01}+\mathbf{x}_i^T\boldsymbol{\beta}\right)\right]+y_{i2}\ln\left[\exp\left(\beta_{02}+\mathbf{x}_i^T\boldsymbol{\beta}\right)\right.\\&\left.-\exp\left(\beta_{01}+\mathbf{x}_i^T\boldsymbol{\beta}\right)\right]+\left(y_{i1}-1\right)\ln\left[1+\exp\left(\beta_{02}+\mathbf{x}_i^T\boldsymbol{\beta}\right)\right]\end{aligned}\tag{16}$$

Hasil turunan parsial pertama dari fungsi *log-likelihood* pada Persamaan (16) terhadap parameter yang akan diestimasi adalah sebagai berikut:

$$\begin{aligned}&\frac{\partial\ell(\boldsymbol{\theta})}{\partial\beta_{01}}=\sum_{i=1}^{n}\left[y_{i1}-\left(y_{i1}+y_{i2}\right)\frac{\exp\left(\beta_{01}+\mathbf{x}_i^T\boldsymbol{\beta}\right)}{1+\exp\left(\beta_{01}+\mathbf{x}_i^T\boldsymbol{\beta}\right)}-y_{i2}\frac{\exp(\beta_{01}+\mathbf{x}_i^T\boldsymbol{\beta}}{1+\exp(\beta_{01}+\mathbf{x}_i^T\boldsymbol{\beta})-\exp(\beta_{01}+\mathbf{x}_i^T\boldsymbol{\beta})}\right]=0\\&\frac{\partial\ell(\boldsymbol{\theta})}{\partial\beta_{02}}=\sum_{i=1}^{n}\left[y_{i2}\frac{\exp(\beta_{02}+\mathbf{x}_i^T\boldsymbol{\beta})}{\exp(\beta_{02}+\mathbf{x}_i^T\boldsymbol{\beta})-\exp(\beta_{01}+\mathbf{x}_i^T\boldsymbol{\beta})}+(y_{i1}-1)\frac{\exp(\beta_{02}+\mathbf{x}_i^T\boldsymbol{\beta})}{1+\exp(\beta_{02}+\mathbf{x}_i^T\boldsymbol{\beta})}\right]=0\\&\frac{\partial\ell(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}}=\sum_{i=1}^{n}\left[y_{i1}\mathbf{x}_i^T-(y_{i1}+y_{i2})\mathbf{x}_i^T\frac{\exp(\beta_{01}+\mathbf{x}_i^T\boldsymbol{\beta})}{1+\exp(\beta_{01}+\mathbf{x}_i^T\boldsymbol{\beta})}+y_{i2}\mathbf{x}_i^T+(y_{i1}-1)x_i^T\frac{\exp(\beta_{02}+x_i^T\beta)}{1+\exp(\beta_{02}+x_i^T\beta)}\right]=0\end{aligned}\tag{17}$$

Turunan parsial orde pertama yang dihasilkan membentuk fungsi yang tidak dapat diselesaikan secara eksplisit, sehingga dibutuhkan pendekatan numerik untuk memperoleh estimasi parameter. Dalam penelitian ini, pendekatan yang digunakan adalah dengan metode Newton-Raphson, yang memerlukan turunan parsial orde kedua dari fungsi *log-likelihood* terhadap parameter yang akan diestimasi. Hasil turunan parsial orde kedua yang didapatkan adalah sebagai berikut:

$$\begin{aligned}&\frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\beta_{01}^2}=\sum_{i=1}^{n}\left\{-(y_{i1}+y_{i2})\frac{\exp(\beta_{01}+\mathbf{x}_i^T\boldsymbol{\beta})}{[1+\exp(\beta_{01}+\mathbf{x}_i^T\boldsymbol{\beta})]^2}-y_{i2}\frac{\exp(\beta_{01}+\mathbf{x}_i^T\boldsymbol{\beta})\exp(\beta_{02}+\mathbf{x}_i^T\boldsymbol{\beta})}{[\exp(\beta_{02}+\mathbf{x}_i^T\boldsymbol{\beta})1-\exp(\beta_{01}+\mathbf{x}_i^T\boldsymbol{\beta})]^2}\right\}\\&\frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\beta_{01}\beta_{02}}=\sum_{i=1}^{n}\left\{y_{i2}\frac{\exp(\beta_{01}+\mathbf{x}_i^T\boldsymbol{\beta})\exp(\beta_{02}+\mathbf{x}_i^T\boldsymbol{\beta})}{[\exp(\beta_{02}+\mathbf{x}_i^T\boldsymbol{\beta})-\exp(\beta_{01}+\mathbf{x}_i^T\boldsymbol{\beta})]^2}\right\}\end{aligned}\tag{18}$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_{01} \partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \left\{ -(y_{i1} + y_{i2}) \mathbf{x}_i^T \frac{\exp(\beta_{01} + \mathbf{x}_i^T \boldsymbol{\beta})}{[1 + \exp(\beta_{01} + \mathbf{x}_i^T \boldsymbol{\beta})]^2} \right\}$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_{02}^2} = \sum_{i=1}^{n} \left\{ -y_{i2} \frac{\exp(\beta_{01} + \mathbf{x}_i^T \boldsymbol{\beta}) \exp(\beta_{02} + \mathbf{x}_i^T \boldsymbol{\beta})}{[\exp(\beta_{02} + \mathbf{x}_i^T \boldsymbol{\beta}) - \exp(\beta_{01} + \mathbf{x}_i^T \boldsymbol{\beta})]^2} + (y_{i1} - 1 \frac{\exp(\beta_{02} + \mathbf{x}_i^T \boldsymbol{\beta})}{[1 + \exp(\beta_{02} + \mathbf{x}_i^T \boldsymbol{\beta})]^2} \right\}$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_{02} \partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \left\{ y_{i2} \mathbf{x}_i^T \frac{\exp(\beta_{02} + \mathbf{x}_i^T \boldsymbol{\beta})}{[1 + \exp(\beta_{02} + \mathbf{x}_i^T \boldsymbol{\beta})]^2} + (y_{i1} - 1) \mathbf{x}_i^T \frac{\exp(\beta_{02} + \mathbf{x}_i^T \boldsymbol{\beta})}{[1 + \exp(\beta_{02} + \mathbf{x}_i^T \boldsymbol{\beta})]^2} \right\}$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} = \sum_{i=1}^{n} \left\{ -(y_{i1} + y_{i2}) \mathbf{x}_i^T \mathbf{x}_i \frac{\exp(\beta_{01} + \mathbf{x}_i^T \boldsymbol{\beta})}{[1 + \exp(\beta_{01} + \mathbf{x}_i^T \boldsymbol{\beta})]^2} + (y_{i1} - 1) \mathbf{x}_i^T \mathbf{x}_i \frac{\exp(\beta_{02} + \mathbf{x}_i^T \boldsymbol{\beta})}{[1 + \exp(\beta_{02} + \mathbf{x}_i^T \boldsymbol{\beta})]^2} \right\}$$

Formula yang digunakan untuk mendapatkan *estimator* ML parameter model regresi logistik ordinal dengan metode Newton-Raphson adalah [10]:

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} - \left[\mathbf{H}(\boldsymbol{\theta}^{(t)}) \right]^{-1} \mathbf{q}(\boldsymbol{\theta}^{(t)}) \tag{19}$$

$\mathbf{H}(\boldsymbol{\theta})$ merupakan matriks Hessian nonsingular dengan elemen-elemennya berupa turunan parsial kedua fungsi *log-likelihood* terhadap parameter yang akan diestimasi, sedangkan $\mathbf{q}(\boldsymbol{\theta})$ adalah vector yang terdiri dari turunan parsial pertama dari fungsi *log-likelihood* terhadap parameter tersebut, dan t menunjukkan banyaknya iterasi yang dilakukan $(t = 0, 1, 2, ...)$. Nilai vektor $\mathbf{q}(\boldsymbol{\theta})$ dan matriks $\mathbf{H}(\boldsymbol{\theta})$ dihitung berdasarkan formula yang dituliskan pada Persamaan (20) sebagai berikut:

$$\mathbf{H}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_{01}^2} & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_{01} \partial \beta_{02}} & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_{01} \partial \boldsymbol{\beta}} \\ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_{01} \partial \beta_{02}} & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_{02}^2} & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_{02} \partial \boldsymbol{\beta}} \\ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_{01} \partial \beta_{02}} & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_{02} \partial \boldsymbol{\beta}} & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \end{bmatrix} \tag{20}$$

Proses iterasi Newton-Raphson akan berhenti apabila terpenuhi kondisi konvergen, yaitu $\left\| \boldsymbol{\theta}^{(t+1)} - \boldsymbol{\theta}^{(t)} \right\| \leq \varepsilon,$ dimana ε adalah bilangan yang sangat kecil. Hasil estimasi yang diperoleh adalah $\boldsymbol{\theta}^{(t+1)}$ pada iterasi terakhir [10].

2.3 Pengujian Multikolinieritas

Menurut [11] multikolinearitas merupakan hubungan linier yang sempurna pada variabel prediktor untuk model regresi. Uji *Pearson Chi-square* adalah metode statistik yang digunakan untuk menguji hubungan antara dua variabel kategori [12]. Syarat uji *Pearson Chi-square* yaitu, data yang digunakan harus dalam bentuk kategori (nominal atau ordinal). Apabila bentuk tabel kontingensi 2 × 2, maka tidak boleh ada 1 sel yang memiliki frekuensi harapan kurang dari 5, dan jika bentuk tabel lebih dari 2 × 2, misal 2 × 3, maka jumlah sel dengan frekuensi harapan yang kurang dari lima tidak boleh lebih dari 20% [13]. Pasangan variabel prediktor yang menunjukkan terdapat hubungan signifikan maka dilanjutkan dengan menghitung koefisien kontingensi untuk mengukur tingkat kekuatan hubungan yang terjadi. Uji Eta yaitu digunakan untuk melihat korelasi antar variabel prediktor lainnya. Eta merupakan metode yang digunakan untuk menentukan hubungan antara variabel kategori dengan variabel rasio atau interval [14].

2.4 Uji Simultan

Uji simultan digunakan untuk mengetahui apakah taksiran parameter variabel prediktor yang diperoleh berpengaruh secara signifikan terhadap variabel respon. Hipotesis yang digunakan sebagai berikut:

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

(Secara simultan variabel prediktor tidak berpengaruh signifikan terhadap variabel respon).

H_1 :Minimal ada satu $\beta_k \neq 0$ dimana $k = 1, 2, ..., p$

(Terdapat minimal satu variabel prediktor yang berpengaruh signifikan terhadap variabel respon).

Statistik uji yang digunakan adalah G^2 yaitu:

$$G^2 = -2\ln\left[\frac{\left(\frac{n_1}{n}\right)^{n_1}\left(\frac{n_2}{n}\right)^{n_2}\left(\frac{n_3}{n}\right)^{n_3}}{\prod_{i=1}^{n}\pi_1(x_i)^{y_{i1}}\pi_2(x_i)^{y_{i2}}\left(1-\pi_1(x_i)-\pi_2(x_i)^{(1-y_{i1}-y_{i2})}\right)}\right] \tag{25}$$

Daerah penolakan H_0 untuk tingkat signifikansi α adalah H_0 ditolak jika nilai $G^2 > \chi^2_{(\alpha,p)}$ dimana p menyatakan jumlah parameter dalam model. Nilai $\chi^2_{(\alpha,p)}$ dapat diperoleh dari tabel distribusi *Chi-Square* [15].

2.5 Uji Parsial

Uji parsial bertujuan untuk mengetahui apakah masing-masing variabel predictor secara individu memiliki pengaruh yang signifikan terhadap variabel respon. Hipotesis yang digunakan dalam uji parsial sebagai berikut:

$H_0 : \beta_k = 0$

$H_1 : \beta_k \neq 0, k = 1, 2, ..., p$

Statistik uji yang digunakan adalah statistik Wald yang dinyatakan pada Persamaan (26)

$$W = \frac{\widehat{\beta}_k}{SE\left(\widehat{\beta}_k\right)} \tag{26}$$

Daerah penolakan H_0 ditolak jika nilai $|W| > Z_{\alpha/2}$. Nilai $Z_{\alpha/2}$ diperoleh tabel distribusi normal standar [15].

2.6 Pemilihan Model Terbaik dengan Metode *Stepwise*

Metode *stepwise* merupakan kombinasi dari teknik seleksi peubah maju (*forward*) dan mundur (*backward*). Metode *stepwise* dilakukan dengan memasukkan semua variabel prediktor dan secara bertahap mengeliminasi variabel prediktor yang tidak signifikan. Proses ini dilakukan terus menerus hingga tidak ada lagi variabel yang tidak signifikan dan memiliki nilai *Akaike Information Criterion* (AIC) yang paling kecil [16].

Akaike Information Criterion (AIC) adalah salah satu pendekatan untuk memilih model terbaik. Model terbaik tersebut merupakan model yang mempunyai hasil nilai AIC terkecil. Perhitungan metode ini didasarkan pada pendekatan *Maximum Likelihood Estimation* (MLE) dengan persamaan sebagai berikut:

$$AIC = -2\ell\left(\hat{\theta}\right) + 2p \tag{27}$$

dimana $\ell(\hat{\theta})$ adalah fungsi *log-likelihood* dan p adalah jumlah parameter yang diestimasi dalam model regresi [17].

2.7 Interpretasi Model Regresi Logistik

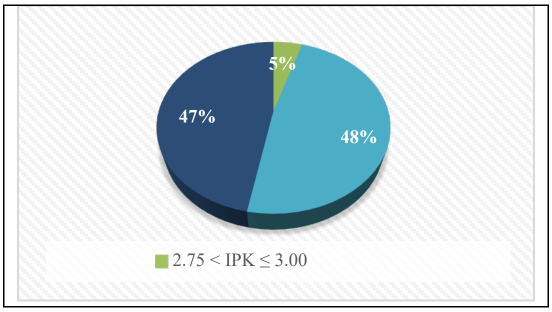
Interpretasi model merupakan suatu bentuk untuk mendefinisikan unit pada variabel respon yang disebabkan oleh variabel prediktor. Untuk memudahkan dalam menginterpretasikan model regresi logistik digunakan nilai (OR) dinotasikan dengan ψ. (OR) digunakan untuk mengukur seberapa kuat hubungan antara variabel respon dan variabel prediktor dalam regresi logistik [18]

$$\psi = \left[\frac{\left[\dfrac{P(Y \le j \mid x=2)}{P(Y > j \mid x=2)}\right]}{\left[\dfrac{P(Y \le j \mid x=1)}{P(Y > j \mid x=1)}\right]}\right] == \exp\left[(\beta_{0j} + 2\beta_1) - (\beta_{0j} + \beta_1)\right] = \exp\left(\beta_1\right) \tag{28}$$

3. Hasil dan Diskusi

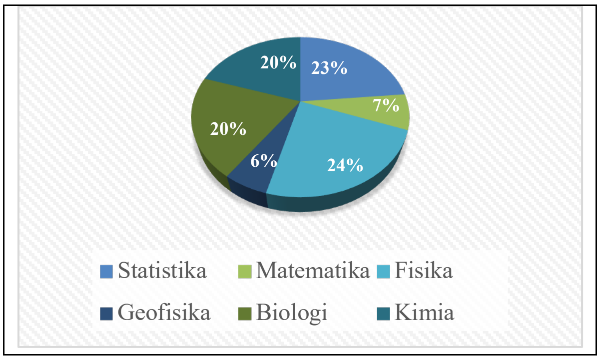
3.1 Analisis Statistik Deskriptif Variabel Penelitian

Deskripsi data penelitian mengenai variable respon, yaitu indeks prestasi kumulatif lulusan program sarjana FMIPA Unmul dapat dilihat pada Gambar 1:



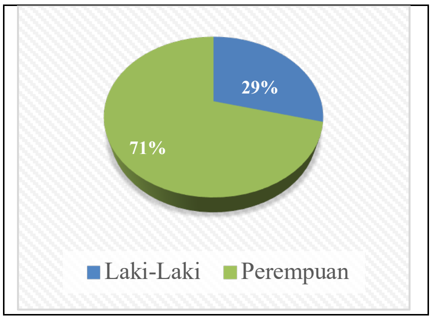
Gambar 1. Proporsi Indeks Prestasi Kumulatif Lulusan Program Sarjana FMIPA Unmul

Berdasarkan Gambar 1, dapat dilihat bahwa dari 251 orang lulusan program sarjana FMIPA Unmul tahun 2023, lulusan terbanyak adalah lulusan yang mendapatkan 3,00 < IPK≤ 3,50 sebanyak 48%. Selanjutnya disajikan pada variabel-variabel prediktor.



Gambar 2. Proporsi Program Studi Lulusan Program Sarjana FMIPA Unmul

Berdasarkan Gambar 2, dapat dilihat bahwa lulusan terbanyak berasal dari Program Studi Fisika sebesar 24%.



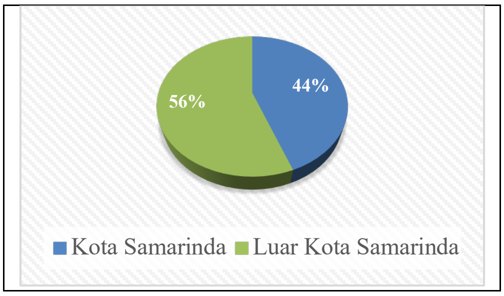
Gambar 3. Proporsi Jenis Kelamin Lulusan Program Sarjana FMIPA Unmul

Berdasarkan Gambar 3, dapat dilihat bahwa lulusan terbanyak jenis kelamin perempuan sebesar 71%.

Tabel 1. Statistika Deskriptif Lulusan FMIPA Unmul Berdasarkan Usia

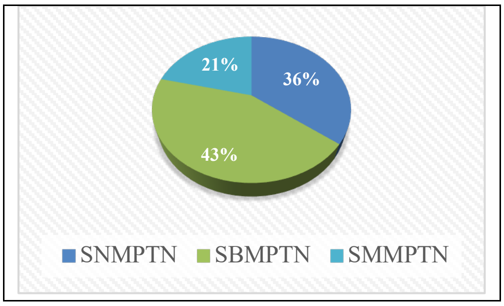
Minimum	Maksimum	Rata-Rata	Simpangan Baku
21	27	22,896	1,199

Berdasarkan Tabel 1, dapat dilihat bahwa rata-rata lulusan program sarjana FMIPA Unmul adalah 22,896, dengan lulusan termuda berusia 21 tahun dan lulusan tertua berusia 27 tahun.



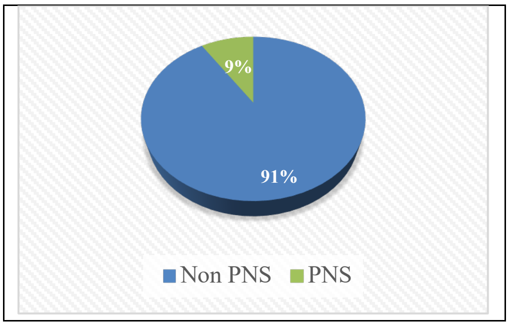
Gambar 4. Proporsi Asal Daerah Lulusan Program Sarjana FMIPA Unmul

Berdasarkan Gambar 4, dapat dilihat bahwa lulusan terbanyak berasal dari Luar Kota Samarinda sebesar 56%.



Gambar 5. Proporsi Jalur Penerimaan Lulusan Program Sarjana FMIPA Unmul

Berdasarkan Gambar 5, dapat dilihat bahwa lulusan terbanyak dari jalur penerimaan SBMPTN sebesar 43%.



Gambar 6. Proporsi Pekerjaan Orang Tua Lulusan Program Sarjana FMIPA Unmul

Berdasarkan Gambar 6, dapat dilihat bahwa lulusan terbanyak dengan pekerjaan orang tua Non PNS sebesar 91%.

3.2 Pengujian Multikolinieritas

Pengujian Multikolinieritas merupakan salah satu pelanggaran terhadap asumsi pada regresi logistik ordinal, yaitu adanya hubungan yang kuat antar variabel prediktor. Pengujian multikolinieritas pada penelitian ini menggunakan uji Pearson untuk variabel X_1, X_2, X_4, X_5, dan X_6. Pasangan variabel prediktor yang menunjukkan terdapat hubungan signifikan maka dilanjutkan dengan menghitung koefisien kontingensi untuk mengukur tingkat kekuatan hubungan yang terjadi. Uji Eta yaitu X_3 digunakan untuk melihat korelasi antar variabel prediktor lainnya.

Tabel 2. Hasil Pengujian *Pearson Chi-Square*

Variabel Prediktor	Statistik *Pearson Chi-square*	*p-value*	Keputusan	Kesimpulan
X_1 dan X_2	6,426	0,266	H_0 gagal ditolak	Tidak terdapat hubungan
X_1 dan X_4	12,037	0,034	H_0 ditolak	Terdapat hubungan
X_1 dan X_5	25,326	0,004	H_0 ditolak	Terdapat hubungan

Tabel 3. Hasil Pengujian *Pearson Chi-Square* (Lanjutan)

Variabel Prediktor	Statistik *Pearson Chi-square*	*p-value*	Keputusan	Kesimpulan
X_1 dan X_6	3,620	0,605	H_0 gagal ditolak	Tidak terdapat hubungan
X_2 dan X_4	0,578	0,447	H_0 gagal ditolak	Tidak terdapat hubungan
X_2 dan X_5	8,366	0,015	H_0 ditolak	Terdapat hubungan
X_2 dan X_6	9,242	0,002	H_0 ditolak	Terdapat hubungan
X_4 dan X_5	0,756	0,684	H_0 gagal ditolak	Tidak terdapat hubungan
X_4 dan X_6	4,524	0,033	H_0 gagal ditolak	Terdapat hubungan
X_5 dan X_6	0,297	0,861	H_0 gagal ditolak	Tidak terdapat hubungan

Kemudian untuk mengetahui keeratan hubungan antara variabel prediktor dapat menghitung nilai koefisien kontingensi pada variabel yang terdapat hubungan yaitu program studi (X_1) dengan asal daerah (X_4), program studi (X_1) dengan jalur penerimaan (X_5); jenis kelamin (X_2) dengan jalur penerimaan (X_5), jenis kelamin (X_2) dengan pekerjaan orang tua (X_6), dan asal daerah (X_4) dengan pekerjaan orang tua (X_6). Perhitungan menggunakan *software R*, diperoleh hasil yang disajikan pada Tabel 4.

Tabel 4. Nilai Koefisien Kontingensi

Variabel Prediktor	$\boldsymbol{C}$	$\boldsymbol{C_{maks}}$	$\boldsymbol{C/C_{maks}}$	Kesimpulan
X_1 dan X_4	0,213	0,707	0,301	Lemah

X_1 dan X_5	0,302	0,816	0,370	Lemah
X_2 dan X_5	0,179	0,707	0,253	Lemah
X_2 dan X_6	0,188	0,707	0,265	Lemah
X_4 dan X_6	0,131	0,707	0,185	Lemah

Selanjutnya menghitung untuk mengetahui hubungan korelasi pada variabel usia dengan variabel prediktor lainnya menggunakan uji Eta.

Tabel.5 Hasil uji Eta

Variabel Prediktor	Statistik Uji	Keputusan	Kesimpulan
X_3 dan X_1	5,688	H_0 ditolak	Terdapat hubungan signifikan
X_3 dan X_2	8,764	H_0 ditolak	Terdapat hubungan signifikan

Tabel 6. Hasil uji Eta (Lanjutan)

Variabel Prediktor	Statistik Uji	Keputusan	Kesimpulan
X_3 dan X_4	14,214	H_0 ditolak	Terdapat hubungan signifikan
X_3 dan X_5	4,764	H_0 ditolak	Terdapat hubungan signifikan
X_3 dan X_6	6,647	H_0 ditolak	Terdapat hubungan signifikan

Berdasarkan Tabel 5 dan Tabel 6 dapat diketahui bahwa variabel-variabel prediktor yang layak digunakan untuk model regresi logistik ordinal adalah X_1, X_2, X_3, X_4, X_5, dan X_6.

3.3 Pemodelan Indeks Prestasi Kumulatif dengan Regresi Logistik Ordinal

Sebelum melakukan penaksiran parameter maka perlu diketahui model awal regresi logistik ordinal dengan enam variabel prediktor yang terbentuk sebagai berikut:

1. $$Logit\left[P(Y_i \le 1 \mid \mathbf{x}_i)\right] = g_1(\mathbf{x}_i) = \beta_{01} + \beta_1 x_{i1(2)} + \beta_2 x_{i1(3)} + \beta_3 x_{i1(4)} + \beta_4 x_{i1(5)} + \beta_5 x_{i1(6)}$$
$$+\beta_6 x_{i2(2)} + \beta_7 x_{i3} + \beta_8 x_{i4(2)} + \beta_9 x_{i5(2)} + \beta_{10} x_{i5(3)} + \beta_{11} x_{i6(2)}$$
$$;i = 1,2,3,...,251$$
2. $$Logit\left[P(Y_i \le 2 \mid \mathbf{x}_i)\right] = g_2(\mathbf{x}_i) = \beta_{02} + \beta_1 x_{i1(2)} + \beta_2 x_{i1(3)} + \beta_3 x_{i1(4)} + \beta_4 x_{i1(5)} + \beta_5 x_{i1(6)}$$
$$+\beta_6 x_{i2(2)} + \beta_7 x_{i3} + \beta_8 x_{i4(2)} + \beta_9 x_{i5(2)} + \beta_{10} x_{i5(3)} + \beta_{11} x_{i6(2)}$$
$$;i = 1,2,3,...,251$$

Berdasarkan model di atas maka dalam penaksiran parameter model regresi logistik ordinal variabel respon kategori $Y = 1$ yaitu 2,75≤IPK≤3,00 digunakan sebagai pembanding. Penaksiran parameter menggunakan metode MLE dan iterasi Newton-Raphson

Tabel 7. Hasil Penaksiran Parameter Model Regresi Logistik Ordinal

Variabel	**Parameter**	**Taksiran Parameter**	***Standard Error***
Konstanta 1	β_{0_1}	-25,928	3,702
Konstanta 2	β_{0_2}	-21,620	3,485
$X_{1(2)}$	β_1	-1,393	0,594
$X_{1(3)}$	β_2	-0,986	0,417
$X_{1(4)}$	β_3	-1,040	0,674
$X_{1(5)}$	β_4	0,925	0,443
$X_{1(6)}$	β_5	0,287	0,452
$X_{2(2)}$	β_6	0,296	0,323

Tabel 8. Hasil Penaksiran Parameter Model Regresi Logistik Ordinal (Lanjutan)

Variabel	**Parameter**	**Taksiran Parameter**	***Standard Error***
X_3	β_7	-0,927	0,150
$X_{4(2)}$	β_8	-0,279	0,301
$X_{5(2)}$	β_9	-0,504	0,334
$X_{5(3)}$	β_{10}	-1,670	0,411
$X_{6(2)}$	β_{11}	0,114	0,509

Berdasarkan Tabel 7 dan 8, dapat diperoleh model regresi logistik ordinal untuk memodelkan data indeks prestasi kumulatif lulusan program sarjana FMIPA Unmul sebagai berikut:

1. $$\begin{aligned} Logit\left[P(Y_i \le 1 \mid \mathbf{x}_i)\right] = g_1(\mathbf{x}_i) = &-25,928 + 1,393x_{i1(2)} + 0,986x_{i1(3)} + 1,040x_{i1(4)} - 0,925x_{i1(5)} \\ &-0,287x_{i1(6)} - 0,296x_{i2(2)} + 0,927x_{i3} + 0,279x_{i4(2)} \\ &+0,504x_{i5(2)} + 1,670x_{i5(3)} - 0,114x_{i6(2)} \\ &; i = 1,2,3,\ldots,251 \end{aligned}$$

2. $$\begin{aligned} Logit\left[P(Y_i \le 2 \mid \mathbf{x}_i)\right] = g_2(\mathbf{x}_i) = &-21,620 + 1,393x_{i1(2)} + 0,986x_{i1(3)} + 1,040x_{i1(4)} - 0,925x_{i1(5)} \\ &-0,287x_{i1(6)} - 0,296x_{i2(2)} + 0,927x_{i3} + 0,279x_{i4(2)} \\ &+0,504x_{i5(2)} + 1,670x_{i5(3)} - 0,114x_{i6(2)} \end{aligned}$$

$; i = 1, 2, 3, ..., 251$

Berdasarkan model logit di atas maka didapatkan model peluang indeks prestasi kumulatif lulusan program sarjana FMIPA Unmul untuk masing-masing kategori sebagai berikut:

- 2,75<IPK≤3,00

$$\pi_1(\mathbf{x}_i) = \frac{\exp(g_1(\mathbf{x}_i))}{1+\exp(g_1(\mathbf{x}_i))}$$

- 3,00<IPK≤3,50

$$\pi_2(\mathbf{x}_i) = \frac{\exp(g_2(\mathbf{x}_i)) - \exp(g_1(\mathbf{x}_i))}{\left(1+\exp(g_2(\mathbf{x}_i))\right)\left(1+\exp(g_1(\mathbf{x}_i))\right)}$$

- IPK>3,50

$$\pi_3(\mathbf{x}_i) = 1 - \pi_1(\mathbf{x}_i) - \pi_2(\mathbf{x}_i)$$

Setelah estimasi parameter model regresi logistik ordinal didapatkan, maka selanjutnya dilakukan pengujian signifikansi parameter yaitu uji simultan dan uji parsial.

Berdasarkan perhitungan *software* R, diperoleh nilai *Wilk's lambda* (G) sebesar 98,457 dan nilai $\chi^2 = 15{,}507$. Karena nilai G lebih besar, maka H_0ditolak, sehingga dapat disimpulkan bahwa terdapat minimal satu parameter yang berpengaruh secara signifikan terhadap variabel indeks prestasi kumulatif.

Pengujian selanjutnya adalah uji parsial. Pengujian parameter secara parsial dilakukan untuk mengetahui apakah variabel prediktor berpengaruh secara individu terhadap variabel respon. Hasil uji parsial variabel prediktor usia (X_3) berpengaruh terhadap indeks prestasi kumulatif lulusan program sarjana FMIPA Unmul. Hal ini ditunjukkan oleh nilai statistik uji $|W| > 1{,}96$. Dengan demikian, variabel (X_3) berpengaruh terhadap model regresi logistik ordinal. Adapun pada variabel prediktor program studi (X_1) dan jalur penerimaan (X_5), terdapat beberapa kategori yang memiliki nilai statistik uji $|W| < 1{,}96$ seperti $X_{1(2)}$, $X_{1(3)}$, $X_{1(5)}$, dan $X_{5(3)}$, yang secara individu tidak signifikan. Namun demikian, variabel program studi (X_1) dan jalur penerimaan (X_5) tetap dinyatakan berpengaruh terhadap model secara keseluruhan karena merupakan bagian dari suatu satu kesatuan dengan variabel-variabel prediktor lainnya yang berpengaruh secara signifikan.

3.4 Pemilihan Model Regresi Logistik Ordinal Terbaik

Pemilihan model regresi logistik ordinal terbaik dilakukan untuk memastikan bahwa variabel prediktor yang disertakan dalam model memberikan kontribusi optimal terhadap variabel respon. Proses pemilihan model terbaik dilakukan menggunakan metode *stepwise* berdasarkan perhitungan dengan *software R*.

Tabel 9. Pemilihan Model Terbaik

Tahap	Model	Variabel Dieliminasi	AIC
1	$X_1, X_2, X_3, X_4, X_5, X_6$	X_6	351,63
2	X_1, X_2, X_3, X_4, X_5	X_2	350,41
3	X_1, X_4, X_3, X_5	X_4	349,22
4	X_1, X_3, X_5	-	349,22

Tahap pertama pemilihan model regresi logistik ordinal terbaik menggunakan metode *stepwise* dimulai dengan memasukkan semua variabel prediktor, yaitu $X_1, X_2, X_3, X_4, X_5, X_6$. Pada tahap kedua metode *stepwise*, variabel prediktor X_6 dieliminasi karena memiliki nilai uji Wald yang tidak signifikan dan menurunkan nilai AIC sebesar 351,63 dapat dilihat pada Lampiran 8, sehingga dilanjutkan dengan model yang mencakup lima variabel prediktor X_1, X_2, X_3, X_4, X_5. Selanjutnya pada tahap ketiga metode *stepwise*, variabel prediktor X_2 dieliminasi karena memiliki nilai uji Wald yang tidak signifikan dan menurunkan nilai AIC sebesar 350,41, sehingga dilanjutkan dengan model yang mencakup 4 variabel prediktor. Pada tahap keempat metode *stepwise*, variabel prediktor X_4 dieliminasi karena memiliki nilai uji wald yang tidak signifikan dan menurunkan nilai AIC sebesar 349,22, sehingga dilanjutkan dengan model yang mencakup 3 variabel prediktor, yaitu X_1, X_3, dan X_5. Tahapan metode *stepwise* berhenti pada model dengan tiga variabel prediktor X_1, X_3, dan X_5, karena variabel prediktor signifikan dan memiliki nilai AIC paling kecil sebesar 349,22.

Variabel X_1, X_3, X_5 tetap dipertahankan dalam model, walaupun beberapa kategorinya seperti $X_{1(4)}$, $X_{1(6)}$, $X_{5(2)}$ tidak signifikan secara individu. Kategori yang tidak signifikan dalam variabel X_1 dan X_5 tidak menghilangkan peran kedua variabel tersebut dalam model. Selain itu, karena variabel prediktor ini dalam bentuk dummy, maka ketika ada salah satu kategori yang tidak signifikan, tetap dipertahankan karena mewakili satu kesatuan variabel prediktor yang secara keseluruhan signifikan. Ketika variabel X_1 dan X_5 dieliminasi, model hanya terbentuk X_3 saja sebesar 374,13 yang mengakibatkan peningkatan nilai AIC. Peningkatan AIC ini menunjukkan bahwa model dengan hanya variabel X_3 memiliki kualitas yang lebih rendah dibandingkan model yang tetap mempertahankan variabel X_1 dan X_5. Oleh karena itu, meskipun ada kategori yang tidak signifikan, X_1 dan X_5 tetap dipertahankan karena secara keseluruhan memberikan kontribusi terhadap model serta menjaga AIC tetap optimal.

Setelah pemilihan model regresi logistik ordinal terbaik diperoleh, ditentukan kembali nilai taksiran parameter. Hasil penaksiran parameter diperoleh berdasarkan perhitungan melalui *software R*

Tabel 10. Hasil Penaksiran Parameter Model Terbaik

Variabel	Parameter	Taksiran Parameter	*Standard Error*
Konstanta 1	β_{01}	-26,745	3,653
Konstanta 2	β_{02}	-22,460	3,427
$X_{1(2)}$	β_1	-1,400	0,585
$X_{1(3)}$	β_2	-0,946	0,414
$X_{1(4)}$	β_3	-1,111	0,669
$X_{1(5)}$	β_4	0,934	0,441
$X_{1(6)}$	β_5	0,322	0,450
X_3	β_6	-0,961	0,148
$X_{5(2)}$	β_7	-0,561	0,329
$X_{5(3)}$	β_8	-1,641	0,410

Berdasarkan Tabel 10, dapat diperoleh model regresi logistik ordinal terbaik untuk memodelkan indeks prestasi kumulatif lulusan program sarjana FMIPA Unmul sebagai berikut:

1. $$\begin{aligned}Logit\left[P(Y_i \le 1 \mid \mathbf{x}_i)\right] = g_1(\mathbf{x}_i) &= -26,745 + 1,400x_{i1(2)} + 0,946x_{i1(3)} + 1,111x_{i1(4)} \\ &\quad -0,934x_{i1(5)} - 0,322x_{i1(6)} + 0,961x_{i3} + 0,561x_{i5(2)} + 1,641x_{i5(3)} \\ &\quad ; i = 1,2,3,\ldots,251\end{aligned}$$

2. $$\begin{aligned}Logit\left[P(Y_i \le 2 \mid \mathbf{x}_i)\right] = g_2(\mathbf{x}_i) &= -22,460 + 1,400x_{i1(2)} + 0,946x_{i1(3)} + 1,111x_{i1(4)} \\ &\quad -0,934x_{i1(5)} - 0,322x_{i1(6)} + 0,961x_{i3} + 0,561x_{i5(2)} + 1,641x_{i5(3)} \\ &\quad ; i = 1,2,3,\ldots,251\end{aligned}$$

Berdasarkan model logit maka didapatkan model peluang indeks prestasi kumulatif lulusan program sarjana FMIPA Unmul untuk masing-masing kategori sebagai berikut:

- 2,75<IPK≤3,00

$$\pi_1(\mathbf{x}_1) = \frac{\exp\left(g_1(\mathbf{x}_1)\right)}{1+\exp\left(g_1(\mathbf{x}_1)\right)} = \frac{\exp(-3,135)}{1+\exp(-3,135)} = \frac{0,043}{1,043} = 0,041$$

$$\pi_1(\mathbf{x}_2) = \frac{\exp\left(g_1(\mathbf{x}_2)\right)}{1+\exp\left(g_1(\mathbf{x}_2)\right)} = \frac{\exp(-2,174)}{1+\exp(-2,174)} = \frac{0,113}{1,113} = 0,102$$

$$\vdots$$

$$\pi_1(\mathbf{x}_{251}) = \frac{\exp\left(g_1(\mathbf{x}_{251})\right)}{1+\exp\left(g_1(\mathbf{x}_{251})\right)} = \frac{\exp(-4,096)}{1+\exp(-4,096)} = \frac{0,016}{1,016} = 0,016$$

- 3,00<IPK≤3,50

$$\pi_2(\mathbf{x}_1) = \frac{\exp\left(g_2(\mathbf{x}_1)\right) - \exp\left(g_1(\mathbf{x}_1)\right)}{\left(1+\exp\left(g_2(\mathbf{x}_1)\right)\right)\left(1+\exp\left(g_1(\mathbf{x}_1)\right)\right)}$$

$$= \frac{\exp(1{,}150) - \exp(-3{,}135)}{(1+\exp(1{,}150))(1+\exp(-3{,}135))} = 0{,}717$$

$$\pi_2(\mathbf{x}_2) = \frac{\exp(g_2(\mathbf{x}_2)) - \exp(g_1(\mathbf{x}_2))}{(1+\exp(g_2(\mathbf{x}_2)))(1+\exp(g_1(\mathbf{x}_2)))}$$

$$= \frac{\exp(2{,}111) - \exp(-2{,}174)}{(1+\exp(2{,}111))(1+\exp(-2{,}174))} = 0{,}789$$

$$\vdots$$

$$\pi_2(\mathbf{x}_{251}) = \frac{\exp(g_2(\mathbf{x}_{251})) - \exp(g_1(\mathbf{x}_{251}))}{(1+\exp(g_2(\mathbf{x}_{251})))(1+\exp(g_1(\mathbf{x}_{251})))}$$

$$= \frac{\exp(0{,}189) - \exp(-4{,}096)}{(1+\exp(0{,}189))(1+\exp(-4{,}096))} = 0{,}530$$

- IPK>3,50

$$\pi_3(\mathbf{x}_1) = 1 - \pi_1(\mathbf{x}_1) - \pi_2(\mathbf{x}_1)$$

$$= 1 - 0{,}041 - 0{,}717 = 0{,}240$$

$$\pi_3(\mathbf{x}_2) = 1 - \pi_1(\mathbf{x}_2) - \pi_2(\mathbf{x}_2)$$

$$= 1 - 0{,}102 - 0{,}789 = 0{,}108$$

$$\vdots$$

$$\pi_3(\mathbf{x}_{251}) = 1 - \pi_1(\mathbf{x}_{251}) - \pi_2(\mathbf{x}_{251})$$

$$= 1 - 0{,}016 - 0{,}530 = 0{,}452$$

Setelah estimasi parameter model regresi logistik ordinal terbaik didapatkan, maka selanjutnya dilakukan pengujian signifikansi parameter model terbaik. Hipotesis uji simultan adalah :

$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = 0$

(Secara simultan variabel program studi, usia, dan jalur penerimaan tidak berpengaruh signifikan terhadap variabel indeks prestasi kumulatif).

H_1 : Minimal ada satu parameter model regresi logistik tidak bernilai nol

(Terdapat minimal satu variabel prediktor yang berpengaruh signifikan terhadap variabel indeks prestasi kumulatif).

Berdasarkan perhitungan *software* R, diperoleh nilai *Wilk's lambda* (G) sebesar $100{,}100$ dan nilai $\chi^2 = 19{,}675$. Karena nilai G lebih besar, maka H_0ditolak, sehingga dapat disimpulkan bahwa terdapat minimal satu parameter yang berpengaruh secara signifikan terhadap variabel indeks prestasi kumulatif.

Pengujian selanjutnya adalah uji parsial. Hipotesis yang digunakan adalah :

$H_0 : \beta_k = 0, k = 1,2,...,8$

(Secara parsial program studi, usia, dan jalur penerimaan tidak berpengaruh signifikan terhadap indeks prestasi kumulatif lulusan program sarjana FMIPA Unmul)

$H_1 : \beta_k \neq 0, k = 1,2,\ldots,8$

(Secara parsial program studi, usia, dan jalur penerimaan berpengaruh signifikan terhadap indeks prestasi kumulatif lulusan program sarjana FMIPA Unmul)

Tabel 11. Hasil Uji Parsial Model Terbaik

Variabel	Parameter	W	$Z_{0,025}$	Keputusan
Konstanta 1	β_{01}	-7,003	1,96	H_0 ditolak
Konstanta 2	β_{02}	-6,203	1,96	H_0 ditolak
$X_{1(2)}$	β_1	-2,391	1,96	H_0 ditolak
$X_{1(3)}$	β_2	-2,285	1,96	H_0 ditolak
$X_{1(4)}$	β_3	-1,660	1,96	H_0 gagal ditolak
$X_{1(5)}$	β_4	2,117	1,96	H_0 ditolak
$X_{1(6)}$	β_5	0,715	1,96	H_0 gagal ditolak
X_3	β_6	-6,475	1,96	H_0 ditolak
$X_{5(2)}$	β_7	-1,703	1,96	H_0 gagal ditolak
$X_{5(3)}$	β_8	-4,004	1,96	H_0 ditolak

Berdasarkan hasil uji parsial model terbaik dapat dilihat bahwa variabel prediktor usia (X_3) berpengaruh terhadap indeks prestasi kumulatif lulusan program sarjana FMIPA Unmul. Hal ini ditunjukkan oleh nilai statistik uji $|W| > 1{,}96$. Dengan demikian, variabel (X_3) berpengaruh terhadap model regresi logistik ordinal. Adapun pada variabel prediktor program studi (X_1) dan jalur penerimaan (X_5), terdapat beberapa kategori yang memiliki nilai statistik uji $|W| < 1{,}96$ seperti $X_{1(2)}, X_{1(3)}, X_{1(5)}$, dan $X_{5(3)}$, yang secara individu tidak signifikan. Namun demikian, variabel program studi (X_1) dan jalur penerimaan (X_5) tetap dinyatakan berpengaruh terhadap model secara keseluruhan karena merupakan bagian dari suatu satu kesatuan dengan variabel-variabel prediktor lainnya yang berpengaruh secara signifikan.

3.5 Interpretasi Model Regresi Logistik Ordinal Terbaik

Setelah didapatkan model regresi logistik ordinal terbaik maka selanjutnya dilakukan interpretasi terhadap model terbaik menggunakan nilai *odds ratio* yang diperoleh dengan rumus exp (β). Berdasarkan perhitungan dengan *software R*, diperoleh nilai *odds ratio* yang disajikan pada Tabel 13:

Tabel 12. Nilai *Odds Ratio*

Variabel Prediktor	Parameter	Taksiran Parameter	*Odds Ratio*
$X_{1(2)}$	β_1	1,400	0,246
$X_{1(3)}$	β_2	0,946	0,388
$X_{1(4)}$	β_3	1,111	0,329
$X_{1(5)}$	β_4	-0,934	2,544
$X_{1(6)}$	β_5	-0,322	1,380
X_3	β_6	0,961	0,382
$X_{5(2)}$	β_7	0,561	0,570
$X_{5(3)}$	β_8	1,641	0,193

Berdasarkan Tabel 12, interpretasi terhadap model regresi logistik ordinal terbaik adalah sebagai berikut:

1. Program Studi
   - Nilai *odds ratio* Program Studi Matematika sebesar 0,246 artinya lulusan Program Studi Matematika memiliki risiko untuk mendapatkan 3,00<IPK≤3,50 atau IPK>3,50 yaitu 0,246 kali lebih kecil dibandingkan lulusan Program Studi Geofisika.
   - Nilai *odds ratio* Program Studi Fisika sebesar 0,388 artinya lulusan Program Studi Fisika memiliki risiko untuk mendapatkan 3,00<IPK≤3,50 atau IPK>3,50 yaitu 0,388 kali lebih kecil dibandingkan lulusan Program Studi Geofisika.
   - Nilai *odds ratio* Program Studi Biologi sebesar 0,329 artinya lulusan Program Studi Biologi memiliki risiko untuk mendapatkan 3,00<IPK≤3,50 atau IPK>3,50 yaitu 0,329 kali lebih kecil dibandingkan lulusan Program Studi Geofisika.
   - Nilai *odds ratio* Program Studi Statistika sebesar 2,544 artinya lulusan Program Studi Statistika memiliki risiko untuk mendapatkan 3,00<IPK≤3,50 atau IPK>3,50 yaitu 2,544 kali lebih besar dibandingkan lulusan Program Studi Geofisika.
   - Nilai *odds ratio* Program Studi Kimia sebesar 1,380 artinya lulusan program studi kimia memiliki risiko untuk mendapatkan 3,00<IPK≤3,50 atau IPK>3,50 yaitu 1,380 kali lebih besar dibandingkan lulusan Program Studi Geofisika.
2. Usia
   - Nilai *odds ratio* usia sebesar 0,382 artinya setiap penambahan 1 tahun usia lulusan program sarjana FMIPA Unmul memiliki risiko sebesar 0,382 kali untuk mendapatkan 3,00<IPK≤3,50 atau IPK>3,50.
3. Jalur Penerimaan
   - Nilai *odds ratio* jalur penerimaan SBMPTN sebesar 0,570 artinya lulusan jalur penerimaan SBMPTN memiliki risiko untuk mendapatkan 3,00<IPK≤3,50 atau

IPK>3,50 yaitu 0,570 kali lebih kecil dibandingkan lulusan jalur penerimaan SNMPTN.

- Nilai *odds ratio* jalur penerimaan SMMPTN sebesar 0,193 artinya lulusan jalur penerimaan SMMPTN memiliki risiko untuk mendapatkan 3,00<IPK≤3,50 atau IPK>3,50 yaitu 0,193 kali lebih kecil dibandingkan lulusan jalur penerimaan SNMPTN.

4. Kesimpulan

1. Model regresi logistik ordinal terbaik berdasarkan metode *stepwise* dari indeks prestasi kumulatif lulusan program sarjana FMIPA Unmul adalah:
   1) $$\begin{aligned} Logit\left[P(Y_i \leq 1 \mid \mathbf{x}_i)\right] = g_1(\mathbf{x}_i) &= -26,745 + 1,400x_{i1(2)} + 0,946x_{i1(3)} + 1,111x_{i1(4)} \\ &\quad -0,934x_{i1(5)} - 0,322x_{i1(6)} + 0,961x_{i3} + 0,561x_{i5(2)} + 1,641x_{i5(3)} \\ &\quad ; i = 1,2,3,\ldots,251 \end{aligned}$$
   2) $$\begin{aligned} Logit\left[P(Y_i \leq 2 \mid \mathbf{x}_i)\right] = g_2(\mathbf{x}_i) &= -22,460 + 1,400x_{i1(2)} + 0,946x_{i1(3)} + 1,111x_{i1(4)} \\ &\quad -0,934x_{i1(5)} - 0,322x_{i1(6)} + 0,961x_{i3} + 0,561x_{i5(2)} + 1,641x_{i5(3)} \\ &\quad ; i = 1,2,3,\ldots,251 \end{aligned}$$
2. Faktor-faktor yang berpengaruh signifikan terhadap indeks prestasi kumulatif lulusan program sarjana FMIPA Unmul berdasarkan model regresi logistik ordinal terbaik adalah program studi (X_1), usia (X_3), dan jalur penerimaan (X_5).

Daftar Pustaka

[1] Bilder, C. R., & Loughin, T. M. *Analysis of Categorical Data with R*. Boca Raton: CRC Press, 2015.

[2] Agresti, A. *An Introduction to Categorical Data Analysis*. Florida: John Wiley & Sons, Inc., 2007.

[3] Utari, A., Islamiyati, A., & Thamrin, S. A. Pemodelan Regresi Logistik Ordinal dengan Dispersi Efek Lokasi. *ESTIMASI: Journal of Statistics and Its Application*, 4(2), 144–152, 2023.

[4] Fariha, F. N., & Subekti, R. Pemilihan Model Regresi Terbaik dalam Kasus Pengaruh Premi, Klaim, Hasil Investasi, dan Hasil Underwriting terhadap Laba Asuransi Jiwa (Studi Kasus PT Asuransi Jiwasraya). *Prosiding KNPMP III*, 674–684, 2018.

[5] Tampil, A. Y., Komalig, H., & Langi, Y. Analisis Regresi Logistik untuk Menentukan Faktor-Faktor yang Mempengaruhi Indeks Prestasi Kumulatif (IPK) Mahasiswa FMIPA Universitas Sam Ratulangi Manado. *JdC*, 6(2), 57–62, 2017.

[6] Amelia, R., Indahwati, & Erfiani. The Ordinal Logistic Regression Model with Sampling Weights on Data from the National Socio-Economic Survey. *BAREKENG: Journal of Mathematics and Its Applications*, 16(4), 1355–1364, 2022. https://doi.org/10.30598/barekengvol16iss4pp1355-1364

[7] Orvanita, Fathurahman, M., & Darnah. Regresi Logistik Ordinal untuk Memodelkan Predikat Lulusan Perguruan Tinggi. *Jurnal Statistika dan Aplikasinya*, 7(2), 116–128, 2023. https://doi.org/10.21009/JSA.07201

[8] Rathore, B., Sengupta, P., Biswas, B., & Kumar, A. Predicting the Price of Taxicab Using Artificial Intelligence: A Hybrid Approach Based on Clustering and Ordinal Regression Models. *Transportation Research Part E*, 2024. https://doi.org/10.1016/j.tre.2024.103530

[9] Gehringer, K. C., Martin, P. G., Van Calster, B., Hyrich, K. L., Verstappen, S. M. M., & Sergeant, J. C. How to Develop, Validate, and Update Clinical Prediction Models Using Multinomial Logistic Regression. *Journal of Clinical Epidemiology*, 2024. https://doi.org/10.1016/j.jclinepi.2024.111481

[10] Agresti, A. *Categorical Data Analysis*. New York: John Wiley & Sons, Inc., 2002.

[11] Gujarati, D. N., & Porter, D. C. *Basic Econometrics*. New York: McGraw-Hill, 2009.

[12] Handoyo, S., Pradianti, N., Nugroho, H. W., & Akri, J. Y. A Heuristic Feature Selection in Logistic Regression Modeling with Newton-Raphson and Gradient Descent Algorithm. *International Journal of Advanced Computer Science and Applications*, 13(3), 119–126, 2022.

[13] Permatasari, D. N., Fajar, A., Nurhaeni, S., Rahmawati, M., & Ramadhani, P. Hubungan Asosiasi antara Inner Child dengan Keharmonisan Keluarga: Pendekatan Menggunakan Uji Chi-Square (Uji Kebebasan). *Journal of Social Science Research*, 3(5), 5339–5349, 2023.

[14] Jones, S. *Learn to Use the Eta Coefficient Test in R with Data from the NIOSH Quality of Worklife Survey*. SAGE Publications, 2014.

[15] Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. *Applied Logistic Regression* (3rd ed.). Hoboken, NJ: John Wiley & Sons, Inc., 2013.

[16] Rahardiantoro, S., & Yanti, Y. Eksplorasi dan Analisis Regresi Logistik terhadap Kondisi Sungai Tercemar Limbah di Desa/Kelurahan Provinsi DKI Jakarta Indonesia. *Jurnal Matematika, Sains, dan Teknologi*, 23(1), 45–56, 2022.

[17] Berhanu, G., Dessalegn, B., Ali, H., & Animut, K. Determinants of Nutritional Status Among Primary School Students in Dilla Town: Application of an Ordinal Logistic Regression Model. *Journal (unspecified)*, 2023.

[18] Mayawi, Nurhayati, Talib, T., Bustan, W. A., & Laamena, S. N. Ordinal Logistic Regression Analysis of Factors Affecting the Blood Sugar Levels of Diabetes Mellitus Patients. *Pattimura International Journal of Mathematics*, 2(1), 33–42, 2023. https://doi.org/10.30598/pijmathvol2iss1pp33-42