PERBANDINGAN *GENARALIZED ESTIMATING EQUATIONS* DAN *GENERALIZED LINEAR MIXED MODELS* DALAM MENGANALISIS GANGGUAN PADA *AUTOMATIC TELLER MACHINE*

GILBERTY RUBEN

H051182326

PROGRAM STUDI STATISTIKA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN

2024

PERBANDINGAN *GENARALIZED ESTIMATING EQUATIONS* DAN *GENERALIZED LINEAR MIXED MODELS* DALAM MENGANALISIS GANGGUAN PADA *AUTOMATIC TELLER MACHINE*

GILBERTY RUBEN
H051181326

Skripsi

Diajukan sebagai salah satu syarat memperoleh gelar Sarjana Statistika

Program Studi Statistika

pada

PROGRAM STUDI STATISTIKA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
2024

SKRIPSI

PERBANDINGAN GENERALIZED ESTIMATING EQUATIONS DAN GENERALIZED LINEAR MIXED MODELS DALAM MENGANALISIS GANGGUAN PADA AUTOMATIC TELLER MACHINE

GILBERTY RUBEN

NIM. H051181326

Skripsi,

telah dipertahankan di depan Panitia Ujian Sarjana pada tanggal 12 Agustus 2024 dan dinyatakan telah memenuhi syarat kelulusan

pada

PROGRAM STUDI STATISTIKA

DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

Pembimbing Utama,	**Pembimbing Pertama,**
Prof. Dr. Dr. Georgina Maria Tinungki, M.Si.	**Siswanto, S.Si., M.Si.**
NIP. 19620926 198702 2001	**NIP. 19920107 201903 1000**

Ketua Departemen Statistika

Dr. Anna Islamiyati, S.Si., M.Si.

NIP. 19770808 200501 2002

PERNYATAAN KEASLIAN SKRIPSI DAN PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa, skripsi berjudul "Perbandingan *Generalizes Estimating Equations* dan *Generalized Linear Mixed Models* Dalam Menganalisis Gangguan Pada *Automatic Teller Machine*" adalah benar karya saya dengan arahan dari pembimbing Ibu Prof. Dr. Dr. Georgina Maria Tinungki, M.Si. sebagai Pembimbing Utama dan Bapak Siswanto, S.Si., M.Si. sebagai Pembimbing Pertama. Karya ilmiah ini belum diajukan dan tidak sedang diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka skripsi ini. Apabila di kemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan skripsi ini adalah karya orang lain, maka saya bersedia menerima sanksi atas pembuatan tersebut berdasarkan aturan yang berlaku.

Dengan ini saya melimpahkan hak cipta (hak ekonomis) dari karya tulis saya berupa skripsi ini kepada Universitas Hasanuddin.

Makassar, 12 Agustus 2024

GILBERTY RUBEN
NIM. H051181326

.

UCAPAN TERIMA KASIH

Puji syukur penulis panjatkan kepada Tuhan Yesus Kristus atas berkat, dan karunia-Nya sehingga penulis dapat menyelesaikan penyusunan skripsi ini. Puji Syukur senantiasa tercurahkan kepada Tuhan Yesus Kristus untuk berkat nikmat kesehatan baik itu secara fisik, akal pikiran, kesabaran, dan kemudahan yang diberikan sehingga penulis mampu menyelesaikan skripsi ini yang berjudul **"Perbandingan *Generalized Estimating Equations* dan *Generalized Linear Mixed Models* Dalam Menganalisis Gangguan Pada *Automatic Teller Machine*"** atas bimbingan dan arahan **Prof.Dr.Dr. Georgina Maria Tinungki, M.Si.** serta **Siswanto, S.Si,. M.Si.** Terima kasih kepada bapak **Andi Kresna Jaya, S.Si., M.Si.** sebagai dosen penguji serta ibu **Sri Astuti Thamrin, S.Si., M.Stat., Ph.D** sebagai sebagai dosen penguji sekaligus dosen pembimbing akademik yang telah meluangkan waktu dalam memberikan saran dan kritikan yang membangun dalam penyempurnaan tugas akhir ini.

Penulis menyadari bahwa dalam penyelesaian skripsi ini tidak lepas dari bantuan dan dorongan dari berbagai pihak yang senantiasa turut membantu dalam bentuk moril maupun materil sehingga dengan segala keterbatasan kemampuan dan pengetahuan, penulis dapat menyelesaikan skripsi ini. Oleh karena itu, penulis menyampaikan ucapan terima kasih yang setulus-tulusnya serta penghargaan yang setinggi-tingginya kepada orangtua penulis Ayahanda **Ruben Tulak Manukallo, SE., M.Si.** Ibunda **Milka Toding, SE** serta tante **dr. Anna Toding** dan tante **Achsa Toding** beserta seluruh keluarga yang telah memberikan dukungan penuh, pengorbanan luar biasa, limpahan cinta dan kasih sayang, kesabaran hati, serta dengan ikhlas menemani setiap langkah penulis dengan doa.

Penghargaan dan ucapan terima kasih yang tulus juga kepada orang-orang terdekat yang selalu memberikan dukungan, semangat, dan selalu ada menemani, mendengarkan setiap keluh kesah penulis dalam masa sulit semasa perkuliahan hingga penulisan skripsi (**Rival, Nehemia, Afick, Wardah, Rael, Fahira, Amel, Peres, Billy, Marsel**). Teman-teman **Statistika Unhas 2018** tak kalah penting kehadirannya, memberikan dukungan, semangat, maupun waktu kepada penulis.

Makassar, 12 Agustus 2024

GILBERTY RUBEN

ABSTRAK

GILBERTY RUBEN**. PERBANDINGAN *GENARALIZED ESTIMATING EQUATIONS* DAN *GENRRALIZED LINEAR MIXED MODELS* DALAM MENGANALISIS GANGGUAN PADA *AUTOMATIC TELLER MACHINE*** (dibimbing oleh Georgina Maria Tinungki dan Siswanto).

Latar Belakang. Mesin Anjungan Tunai Mandiri (ATM) memainkan peran penting dalam mendukung layanan perbankan yang efisien dan tanpa gangguan. Namun, gangguan yang terjadi pada mesin ATM dapat menyebabkan ketidaknyamanan bagi nasabah dan berdampak pada reputasi bank. Oleh karena itu, diperlukan analisis yang mendalam untuk memahami faktor-faktor yang mempengaruhi kinerja ATM dan mengidentifikasi langkah-langkah preventif untuk meminimalkan gangguan. **Tujuan.** Penelitian ini bertujuan untuk mengkaji meotde *Generalized Estimating Equations* (GEEs) yang berfokus pada estimasi efek rata-rata populasi dan tahan terhadap spesifikasi yang salah dalam struktur korelasi dan metode *Generalized Linear Mixed Models* (GLMMs) yang memberikan wawasan tentang efek rata-rata populasi dan efek spesifik individu dengan menggabungkan efek acak. **Metode**. Studi kasus dalam menganalisis gangguan yang terjadi pada mesin ATM membandingkan dua metode statistik, GEEs dan GLMMs dengan membandingkan nilai kriteria informasi pada tingkat keakuratan dan kompleksitas. **Hasil**. Hasil penelitian ini menunjukkan bahwa GEEs lebih efisien secara komputasional dan cocok untuk analisis tren keseluruhan, hal ini berdasarkan model terbaik dari GEEs dalam menganalisis gangguan pada mesin ATM menghasilkan nilai kriteria informasi terkecil yaitu 366,450. Dalam pemodelan GEEs diperoleh faktor yang berpengaruh secara signifikan yaitu pada variabel selisih hari terjadi gangguan dengan resiko peluang terjadi gangguan yaitu 0,000831, variabel status ATM dengan resiko peluang terjadi yaitu 0,0738, dan variabel status jaringan dengan resiko peluang terjadi yaitu 0,6072 saat terjadi gangguan ATM. **Kesimpulan**. Analisis perbandingan ini menunjukkan kelebihan dan keterbatasan kedua metode, sehingga memberikan panduan untuk memilih pendekatan yang tepat berdasarkan tujuan penelitian dan struktur data.

Kata Kunci: Anjungan Tunai Mandiri, Data Terkorelasi, Efek Acak, *Generalized Estimating Equations, Generalized Linear Mixed Models*.

ABSTRACT

GILBERY RUBEN. **COMPARISON OF GENERALIZED ESTIMATING EQUATIONS AND GENERALIZED LINEAR MIXED MODELS IN ANALYZING AUTOMATIC TELLER MACHINE MALFUNCTIONS** (supervised by Georgina Maria Tinungki and Siswanto).

Background. Automatic Teller Machines (ATMs) play a crucial role in supporting efficient and uninterrupted banking services. However, disruptions in ATM operations can cause inconvenience for customers and negatively impact the bank's reputation. Therefore, an in-depth analysis is necessary to understand the factors affecting ATM performance and to identify preventive measures to minimize disruptions. **Objective**. The objective of this study is to evaluate the Generalized Estimating Equations (GEEs) method, which focuses on estimating population-average effects and is robust against misspecification of the correlation structure, and the Generalized Linear Mixed Models (GLMMs) method, which provides insights into both population-average and individual-specific effects by incorporating random effects. **Method.** The case study in analyzing ATM disruptions compares these two statistical methods, GEEs and GLMMs, by comparing the information criteria values to assess accuracy and complexity. **Results.** The findings of this study indicate that GEEs are more computationally efficient and suitable for overall trend analysis, with the best GEE model for analyzing ATM disruptions yielding the lowest information criterion value of 366.450. In the GEE modeling, the factors that significantly influence the likelihood of ATM disruptions include the variable representing the difference in days when disruptions occur, with an odds ratio of 0.000831, the ATM status variable with an odds ratio of 0.0738, and the network status variable with an odds ratio of 0.6072 during ATM disruptions. **Conclusion.** This comparative analysis highlights the strengths and limitations of both methods, providing guidance for selecting the appropriate approach based on research objectives and data structure.

Keywords: Automatic Teller Machines, Correlated Data, Random Effects, Generalized Estimating Equations, Generalized Linear Mixed Models.

DAFTAR ISI

Halaman

DAFTAR TABEL

Tabel **Halaman**

DAFTAR GAMBAR

Gambar **Halaman**

DAFTAR LAMPIRAN

DAFTAR ISTILAH

Istilah	Arti dan Penjelasan
Asimtotis	Sifat perilaku suatu fungsi atau estimasi saat variabel atau ukuran sampel mendekti tak terhingga
Chi-square	Uji statistik yang digunakan untuk menguji hubungan antara dua variabel kategorikal atau untuk menguji kesesuaian distribusi data dengan distribusi yang di harapkan.
Cross-Sectional	Desain penelitian yang mengumpulkan data dari suatu populasi atau sampel pada satu tutuk waktu tertentu
Fixed effect	Asumsi dalam statistik bahwa variabel pengamatan yang berbeda memiliki padameter yang tetap dan konstan
Fungsi Densitas	Distribusi probabilitas dari variabel acak kontinu.
Global Minimizer	Titik pada domain fungsi dimana nilai fungsi mencapai minimum absolutnya, lebi rendah daripada di titik-titik lain di seluruh domain tersebut.
Link Function	Fungsi yang menghubungkan nilai ekspektasi dari variabel respon dengan kombinasi linear dari prediktor atau kovariat
Logit	Menghubungkan probabilitas kejadian dengan kombinasi linear dari prediktor.
Odds Ratio	Ukuran statistik yang digunakan untuk menentukan kekuatan atau besarnya hubungan antara dua kejadian atau variabel
Random effect	Variabel yang nilai-nilainya diambil dari suatu distribusi probabilitas tertentu
Scoring algorithm	Prosedur rumus yang digunakan untuk menghasilkan skor dari sekumpulan data atau respon

DAFTAR SINGKATAN DAN LAMBANG

LAMBANG/SINGKATAN	Arti dan Penjelasan
$\boldsymbol{B}_*$	Vektor parameter atau koefisien dari model referensi atau model alternatif
E_{M*}	Ekspektasi atau nilai harapan yang diambil terhadap distribusi dari model alternatif atau referensi
$\boldsymbol{M}_0$	Matriks informasi atau matriks Fisher, seringkali merupakan matriks informasi dari model yang diestimasikan
$\boldsymbol{M}_1$	Matriks informasi empiris yang diperoleh dari data aktual.
χ^2	*Chi-Square,* digunakan dalam pengujian hipotesis
∅	Parameter dispersi yang menggambarkan skala variabilitas dalam data.
Δ	Perubahan atau selisih
E	Ekspektasi atau nilai harapan
$L(\beta)$	Fungsi *likelihood*
Q	Nilai statistik *quasi-likelihood*
$\boldsymbol{R}$	Matriks varians kovarians dari error atau residual
Var	Varian, mengukur penyebaran data
W	Nilai Wald
b	Nilai efek acak
exp	Fungsi eksponensial
$l(\beta)$	*Log-Likelihood,* logaritma dari fungsi *likelihood*
$\boldsymbol{D}$	Matriks derivatif, pada GLMM berupakan Matriks varians kovarians dari efek acak
$\boldsymbol{I}$	Matriks identitas
$\boldsymbol{R}(\alpha)$	Struktur Korelasi
$\boldsymbol{Z}$	Matriks desain, berisi variabel prediktor
$\boldsymbol{\Omega}$	Matrik informasi Fisher
ε	*Residual error*
∂	Operatir diferensial, digunakan dalam konteks turunan
$\boldsymbol{\beta}$	Vektor koefisien parameter tetap yang diestimasi

LAMBANG/SINGKATAN	**Arti dan Penjelasan**
ψ	Fungsi kumulatif log-partisi untuk menentukan bentuk distribusi dari keluarga eksponensial
AIC	*Akaike Information Criterion*
ATM	*Automatic Teller Machine* (Anjungan Tunai Mandiri)
GEE	*Generalized Estimating Equations*
GLM	*Generalized Linear Model*
GLMM	*Generalized Linear Mixed Models*
LAN	*Local Area Network*
OR	*Odds Ratio*
SE	Standar Eror
QIC	*Quasi-Likelihood under the Independence Criterion*

BAB I
PENDAHULUAN

1.1 Latar Belakang

Generalized Linear Models (GLM) yang merupakan perluasan dari regresi linear klasik dengan variabel respon mengikuti distribusi dalam keluarga eksponensial dan mencakup berbagai distribusi yang umum digunakan seperti distribusi normal, binomial, dan poisson yang akan menghasilkan bentuk regresi seperti regresi linear, regresi logistik, dan regresi poisson. Hal ini dapat diartikan bahwa variabel respon pada GLM dapat bersifat kategorik, kontinu, diskrit, dan hitung. Hubungan antara distribusi yang terbentuk pada variabel dengan GLM adalah untuk dapat menentukan *link function* yang merupakan salah satu komponen dalam proses analisis dalam GLM (Hardin & Hilbe, 2018). Dalam kondisi tertentu saat dilakukan pengamatan yang dilakukan secara berulang terdapat struktur korelasi yang tidak diperhitungkan pada proses dalam GLM namun terdapat beberapa pengembangan dari bentuk GLM untuk mengatasi data yang berkorelai yaitu *Generalized Estimating Equations* (GEE) dan *Generalized Linear Mixed Models* (GLMM) (Fatemi dkk., 2018).

Metode GEE merupakan pengembangan dari model GLM yang dikembangkan oleh Liang dan Zeger (1986) dengan melakukan pendekatan alternatif dari model data yang berkorelasi. Pengembangan GEE dilakukan karena banyak metode regresi yang hanya dapat digunakan pada data independen sehingga dibutuhkan sebuah solusi untuk mengatasi masalah korelasi antar observasi dengan menggunakan asumsi korelasi dan *quasi-likelihood.* Menurut Hudecová & Pešta (2013), GEE dapat dilakukan dengan memenuhi tiga komponen dalam GEE yaitu fungsi penghubung, funsi varians, dan penentuan struktur korelasi. Sedangkan metode GLMM merupakan perluasan model GLM yang memiliki efek tetap dan ditambah suatu komponen acak. Model GLMM memiliki inti, yaitu menggabungkan *fixed effect* dan *random effect* (Datta & Ghost, 1991). Seperti GLM dan GEE, pembentukan model GLMM membutuhkan tiga komponen utama yaitu asumsi distribusi, komponen sistematik dan fungsi penghubung (*link function*) (Dobson, 2001). Model GLMM merupakan model yang fleksibel dan dapat diterapkan secara luas, salah satunya untuk menganalisis data longitudinal dan data klaster (Säfken et al., 2018).

Pengamatan longitudinal akan menghasilkan respon yang multivariat serta kemungkinan besar tidak saling independen antar pengamatan. Penyebabnya yaitu respon yang berulang cenderung berkorelasi antar respon satu dengan lainnya. Hal ini bertentangan dengan asumsi GLM, metode yang dapat mengatasi permasalahan ini yaitu GEE dan GLMM. Model dalam metode ini dapat mengatasi ketidak saling bebasan dengan struktur korelasi yang berbeda-beda tergantung pada matriks varian-kovariannya, dengan tujuan untuk mengetahui pola korelasi yang terbentuk dari respon (Hidayati et al., 2014). Berdasarkan struktur data longitudinal menurut Mardianto dkk., (2021) pengamatan yang dilakukan lebih dari satu untuk setiap subjek dan hubungan antar subjek diasumsikan saling independen namun antar observasi pada subjek saling bergantung sehingga terjadi korelasi. Maka salah satu penerapan data longitudinal yaitu dengan melakukan pengamatan pada sejumlah mesin *Auto Teller Machine* (ATM) dengan pengamatan yang dilakukan pada gangguan yang terjadi sebanyak titik waktu yang telah di tentukan.

ATM merupakan salah satu produk dan layanan perbankan yang paling umum digunakan oleh masyarakat. Perkembangan teknologi menyebabkan mesin ATM ikut berkembang dalam memberikan kemudahan layanan kepada nasabah. Saat ini, mesin ATM tidak hanya memberikan layanan tarik tunai dan cek saldo saja, melainkan juga layanan *transfer* antar rekening, setor tunai, pembelian pulsa, pembayaran tagihan, dan lain-lain (Herdi & Dores, 2020). Selain layanan yang semakin bertambah, persebaran mesin ATM yang semakin merata di Indonesia juga merupakan dampak perkembangan teknologi yang memudahkan nasabah untuk melakukan transaksi tanpa batas ruang dan waktu. Seiring dengan kemudahan dalam layanan ATM, penggunaan mesin ATM yang terlalu sering dapat menyebabkan terjadinya gangguan pada mesin ATM. Beberapa faktor yang mengakibatkan mesin ATM menjadi terganggu yaitu uang sangkut (*cash handler fatal error*), kartu sangkut (*card reader fatal error*), kertas struk sangkut (*receipt fatal error*), jaringan komunikasi terputus serta beberapa faktor lainnya (Rinaldi, 2019).

Penelitian ini menggunakan variabel respon yaitu Gangguan ATM (terjadi gangguan *hardware* atau tidak), sedangkan variabel prediktor yaitu selisih hari terjadinya gangguan, status jaringan dan status ATM untuk mengetahui pengaruh dan perubahan yang terjadi pada ATM saat gangguan terjadi. Pemilihan variabel merujuk pada penelitian terdahulu oleh Lestari (2018) melakukan perbandingan metode regresi logistik dan GEE dalam mengatasi penyebab gangguan ATM, melalui penelitian tersebut metode GEE lebih tepat digunakan dalam menganalisis penyebab gangguan ATM. Penelitian terkait GLMM juga diteliti oleh Sastri & Setiadi (2018) untuk data kematian bayi di Indonesia, melalui penelitian tersebut model GLMM yang terbentuk sesuai untuk memodelkan data kematian bayi di Indonesia. Serta penelitian Sihombing dkk., (2022) yang melakukan perbandingan metode GEE dan GLMM pada data longitudinal, studi kasus persentase penduduk miskin di Indonesia, melalui penelitian tersebut metode GLMM lebih baik dibandingkan GEE dalam memodelkan presentase penduduk miskin di Indonesia. Berdasarkan beberapa penelitian terdahulu dan uraian latar belakang tersebut, peneliti tertarik untuk melakukan perbandingan metode GEE dan GLMM dalam menganalisis gangguan pada ATM.

1.2 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Bagaimana perbandingan penerapan metode GEE dan GLMM dalam menganalisis gangguan Pada mesin ATM ?
2. Faktor-faktor apa saja yang berpengaruh saat terjadi gangguan pada mesin ATM?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Memperoleh hasil perbandingan penerapan metode GEE dan GLMM dalam menganalisis gangguan pada mesin ATM.
2. Memperoleh faktor-faktor apa saja yang berpengaruh saat gangguan terjadi pada mesin ATM.

1.4 Manfaat Penelitian

1. Untuk menambah pengetahuan dalam penerapan metode GEE dan GLMM terkait dengan menganalisis dan mengestimasi.
2. Menambah wawasan bagi masyarakat tentang apa saja gangguan yang sering terjadi pada ATM dan apa saja yang memberikan pengaruh saat terjadi gangguan pada mesin ATM,
3. Sebagai saran, masukan, dan sebagai sumber informasi bagi semua bank terlebih khusus untuk Bank Negara Indonesia dalam mengambil keputusan saat terjadi gangguan pada mesin ATM.

1.5 Teori

1.5.1 Data Longitudinal

Data longitudinal adalah data yang didapatkan dengan pengamatan yang dilakukan secara berulang pada beberapa *cross-sectional,* pengamatan yang dilakukan yaitu terhadap individu atau dapat diartikan sebagai jumlah sampel dalam beberapa titik waktu. Data *cross-sectional* berasal dari pengamatan yang dilakukan pada beberapa individu berbeda (Twisk, 2003). Data longitudinal banyak diaplikasikan dalam bidang pendidikan, kesehatan, sosial dan ekonomi. Data longitudinal lebih kompleks dan lebih baik dalam menemukan jawaban tentang dinamika perubahan dan menyediakan informasi yang lengkap, tergantung pada pengoperasian teori dan metodologi penelitian (Ramli et al., 2020).

D Dalam data longitudinal dipresentasikan dalam bentuk formulasi matematis untuk mempermudah dalam menganalisis. Notasi yang di perlukan dalam pemodelan data longitudinal sebagai berikut. Ada sebanyak $i = 1, \dots, m$ individu, yang masing-masing dilakukan pengamatan berulang $j = 1,2, \dots, n_i$. Sehingga untuk keseluruhan ada $N = \sum_{i=1}^{m} n_i$ observasi. Waktu observasi aktual, yaitu pada pengamatan yang dilakukan dinotasikan dengan t_{ij}. Variabel respon dinyatakan sebagai Y_{ij} dengan nilai observasi y_{ij} atau dapat dinyatakan sebagai :

$$\boldsymbol{Y}_i = \begin{bmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{in_i} \end{bmatrix}$$

atau $\boldsymbol{Y}_i = (\boldsymbol{Y}_{i1} \cdots \boldsymbol{Y}_{in})^t$ dengan nilai observasi $\boldsymbol{y}_i = (\boldsymbol{y}_{i1} \cdots \boldsymbol{y}_{in})^t$. Variabel penjelas dinyatakan sebagai

$$\boldsymbol{X}_i = \begin{bmatrix} x_{i11} & \cdots & x_{i1p} \\ x_{i21} & \cdots & x_{i1p} \\ \vdots & \ddots & \vdots \\ x_{in1} & \cdots & x_{inp} \end{bmatrix}$$

Matriks dengan ukuran $p \times n_i$ dengan p adalah banyaknya kovariat. Untuk menunjukkan semua nilai kovariat pada satu observasi dapat ditulis sebagai vektor $\boldsymbol{x}_i = (\boldsymbol{x}_{ij1} \quad \cdots \quad \boldsymbol{x}_{ijp})^t$ berukuran $p \times 1$. Variabel penjelas pada data longitudinal yang diamati sekali dan nilainya sama dari awal studi hingga akhir studi disebut dengan kovariat awal (*baseline covariate*), untuk variabel penjelas dapat juga diamati lebih dari satu kali selama studi berjalan yang disebut dengan kovariat bergantung waktu (*time-varying covariate*) (Hermawan, 2020).

1.5.2 Generalized Linear Model

Generalized Linear Model (GLM) merupakan pengembangan dari model linear klasik yang diasumsikan memiliki distribusi yang termasuk ke dalam keluarga eksponensial. Diperkenalkan pertama kali oleh Nelder & Wedderburn (1972) GLM di kembangkan untuk mengatasi masalah peubah respon yang tidak berdistribusi normal. Menurut Agresti (2002), terdapat tiga komponen utama dalam menjalankan GLM, yaitu :

1. *Random component,* yaitu peubah respon Y saling bebas dan termasuk keluarga eksponensial.
2. *Systematic component,* merupakan fungsi dari peubah penjelas.
3. *Link Function,* merupakan fungsi dari nilai tengah *Random component* dengan *Systematic component*, menghubungkan μ_i pada η_i dengan $\eta_i = g(\mu_i)$ Maka g menghubungkan $E(Y_i)$ pada peubah penjelas seperti pada persamaan 1 berikut:

$$g(\mu_i) = \sum_j \beta_j x_{ij}\,, i = 1, \ldots n \tag{1}$$

dengan,

$g(\mu_i)$: *link function*

μ_i : nilai tengah ke$-i$

β_j : koefisien regresi untuk masing-masing variabel independen x_{ji}

x_{ij} : prediktor ke$-i$ pengamatan ke$-j$

1.5.3 Regresi Logistik

Regresi logistik merupakan model yang dapat digunakan untuk mengetahui hubungan antara variabel tak bebas yang bersifat kategorik maupun variabel bebas (Prasad & Muralidhar, 2022). Variabel tak bebas y terdiri dari 2 kategori yang dinotasikan dengan angka 0 dan 1 yaitu kategori “gagal atau tidak” dinotasikan sebagai angka 0 sedangkan untuk kategori “sukses atau ya” di notasikan sebagai 1 (Newita & Martha, 2018). Regresi logistik dibagi menjadi beberapa jenis dan penggunaannya tergantung pada jenis data dan situasi yang digunakan yaitu regresi logistik biner, regresi logistik multinominal, dan regresi logistik ordinal. Regresi logistik biner digunakan untuk menganalisis hubungan antara variabel respon dan variabel respon biner atau berdistribusi binomial (Huang, 2022). Dalam pemodelan regresi logistik, *link function* digunakan untuk memodelkan bagaimana respon yang diukur pada skala yang berbeda berkaitan dengan prediktor. *Link function* memungkinkan untuk memodelkan respon yang tidak terdistribusi normal dan memiliki varians yang tidak konstan (Uanhoro et al., 2021). Teberapa jenis *link function* yang digunakan dalam pemodelan regresi. Namum untuk regresi logistik digunakan *logit link function.* Bentuk umum dari regresi logistik dengan *logit link function* dapat dilihat pada Persamaan 2 berikut:

$$logit(\mu_i) = \ln \frac{\mu_i}{1-\mu_i} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_i x_{ij} \tag{2}$$

atau

$$\frac{\mu_i}{1-\mu_i} = \exp\left(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_i x_{ij}\right)$$

atau bentuk *inverse link* nya pada Persamaan 3:

$$\mu_i = \frac{\exp\left(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_i x_{ij}\right)}{1 + \exp\left(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_i x_{ij}\right)} \tag{3}$$

dengan,

β_0 : nilai konstanta dari model regresi

β_i : koefisien regresi untuk masing-masing variabel independen x_{ij}

x_{ij} : prediktor ke$-i$ pengamatan ke$-j$

μ_i : probabilitas sukses dalam regresi logistik. $\mu_i = P(Y=1|j) = 1 - P(Y=0|j)$

1.5.4 Generalized Estimating Equations

Generalized Estimating Equations (GEE) adalah perluasan dari *Generalized Linear Model* (GLM) yang digunakan untuk mengolah data yang berkorelasi. Model GLM mengasumsikan multivariat normalitas, sehingga asumsi normal akan menjadi faktor pembeda utama antara GEE dan GLM saat menggunakan data longitudinal yang bersifat kontinu (Tantular et al., 2018). Bentuk persamaan umum GEE pada persamaan 4 berikut :

$$\mathrm{U}(\beta) = \sum_{i=1}^{n} \boldsymbol{D}_i^T \boldsymbol{V}_i^{-1} (Y_i - \mu_i) \tag{4}$$

dengan,

$\mathrm{U}(\beta)$: Persamaan skor yang harus memenuhi nilai nol untuk estimasi parameter β

$\boldsymbol{D}_i^T$:Transpose dari matriks $\boldsymbol{D}_i$ yang mencerminkan hubungan antara parameter dan respon

$\boldsymbol{V}_i^{-1}$: Invers dari matriks varian-kovarian $\boldsymbol{V}_i$, yang menggambar-kan struktur variabilitas dalam data

Y_i : Variabel dependen pada observasi ke$-i$

μ_i : Mean dari distribusi keluarga yang bergantung pada distribusi yang dipilih dalam GLM

$\boldsymbol{D}_i = \frac{\partial \mu_i}{\partial \beta}$ didefinisikan sebagai matriks yang memiliki turunan sementara $\boldsymbol{V}_i = \boldsymbol{A}^{1/2}\boldsymbol{R}(\boldsymbol{\alpha})\boldsymbol{A}^{1/2}$ sebagai matriks kovarians dan jika memiliki varian konstan pada titik waktu yang berbeda kita dapat menganggap bentuk ini sebagai $\sigma^2\boldsymbol{R}(\boldsymbol{\alpha})$ dimana $\boldsymbol{R}(\boldsymbol{\alpha})$ merupakan struktur korelasi yang akan digunakan. μ_i merupakan *link function* dari bentuk GLM yaitu $g(\mu_i) = \boldsymbol{X}'\beta$. Jadi dalam menentukan GEE perlu untuk menentukan *link function* yang telah dideskripsikan kemudian menentukan fungsi varians $\phi V(\mu_i)$ yaitu data kontinu digunakan berdasarkan distribusi normal dan distribusi normal menyatakan fungsi varians konstan sehingga dapat disebutkan $V(\mu_i) = 1$ (Hardin & Hilbe, 2003).

Korelasi dibentuk dalam matriks korelasi $\boldsymbol{R}(\boldsymbol{\alpha})$ dengan ukuran $n \times n$, dan struktur dari matriks korelasi ini tidak diketahui dan harus diduga. Pada umumnya struktur korelasi yang sering digunakan dalam GEE adalah *independent*, *exchangeable, autoagressive AR(1),* dan *unstructured* yaitu bentuk korelasi masing-masing adalah sebagai berikut (Pekár & Brabec, 2018) :

1. Struktur korelasi *independent*

   Setiap subjek atau setiap observasi dari setiap subjek yang sama pada dasarnya bersifat independen atau tidak ada hubungan korelasi terhadap observasi lainnya. Bentuk struktur korelasi *independent* dapat dilihat pada Persamaan 5 berikut:

$$\boldsymbol{R}(\boldsymbol{\alpha}) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} Corr\left(Y_{ij}, Y_{ik}\right) = \begin{cases} 1 & j = k \\ 0 & j \neq k \end{cases} \tag{5}$$

2. Struktur korelasi *exchangeable*

   Ada korelasi tetap yang terjadi antara observasi dari setiap subjek. Observasi yang dilakukan di awal hingga akhir memiliki korelasi yang sama. Bentuk struktur korelasi *exchangeable* dapat dilihat pada Persamaan 6 berikut:

$$\boldsymbol{R}(\boldsymbol{\alpha}) = \begin{bmatrix} 1 & \alpha & \alpha \\ \alpha & 1 & \alpha \\ \alpha & \alpha & 1 \end{bmatrix} Corr\left(Y_{ij}, Y_{ik}\right) = \begin{cases} 1 & j = k \\ \alpha & j \neq k \end{cases} \tag{6}$$

3. Struktur korelasi *autoagressive AR(1)*

   Korelasi antara pengamatan dalam kelompok yang sama bergantung pada jarak waktu antara pengamatan, semakin banyak jumlah observasi maka korelasinya akan berkurang. Bentuk struktur korelasi *AR(1)* dapat dilihat pada Persamaan 7 berikut:

$$\boldsymbol{R}(\boldsymbol{\alpha}) = \begin{bmatrix} 1 & \alpha & \alpha^2 \\ \alpha & 1 & \alpha \\ \alpha^2 & \alpha & 1 \end{bmatrix} Corr\left(Y_{ij}, Y_{ik}\right) = \alpha^m, m = 0,1,2, \dots, n_i - j \tag{7}$$

4. Struktur korelasi *unstructured*

   Semua observasi dalam kelompok yang sama memiliki korelasi yang berbeda-beda. Bentuk struktur korelasi *unstructured* dapat dilihat pada Persamaan 8 berikut:

$$\boldsymbol{R}(\boldsymbol{\alpha}) = \begin{bmatrix} 1 & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & 1 & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & 1 \end{bmatrix} Corr(Y_{ij}, Y_{ik}) = \begin{cases} 1 & j = k \\ \alpha_{jk} & j \neq k \end{cases} \tag{8}$$

Menurut Burton dkk., dalam Twisk (2003), pendugaan parameter meliputi tiga tahapan yaitu:

1. Pendugaan parameter $\hat{\beta}$ dengan GLM.
2. Pembentukan matriks korelasi $\boldsymbol{R}(\boldsymbol{\alpha})$
3. Iterasi pendugaan parameter. Pendugaan parameter tidak dapat dilakukan secara langsung, sehingga memerlukan iterasi *Newton Raphson*. Bentuk umum iterasi *Newton Raphson* dapat dilihat pada Persamaan 9 berikut:

$$\beta^{t+1} = \beta^{t} + \left[\sum_{i=1}^{N} \left(\frac{\partial \mu_i}{\partial \beta}\right)' \boldsymbol{V}_{\boldsymbol{i}}^{-1} \frac{\partial \mu_i}{\partial \beta}\right]^{-1} \left[\sum_{i=1}^{N} \left(\frac{\partial \mu_i}{\partial \beta}\right)' \boldsymbol{V}_{\boldsymbol{i}}^{-1} (y_i - \mu_i)\right] \tag{9}$$

   dengan,

   β : parameter β

   $\boldsymbol{V}_i$: matriks ragam peragam bagi y_i

   μ_i : nilai tengah ke$-i$

   y_i : nilai pengamatan ke$-i$

1.5.5 *Quasi-Likelihood under the Independence Criterion*

Quasi-likelihood under the Independence Criterion (QIC) merupakan metode untuk memilih model dalam analisis data. QIC merupakan hasil modifikasi dari AIC untuk digunakan dalam proses analisis GEE untuk memilih struktur korelasi yang terbaik. Nilai dari QIC juga digunakan untuk memastikan *goodness of fit* dari model dengan melihat nilai QIC terkecil dari masing-masing model yang dibandingkan karena memiliki kesalahan estimasi parameter yang paling kecil dan jumlah parameter paling sedikit. Bentuk QIC untuk GEE dapat dilihat pada Persamaan 10 (Nyabwanga et al., 2019) :

$$QIC = -2Q\left(\left(\hat{\beta}(\boldsymbol{R}); I, D\right)\right) + 2trace\left(\widehat{\boldsymbol{\Omega}}_{\boldsymbol{I}} \widehat{\boldsymbol{V}}_{\boldsymbol{r}}\right) \tag{10}$$

dengan $\widehat{\boldsymbol{\Omega}}_I$ dan $\widehat{\mathbf{V}}_r$ pada persamaan 11 dan 12 berikut :

$$\widehat{\boldsymbol{\Omega}}_I = \sum_{i=1}^{n} \boldsymbol{D}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{D}_i \tag{11}$$

$$\widehat{\mathbf{V}}_r = \widehat{\boldsymbol{\Omega}}_I \sum_{i=1}^{n} \boldsymbol{D}_i^T \boldsymbol{V}_i^{-1} Cov[y_i] \boldsymbol{V}_i^{-1} \boldsymbol{D}_i \, \widehat{\boldsymbol{\Omega}}_I \tag{12}$$

dengan,

$\widehat{\boldsymbol{\Omega}}_I$: perkiraan matriks varians kovarian antar observasi

$\widehat{\mathbf{V}}_r$: perkiraan matriks varians kovarian dari estimasi parameter menggunakan asumsi matriks korelasi $\boldsymbol{R}(\boldsymbol{\alpha})$

$Q\left(y_i; \hat{\beta}(\boldsymbol{R}(\boldsymbol{\alpha}))\right)$: nilai *quasi-likelihood* yang diestimasi dengan parameter $\hat{\beta}$ berdasarkan struktur korelasi $\boldsymbol{R}(\boldsymbol{\alpha})$

$trace(\widehat{\boldsymbol{\Omega}}_I \widehat{\mathbf{V}}_r)$: penjumlahan elemen-elemen dari hasil perkalian $\widehat{\boldsymbol{\Omega}}_I$ dan $\widehat{\mathbf{V}}_r$

$Cov[y_i]$: matriks kovarians respon (y_i)

1.5.6 *Generalized Linear Mixed Models*

Generalized Linear Mixed Models (GLMM) merupakan perluasan dari *Generalized Linear Model* untuk data longitudinal dengan penambahan efek acak pada persamaan modelnya (Fokkema et al., 2020). Secara umum model GLMM dapat dilihat pada Persamaan 13 berikut:

$$g(\mu_{ij}) = \boldsymbol{X}^T\boldsymbol{\beta} + \boldsymbol{Z}^T b_i + \varepsilon_{ij} \tag{13}$$

dengan,

$g(\mu_{ij})$: *link function* yang menghubungkan mean μ_{ij} dengan kombinasi linear dari parameter β, matriks desain $\boldsymbol{X}$,efek acak $\boldsymbol{Z}$, efek acak b_i, dan *error* ε_{ij}

μ_{ij} : nilai tengah untuk subyek ke$-i$, pengamatan ke$-j$

$\boldsymbol{X}$: matriks kovariat efek tetap untuk subyek ke$-i$, pengamatan ke$-j$

$\boldsymbol{\beta}$: vektor parameter efek tetap

$\boldsymbol{Z}$: vektor kovariat efek acak untuk subyek ke$-i$, pengamatan ke$-j$

b_i : efek acak untuk subyek ke$-i$

ε_{ij} : *error* untuk subyek ke$-i$, pengamatan ke$-j$

Pada efek acak b_0 diasumsikan menyebar normal dengan rata-rata adalah 0 dan varians σ_b^2 untuk efek acak intersep, atau $b_0 \sim N(0, \sigma_b^2)$. Regresi logistik (data binomial) pada model GLMM untuk penelitian ini pada Persamaan 14 berikut:

$$log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) = \boldsymbol{X}_{ij}^T\widetilde{\boldsymbol{\beta}} + \boldsymbol{Z}_{ij}^T b_i + \varepsilon_{ij}\,, \qquad 1 < i < n, 1 < j < m \tag{14}$$

dengan,

$log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right)$: model *link function* logit

$\boldsymbol{X}$: vektor kovariat efek tetap untuk subyek ke$-i$, pengamatan ke$-j$

$\widetilde{\boldsymbol{\beta}}$: parameter efek tetap (koefisien regresi)

$\boldsymbol{Z}$: vektor kovariat efek acak untuk subyek ke$-i$, pengamatan ke$-j$

b_i : efek acak untuk subyek ke$-i$

ε_{ij} : *error* untuk subyek ke$-i$, pengamatan ke$-j$

Estimasi parameter GLMM dapat dilakukan menggunakan metode maksimum *likelihood* untuk mengestimasi parameter β. GLMM mempunyai asumsi pada persamaan 15 hingga 18 sebagai berikut (Rohmaniah & Chandra, 2018):

$$E(b_i) = 0 \tag{15}$$

$$E(\varepsilon_{ij}) = 0 \tag{16}$$

$$Var(b_i) = E(b_i b_i') = \boldsymbol{D} \quad (17)$$
$$Var(\varepsilon_{ij}) = E(\varepsilon_{ij}\varepsilon_{ij}') = \boldsymbol{R} \quad (18)$$

dengan,

b_i dan ε_{ij} : *independently distributed*

D dan *R* : matriks varians kovarians

Karena b_i dan ε_{ij} saling independen, maka $corr(b_i \varepsilon_{ij}) = 0$, sehingga $cov(b_i \varepsilon_{ij}) = E(b_i \varepsilon_{ij}') = E(b_i', \varepsilon_{ij}) = 0$. Mean kovarian $g(\mu_{ij})$ adalah pada persamaan 19 berikut:

$$\begin{aligned} E\left(g(\mu_{ij})\right) &= E(\boldsymbol{X}^T\boldsymbol{\beta} + \boldsymbol{Z}^T b_i + \varepsilon_{ij}) \\ &= E(\boldsymbol{X}^T\boldsymbol{\beta}) + E(\boldsymbol{Z}^T b_i) + E(\varepsilon_{ij}) \\ &= \boldsymbol{X}^T\boldsymbol{\beta} + \boldsymbol{Z}^T E(b_i) \\ &= \boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{\beta} \end{aligned} \quad (19)$$

$$\begin{aligned} cov\left(g(\mu_{ij})\right) &= cov(\boldsymbol{X}^T\boldsymbol{\beta} + \boldsymbol{Z}^T b_i + \varepsilon_{ij}) \\ &= cov(\boldsymbol{Z}^T b_i + \varepsilon_{ij}) \\ &= cov(\boldsymbol{Z}^T b_i) + cov(\varepsilon_{ij}) + cov(\boldsymbol{Z}^T b_i, \varepsilon_{ij}) + \\ &\quad cov(\varepsilon_{ij}, \boldsymbol{Z}^T b_i) \\ &= \boldsymbol{Z}cov(b_i)\boldsymbol{Z}^T + cov(\varepsilon_{ij}) + \boldsymbol{Z}cov(b_i, \varepsilon_{ij}) + \\ &\quad cov(\varepsilon_{ij}, b_i)\boldsymbol{Z}^T \\ &= \boldsymbol{ZDZ}' + \boldsymbol{R} \end{aligned} \quad (20)$$

Estimasi parameter dalam GLMM menggunakan maksimum *likelihood* mengasumsikan bahwa (Rohmaniah & Chandra, 2018):

1. $g(\mu_i)$ mengikuti distribusi dari keluarga eksponensial dengan densitas $f\left(g(\mu_i)\right)$.
2. $g(\mu_{ij})$ independen satu sama lain dengan diberikannya b_i.
3. b_i adalah independen dan berdistribusi identik Gaussian dengan mean nol dan $Var = (b_i) = \sigma_b^2 = D$

Estimasi maksimum *likelihood* dari β diperoleh dengan menentukan pembuat nol dari derivatif parsial persamaan terhadap β yang dapat kita lihat pada Persamaan 22 berikut:

$$\frac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = 0 \quad (21)$$

$$\begin{aligned} \boldsymbol{X}^T(\boldsymbol{ZDZ}^T + \boldsymbol{R})^{-1}\Delta^{-1}(Y - \mu) &= 0 \\ \boldsymbol{X}^T(\boldsymbol{ZDZ}^T + \boldsymbol{R})^{-1}\Delta^{-1}Y &= \boldsymbol{X}^{\boldsymbol{T}}(\boldsymbol{ZDZ}^T + \boldsymbol{R})^{-1}\Delta^{-1}\mu \end{aligned} \quad (22)$$

Persamaan di atas merupakan fungsi nonlinear sehingga tidak dapat diselesaikan secara analitik, sehingga digunakan suatu metode numerik salah satunya *Newton-Raphson*. Turunan kedua fungsi *log-likelihood* $l(\beta)$ pada Persamaan 23 berikut:

$$\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \beta \partial \boldsymbol{\beta}'} = -\boldsymbol{X}^T(\boldsymbol{ZDZ}^T + \boldsymbol{R})^{-1}\Delta^{-1}\frac{\partial \mu}{\partial \beta'} + \boldsymbol{X}^T\frac{\partial(\boldsymbol{ZDZ}^T + \boldsymbol{R})^{-1}\Delta^{-1}}{\partial \beta^T}(Y - \mu) \quad (23)$$

maka,

$$E\left[\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}\right] = -\boldsymbol{X}^T(\boldsymbol{ZDZ}^T + \boldsymbol{R})^{-1}\Delta^{-1}\frac{\partial \mu}{\partial \boldsymbol{\beta}'} + 0$$

$$= -\boldsymbol{X}^T(\boldsymbol{Z}D\boldsymbol{Z}^T + \boldsymbol{R})^{-1}\Delta^{-1}\left(\frac{\partial g(\mu_i)}{\partial \mu_i}\right)\boldsymbol{X}$$
$$= -\boldsymbol{X}^T(\boldsymbol{ZDZ}^T + \boldsymbol{R})^{-1}\Delta^{-1}\Delta\boldsymbol{X}$$
$$= -\boldsymbol{X}'(\boldsymbol{ZDZ}' + \boldsymbol{R})^{-1}\boldsymbol{X} \quad (24)$$

1.5.7 *Wald Test*

Uji wald menunjukkan seberapa berpengaruh variabel independen secara parsial dalam menerangkan variabel dependen, dalam artian uji wald sebagai uji signifikansi parameter dalam model yang digunakan. Uji wald didasarkan pada perbandingan antara perbedaan koefisien dan standar *error* dari koefisien tersebut dengan distribusi standar normal. Dapat dituliskan sebagai berikut:

H_0 : $\beta_j = 0$ (parameter β tidak berpengaruh secara signifikan terhadap variabel respon secara individu)

H_1 : $\beta_j \neq 0; j = 1,2, \dots, p$ (parameter β berpengaruh secara signifikan terhadap variabel respon secara individu). Dengan bentuk statistik uji wald pada Persamaan 25 berikut (Putri et al., 2021) :

$$W = \frac{(\hat{\beta}_i)}{[SE(\hat{\beta}_i)]} \quad (25)$$

dengan,

β_i : nilai koefisien untuk variabel ke$-i$

$\hat{\beta}_i$: penduga parameter β

$SE(\hat{\beta}_i)$: nilai standar *error* untuk variabel ke$-i$

Tolak H_0 , jika $|W| > Z$ yang berarti parameter β berpengaruh secara signifikan terhadap variabel respon secara individu.

1.5.8 Uji Kesesuaian Model

Uji kesesuaian model dilakukan untuk mengetahui apakah model yang dihasilkan sudah sesuai atau belum.

Hipotesis:

H_0: Model sesuai

H_1 : Model tidak sesuai

Menurut Nugraha (2016), statistik uji untuk kesesuaian model menggunakan statistik devians (D) seperti pada Persamaan 26 sebagai berikut.

$$D = -2\sum_{i=1}^{p}\left[y_{ij}\, ln\left(\frac{n_i\hat{\mu}_{ij}}{y_{ij}}\right) + (n_i - y_{ij}) ln\left(\frac{n_i - \hat{\mu}_{ij}}{(n_i - y_{ij})}\right)\right] \quad (26)$$

Keterangan:

$\hat{\mu}_{ij}$: peluang prediksi Y ke-i pada kategori-j

y_{ij} : pengamatan/observasi Y ke-i pada kategori-j

p : jumlah parameter pada model

n_i : jumlah pengamatan/observasi

Statistik deviasns (D) menyebar menurut distribusi khi-kuadrat dengan Keputusan tolak H0, jika $\chi^2{}_{\mathrm{p}} > \chi^2{}_{\mathrm{df}}$. Hipotesis nol ditolak maka disimpulkan model yang diperoleh tidak sesuai, dan sebaliknya Hipotesis nol diterima yang maka disimpulkan model yang diperoleh sudah sesuai.

1.5.9 Pemilihan Model Terbaik

Pemilihan model terbaik menggunakan *Akaike Information Criterion* (AIC) yang merupakan teknik pemilihan model, diusul oleh Hirotogu Akaike (1973). Dalam beberapa hal, AIC memperluas maksimum *likelihood* dengan membuat *framework* yang mencakup estimasi parameter dan pemilihan model. Prinsip AIC yaitu berdasarkan jarak *Kullback-Leibler* (KLD; Kullback dan Leibler, 1995), yang dapat digunakan untuk menghitung jarak antara kepadatan sebenarnya (biasanya tidak diketahui) $g(\boldsymbol{y})$ dan model parametrik $f(\boldsymbol{y}|\boldsymbol{v})$. Paramater yang tidak diketahui $\boldsymbol{v}$ biasnya diestimasi menggunakan maksimum *likelihood* $\widehat{\boldsymbol{v}}(\boldsymbol{y})$. Perhitungan AIC dijabarkan pada Persamaan 27 berikut (Säfken et al., 2018) :

$$AIC(\boldsymbol{y}) = -2logf\left(\boldsymbol{y}\middle|\widehat{\boldsymbol{v}}(\boldsymbol{y})\right) + 2\dim(\boldsymbol{v}) \qquad (27)$$

dengan,

$AIC(\boldsymbol{y})$: nilai AIC
$logf\left(\boldsymbol{y}\middle|\widehat{\boldsymbol{v}}(\boldsymbol{y})\right)$: *log-likelihood* maksimal
$\dim(\boldsymbol{v})$: dimensi $\boldsymbol{v}$

1.5.10 *Odds Ratio*

Odds merupakan penyajian probabilitas yang menjelaskan probabilitas suatu kejadian akan terjadi akan dibagi dengan probabilitas suatu kejadian tidak akan terjadi. Rasio probabilitas *odds* untuk sukses diberi notasi (π), sedangkan rasio probabilitas untuk gagal diberi notasi $(1-\pi)$. Perhitungan *odds* untuk populasi dijabarkan pada Persamaan 28 sebagai berikut (Nugraha, 2016).

$$\pi = \frac{\pi}{1-\pi} \qquad (28)$$

Sedangkan untuk sampel dijabarkan pada persamaan 29 sebagai berikut.

$$\pi = \frac{p}{1-p} \qquad (29)$$

Misalkan model regresi logistik pada Persamaan 30 sebagai berikut.

$$\log\left(\frac{p}{1-p}\right) = \propto + \beta x \qquad (30)$$

Jika model dieksponensialkan pada kedua sisi, maka membentuk Persmaan 31 sebagai berikut.

$$\frac{p}{1-p} = \exp(\propto + \beta x) \qquad (31)$$

Misalkan $x = 1$ dan $x = 0$. Ketika $x = 1$, maka :

$$\frac{p}{1-p} = \exp(\propto + \beta) \tag{32}$$

Sedangkan ketika $x = 0$, dimisalkan notasi *p* menjadi p^* untuk membedakan notasi probabilitas:

$$\frac{p^*}{1-p^*} = \exp(\propto) \tag{33}$$

Ketika $x = 1$ maka probabilitasnya adalah *p*. Sedangkan ketika $x = 0$ maka probabilitasnya adalah p^*. Sehingga, berdasarkan Persamaan 32 dan Persamaan 33, dapat dihitung *odds ratio* yaitu sebagai berikut:

$$\frac{\frac{p}{1-p}}{\frac{p^*}{1-p^*}} = \frac{\exp(\propto+\beta)}{\exp(\propto)} = \frac{\exp(\propto).\exp(\beta)}{\exp(\propto)} = \exp(\beta) \tag{34}$$

Sehingga berdasakan perhitungan pada Persamaan (34), nilai *odds* ratio adalah $\exp(\beta)$. Nilai *odds ratio* sama dengan satu artinya *odds* pada grup satu sama dengan *odds* pada grup dua. Nilai *odds ratio* kurang dari satu artinya *odds* pada grup satu lebih kecil dibandingkan odds grup dua. Nilai *odds ratio* lebih dari satu artinya *odds* pada grup satu lebih besar dibandingkan *odds* grup dua.

1.5.11 *Automatic Teller Machine*

Automatic Teller Machine atau Anjungan Tunai Mandiri yang biasa disingkat ATM adalah mesin yang beroperasi secara sistematik menggunakan teknologi komputerisasi dan komunikasi nirkabel sehingga nasabah dapat melakukan transaksi yang pada awalnya hanya dapat dilakukan melalui pelayanan dengan berhadapan langsung oleh *teller* dengan batasan waktu jam kerja sekarang dengan adanya mesin ATM transaksi dapat dilakukan secara mudah dan dapat diakses 24 jam (Syahrizal et al., 2022). Secara umum fungsi ATM yaitu untuk dapat melakukan penarikan tunai namun selain itu ATM memiliki beberapa fungsi lain yaitu :

1. Setoran Tunai
2. Pemindahan dana
3. Pengecekan dana
4. Penbayaran tagihan listrik, telepon, air, internet, dan lainnya.
5. Pembelian pulsa, paket data, token listrik, dan produk lainnya
6. Informasi produk dan layanan bank

Dengan kinerja masin ATM yang berfungsi selama 24 jam tentu perangkat keras maupun perangkat lunak yang bekerja di dalam mesin ATM dapat mengalami berbagai gangguan yang dapat mempengaruhi fungsionalitasnya. Berikut adalah beberapa gangguan yang sering terjadi pada mesin ATM:

1. Gangguan jaringan ATM yang biasanya disebabkan oleh gangguan pada satelit atau masalah teknis lainnya.
2. *Cash handler fatal error,* kesalahan yang biasanya terjadi pada pengaturan *fast cash*, kehabisan uang tunai, gangguan pada memori, gangguan pada sensor.

3. *Bunch note acceptor faulted fatal error,* kegagalan dalam menerima dan memvalidasi uang kertas yang dimasukkan oleh nasabah.
4. *Receipt printer fatal error,* gangguan ini dapat disebabkan karena koneksi antara mesin ATM dan pencetak struk, kehabisan kertas struk, dan kerusakan pada perangkat itu sendiri.

BAB II
METODOLOGI PENELITIAN

2.1 Jenis dan Sumber Data

Data yang digunakan dalam penelitian ini merupakan sebuah data sekunder yang diperoleh dari bank BNI pada bulan juli 2023 berupa laporan monitoring mesin ATM. Data yang diperoleh melalui kantor wilayah Makassar bank BNI dengan pengamatan sebanyak 30 mesin ATM dan jumlah data observasi sebanyak 564 yang terlapir pada lampiran 1. Gangguan yang terjadi pada ATM dikategorikan menjadi 2 yaitu gangguan yang terjadi karena *hardware* yang bermasalah dan gangguan yang terjadi bukan karena pengaruh dari *hardware*.

2.2 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini ada pada tabel 1 sebagai berikut:

Tabel 1 Variabel Penelitian

Variabel	Keterangan	Definisi Operasional
Y	Gangguan ATM	0 : sebagai gangguan yang terjadi bukan karena *Hardware* 1 : sebagai gangguan yang terjadi karena *Hardware*
X_1	Selisih Hari Terjadinya Gangguan	Selisih hari gangguan terjadi pada pengamatan ke-n terhadap gangguan sebelumnya
X_2	Status jaringan	0 : sebagai nilai untuk mesin ATM tidak terhubung dengan jaringan saat gangguan terjadi 1 : sebagai mesin ATM terhubung dengan jaringan saat gangguan terjadi
X_3	Status ATM	0 : sebagai status mesin ATM mati saat gangguan terjadi 1 : sebagai status mesin ATM menyala saat gangguan terjadi

2.3 Metode Analisis

Adapun tahapan analisis yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Menentukan jenis data yang digunakan, dalam penelitian ini digunakan data longitudinal berdistribusi binomial.
2. Melakukan eksplorasi data menggunakan statistika desktiptif.
3. Proses analisis GEE:
   a. Menentukan *link function* untuk regresi logistik berdistribusi binomial dengan Persamaan 2.

$$logit(\mu_i) = \ln \frac{\mu_i}{1-\mu_i}$$

b. Menentukan *working correlation structure* yang terbaik dengan membandingkan nilai QIC dari masing-masing jenis struktur korelasi menggunakan Persamaan 10

$$QIC = -2Q\left(\widehat{\boldsymbol{\beta}}(\boldsymbol{R}(\boldsymbol{\alpha})); I\right) + 2trace\left(\widehat{\boldsymbol{\Omega}}_I \widehat{\mathbf{V}}_r\right)$$

c. Melakukan iterasi pendugaan parameter. berdasarkan iterasi *Newton Raphson* menggunakan Persamaan 9.

$$\hat{\beta}^{t+1} = \hat{\beta}^t + \left[\sum_{i=1}^{N} \left(\frac{\partial \mu_i}{\partial \beta}\right)' \boldsymbol{V_i}^{-1} \frac{\partial \mu_i}{\partial \beta}\right]^{-1} \left[\sum_{i=1}^{N} \left(\frac{\partial \mu_i}{\partial \beta}\right)' \boldsymbol{V_i}^{-1} (y_i - \mu_i)\right]$$

d. Melakukan uji hipotesis signifikan parameter secara pasial (uji *Wald*) menggunakan Persamaan 25.

$$W = \frac{(\hat{\beta}_i)}{[SE(\hat{\beta}_i)]}$$

Jika terdapat parameter tidak signifikan, maka kembali ke tahap 3c dan melakukan uji parameter kembali tanpa mengikutsertakan parameter yang tidak signifikan.

e. Melakukan pemodelan akhir GEE.

4. Proses analisis GLMM:

a. Menentukan *link function* untuk regresi logistik berdistribusi binomial dengan Persamaan 2.

$$logit(\mu_i) = \ln \frac{\mu_i}{1-\mu_i}$$

b. Melakukan pendugaan parameter menggunakan iterasi *Newton-Rapshon* menggunakan Persamaan 24.

$$E\left[\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}\right] = -\boldsymbol{X^T}(\boldsymbol{Z}D\boldsymbol{Z^T} + R)^{-\mathbf{1}}\boldsymbol{X}$$

c. Melakukan uji hipotesis signifikan parameter secara parsial (uji Wald) menggunakan Persamaan 25.

$$W = \frac{(\hat{\beta}_i)}{[SE(\hat{\beta}_i)]}$$

Jika terdapat parameter tidak signifikan, maka kembali ke tahap 4b dan melakukan uji parameter kembali tanpa mengikutsertakan parameter yang tidak signifikan.

d. Melakukan uji kesesuaian model menggunakan uji *Pearson* menggunakan Persamaan 26.

$$\chi_p^2(y_i, \hat{\pi}_i) = n_i \sum_{j=1}^{k} \frac{(y_{ij} - \hat{\pi}_{ij})^2}{\hat{\pi}_{ij}}$$

Jika model telah sesuai maka dilanjutkan ke tahap 4e, namun jika model belum sesuai maka kembali ke tahap 4b untuk penyesuaian model dengan menambahkan efek acak atau pertimbangan dalam transformasi pada variabel respon atau variabel prediktor.

e. Melakukan interpretasi model akhir GLMM.

5. Memilih model terbaik dari Metode GEE dan GLMM berdasarkan nilai AIC menggunakan Persamaan 27.

$$AIC(\boldsymbol{y}) = -2logf\left(\boldsymbol{y}\middle|\hat{\boldsymbol{v}}(\boldsymbol{y})\right) + 2\dim(\boldsymbol{v})$$

6. Melakukan interpretasi hasil olah data.

7. Menarik kesimpulan dan saran.