UIN IMAM BONJOL PADANG
MAP Journal
Mathematics & Applications
Web : ejournal.uinib.ac.id/jurnal/index.php/MAp
Email : map_journal@uinib.ac.id

ANALISIS DATA LONGITUDINAL DENGAN RESPON BINER MENGGUNAKAN *GENERALIZED ESTIMATING EQUATION* (*GEE*)

ANALYSIS OF LONGITUDINAL DATA WITH BINARY RESPONSE USING GENERALIZED ESTIMATING EQUATION (GEE)

Syarto Musthofa[1§], Lilis Harianti Hasibuan[2], Darvi Mailisa Putri[3], Miftahul Jannah[4], Ilham Dangu Rianjaya[5]

[1]UIN Imam Bonjol Padang [Email: syartom@uinib.ac.id]
[2] UIN Imam Bonjol Padang [Email: lilisharianti@uinib.ac.id]
[3] UIN Imam Bonjol Padang [Email: darvimailisa@uinib.ac.id]
[4] UIN Imam Bonjol Padang [Email: miftahuljannah@uinib.ac.id]
[5] UIN Imam Bonjol Padang [Email: ilham.rianjaya@uinib.ac.id]
[§]*Corresponding Author*

Received 28th Dec 2023; Accepted 29th Dec 2023; Published 01th Dec 2023;

Abstrak

Data longitudinal adalah data yang diperoleh dari hasil pengukuran sejumlah individu secara berulang dalam beberapa waktu yang berbeda. Data longitudinal menunjukkan bagaimana perubahan nilai pada individu yang diamati relatif terhadap waktu dan beberapa variabel yang menjadi perhatian. Variabel respon pada data longitudinal dimungkinkan dalam bentuk biner. Data dengan respon biner pada dasarnya bisa dianalisis dengan regresi logistik. Namun, regresi logistik tidak mempertimbangkan korelasi antar pengamatan yang mungkin terjadi pada satu individu. Dalam penelitian ini *Generalized Estimating Equation* (GEE) digunakan dalam melakukan estimasi parameter pada model data longitudinal. GEE memberi ruang pembahasan pada adanya kemungkinan korelasi antar pengamatan pada satu individu untuk data longitudinal yang memiliki variabel respon biner. Studi kasus dalam penelitian ini menganalisis probabilitas terjadinya kondisi suhu di atas normal berdasarkan lamanya penyinaran matahari (X_1). Estimasi parameter yang dilakukan menghasilkan model $\pi_i = \frac{1}{1+e^{-(-2.427+0.553x_{1i})}}$ dengan struktur korelasi *exchangeable* ($\alpha = 0{,}607$) yang menunjukkan bahwa semakin lama penyinaran matahari akan semakin memperbesar probabilitas kondisi suhu di atas normal.

Kata Kunci: Data Longitudinal, Regresi Logistik, Generalized Estimating Equation (GEE)

Abstract

Longitudinal data is data obtained from repeated measurements of a number of subjectss at different times. Longitudinal data show how values change in observed subjects relative to time and some independent variables of interest. Response variables in longitudinal data may be binary. Data with binary responses can basically be analyzed using logistic regression. However, logistic regression does not consider correlations between observations that may occur in one subject. In this research, Generalized Estimating Equation (GEE) is used to estimate longitudinal model parameters. GEE provides space for discussion regarding the possibility of correlation between observations of one subject on longitudinal data that has a binary response variable. The case study in this research analyzes the probability of above normal temperature conditions based on the duration of sunlight exposure (X_1). The

parameter estimation carried out produces a model $\pi_i = \frac{1}{1+e^{-(-2.427+0.553x_{1i})}}$ *with an exchangeable correlation structure (α=0.607) which shows that the longer the sunlight exposure, the greater the possibility of above normal temperature condition.*
***Keywords**: Longitudinal Data, Logistic Regression, Generalized Estimating Equation (GEE)*

1. Pendahuluan

Data longitudinal adalah data yang individu dalam sampelnya diamati dalam periode waktu tertentu lebih dari satu kali dan dilakukan pengukuran berulang pada individu tersebut. Inferensi pada model data longitudinal didasarkan pada data individu, dengan asumsi masing-masing individu independen, namun dengan memperhatikan bahwa observasi berulang untuk tiap-tiap individu tidak independen.

Informasi yang diambil dari tiap objek penelitian dalam penelitian longitudinal biasanya lebih dari satu variabel, yang dikategorikan sebagai variabel respon (dependen) dan variabel penjelas (variabel independen). Salah satu tujuan penggunaan data longitudinal dalam penelitian adalah untuk mengetahui apakah ada pengaruh variabel penjelas terhadap variabel respon, termasuk meneliti pengaruh variabel penjelas pada besarnya perubahan pada variabel respon.

Model statistik untuk menganalisis data longitudinal akan lebih ringkas dan jelas apabila dideskripsikan dan direpresentasikan dalam bentuk formulasi matematis. Misal $i = 1,2,\dots,m$ adalah indeks yang menyatakan banyaknya objek penelitian/individu yang diamati, dimana masing-masing memiliki pengamatan berulang $j = 1,2,\dots,n_i$. Banyaknya pengamatan berulang untuk setiap individu tidak selalu sama sehingga secara total ada $N = \sum_{i=1}^{m} n_i$ observasi. Waktu observasi aktual, yaitu observasi saat pengamatan, dinotasikan dengan t_{ij}. Selanjutnya, variabel respon dalam data longitudinal dinyatakan sebagai Y_{ij} dengan nilai observasinya adalah y_{ij}. Variabel tersebut dapat dinyatakan dalam bentuk matriks sebagai berikut,

$$Y_i = \begin{bmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{in_i} \end{bmatrix} \quad (1)$$

atau $Y_i = [Y_{i1} \quad Y_{i2} \quad \cdots \quad Y_{in_i}]^T$ dengan nilai observasi $y_i = [y_{i1} \quad y_{i2} \quad \cdots \quad y_{in_i}]^T$. Apabila dinotasikan sebagai satu vektor observasi untuk seluruh individu, maka faktor Y dapat pula ditulis sebagai berikut

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{bmatrix} \quad (2)$$

atau $Y_i = [Y_1 \quad Y_2 \quad \cdots \quad Y_m]^T$ dengan nilai observasi $y_i = [y_1 \quad y_2 \quad \cdots \quad y_m]^T$. Sedangkan variabel penjelas dapat dinyatakan sebagai berikut

$$X_i = \begin{bmatrix} x_{i11} & \cdots & x_{i1p} \\ \vdots & \ddots & \vdots \\ x_{in_i1} & \cdots & x_{in_ip} \end{bmatrix} \quad (3)$$

yakni sebuah matriks berukuran $n_i \times p$ dengan p adalah banyaknya variabel penjelas. Variabel penjelas dalam analisis longitudinal dapat diamati sekali saja dan nilainya sama sampai akhir studi, yang sering dinamakan sebagai kovariat awal (*baseline covariate*). Variabel penjelas dapat pula diamati lebih dari satu kali selama waktu studi

berjalan, yang sering dinamakan kovariat bergantung waktu (*time varying covariate*) [1].

Mean atau nilai harapan dari variabel respon adalah $E(Y_{ij}) = \mu_{ij}$. Jika variabel respon ditulis sebagai Y_i , maka nilai harapannya adalah $E(Y_i) = \mu_i$. Untuk individu ke-i, variansi dari Y_i berupa matriks kovariansi berukuran $n_i \times n_i$, yaitu,

$$Var(Y_i) = \begin{pmatrix} v_{i11} & \cdots & x_{i1n_i} \\ \vdots & \ddots & \vdots \\ v_{in_i1} & \cdots & v_{in_in_i} \end{pmatrix} \quad (4)$$

dengan $V_{ijk} = Cov(Y_{ij}, Y_{ik})$. Dalam banyak aplikasi, pengamatan berulang untuk setiap individu biasanya dibuat sama, yaitu $n_i = n,$ untuk $i = 1,2,\ldots,m$ sehingga semua persamaan dan notasi di atas menjadi lebih sederhana [2].

2. Landasan Teori

2.1 Regresi Linear

Regresi linear merupakan suatu teknik analisis yang digunakan untuk melakukan prediksi nilai suatu variabel dependen (Y) berdasarkan hubungannya dengan variabel independen (X). Regresi linear yang melibatkan sebuah variabel independen disebut sebagai regresi linear sederhana. Model regresi linear sederhana dapat dilihat sebagai berikut

$$y = \beta_0 + \beta_1 x.$$

Sebagaimana disebutkan dalam [3] estimator untuk β_0 dan β_1 diperoleh dengan meminimumkan jumlah kuadrat error sehingga diperoleh sebagai berikut

$$\hat{\beta}_1 = \frac{n\sum_{i=1}^n x_i y_i - (\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}$$

$$\hat{\beta}_0 = \frac{\sum_{i=1}^n y_i - \hat{\beta}_1 \sum_{i=1}^n x_i}{n}.$$

Apabila variabel independen yang dilibatkan sebanyak p, misal $X_1, X_2, \cdots, X_p$, maka model tersebut akan menjadi regresi linear berganda dengan bentuk

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$$

Dalam [4] disebutkan bahwa dengan meminimumkan jumlah kuadrat error, estimator untuk $\beta = (\beta_0, \beta_1, \cdots, \beta_p)$ diperoleh sebagai berikut

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

2.2 Regresi Logistik

Banyak kasus dalam analisis regresi dimana variabel dependennya bersifat kualitatif. Variabel dependen ini bisa mempunyai dua kelas atau kategori (biner) dan lebih dari dua kelas (multinomial). Salah satu regresi logistik yang paling sederhana digunakan adalah regresi logistik untuk respon biner. Regresi logistik untuk respon biner merupakan suatu metode analisis data yang digunakan untuk mencari hubungan antara variabel dependen (Y) yang bersifat biner atau dikotomus dengan variabel independen (X) yang bersifat polikotomus [2].

Outcome dari variabel dependen Y terdiri dari dua kategori yaitu "terjadi" dan "tidak terjadi" yang dinotasikan dengan $Y = 1$ (terjadi) dan $Y = 0$ (tidak terjadi). Dalam keadaan demikian, variabel Y mengikuti distribusi Bernoulli untuk setiap observasi tunggal. Misalkan Y_i adalah variabel random Bernoulli untuk individu i, distribusi probabilitas Y_i adalah

$$P(Y_i = y) = \pi_i^{y_i}(1-\pi)^{1-y_i}, \quad y_i = 0,1$$

Setiap individu i mempunyai karakteristik berupa variabel x_i yang mempengaruhi π_i dalam bentuk:

$$\pi_i = \frac{1}{1+\exp\left(-(\beta_0+\beta_1 x_i)\right)}$$

Fungsi seperti π_i dalam persamaan disebut fungsi logistik. Untuk variabel independen atau faktor yang lebih dari satu, fungsi untuk π_i dapat diperluas menjadi:

$$\pi_i = \frac{1}{1+e^{-Z}}, \quad \text{atau} \quad \pi_i = \frac{e^Z}{1+e^Z} \tag{5}$$

dengan $Z = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi}$ adalah fungsi linear dari sebanyak p variabel independen . Persamaan π_i tersebut dapat dituliskan sebagai kombinasi linear dari variabel independen seperti halnya pada model linear sebagai berikut:

$$\log\frac{\pi_i}{1-\pi_i} = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi}$$

atau

$$\text{logit}(\pi_i) = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi} \tag{6}$$

dengan $x_{1i}, x_{2i}, \ldots, x_{pi}$ adalah variabel independen; dan $\beta_0, \beta_1, \ldots, \beta_p$ adalah parameter model [5].

Estimasi untuk $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)$ dapat diperoleh dengan MLE untuk fungsi *likelihood* berikut ini [6]

$$L(\beta) = \prod_{i=1}^{n} P(Y_i = y_i)$$

$$= \prod_{i=1}^{n} \frac{\left[\exp\left(\beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi}\right)\right]^{y_i}}{1+\exp\left(\beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi}\right)}$$

2.3 Generalized Estimating Equation

Model *Generalized Estimating Equation* (GEE) merupakan perluasan dari model linear biasa dalam hal spesifikasi korelasi antara dua respon yang berbeda, yaitu y_{ij} dan y_{ik}. GEE memiliki spesifikasi linear prediktor sebagai $\eta_{ij} = x_i\beta$ dengan fungsi penghubungnya adalah $g(\mu_{ij}) = \eta_{ij}$.

Variansi dari respon y adalah $Var(y_{ij}) = \phi v(\mu_{ij})$ dengan $v(\mu_{ij})$ adalah fungsi variansi dan ϕ merupakan parameter skala yang diketahui atau diestimasi. Di dalam GEE terdapat spesifikasi korelasi antara dua respon yang berbeda, atau sering disebut sebagai *working correlation matrix*, yang dinotasikan sebagai R_i, yaitu matriks berukuran $n_i \times n_i$ yang bergantung pada suatu parameter α sehingga matriks korelasi tersebut sering ditulis sebagai $R_i(\alpha)$ [2]. Diantara bentuk struktur korelasi yang ada yaitu *independence*, *exchangeable*, *unstructured*, dan *autoregressive* [7].

Independence

Struktur korelasi ini mengasumsikan bahwa seluruh pengamatan adalah saling independen, termasuk antar pengamatan berulang pada individu yang sama. Artinya tidak ada parameter struktur korelasi yang perlu diestimasi. Bentuknya adalah sebagai berikut

$$R_i(\alpha)_{ind} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix},$$

dapat juga dinyatakan sebagai

$$Corr(Y_{ij}, Y_{ik}) = \begin{cases} 1 & ; j = k \\ 0 & ; j \neq k \end{cases} \tag{7}$$

Exchangeable

Struktur korelasi ini mengasumsikan bahwa nilai korelasi antar pengamatan pada individu yang sama adalah sama, yaitu α. Artinya ada satu parameter yang akan diestimasi. Bentuknya adalah sebagai berikut

$$R_i(\alpha)_{exc} = \begin{bmatrix} 1 & \alpha & \cdots & \alpha \\ \alpha & 1 & \cdots & \alpha \\ \vdots & \vdots & \ddots & \vdots \\ \alpha & \alpha & \cdots & 1 \end{bmatrix},$$

dapat juga dinyatakan sebagai

$$Corr(Y_{ij}, Y_{ik}) = \begin{cases} 1 & ; j = k \\ \alpha & ; j \neq k \end{cases} \quad (8)$$

Unstructured

Pada jenis korelasi ini, nilai korelasi antar pengamatan cenderung tidak mengikuti pola tertentu. Oleh karena itu bentuk korelasi tersebut dinamakan *unstructured*, yaitu tidak terstruktur. Dengan demikian jika maksimum pengamatan berulang pada semua individu adalah *M*, maka banyak parameter yang diestimasi adalah $\frac{M(M-1)}{2}$. Bentuknya adalah sebagai berikut:

$$R_i(\alpha)_{uns} = \begin{bmatrix} 1 & \alpha_{1,2} & \cdots & \alpha_{1,M} \\ \alpha_{1,2} & 1 & \cdots & \alpha_{2,M} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{1,M} & \alpha_{2,M} & \cdots & 1 \end{bmatrix},$$

dapat juga dinyatakan sebagai

$$Corr(Y_{ij}, Y_{ik}) = \begin{cases} 1 & ; j = k \\ \alpha_{jk} & ; j \neq k \end{cases} \quad (9)$$

Autoregressive

Pada struktur korelasi *Autoregressive*, diasumsikan bahwa nilai korelasi antar pengamatan akan semakin meluruh sesuai jarak pengamatannya, dengan kata lain memiliki nilai yang semakin kecil jika jarak antar pengamatan semakin jauh. Oleh karena itu parameter yang diestimasi ada satu dengan bentuk strukturnya sebagai berikut

$$R_i(\alpha)_{aut} = \begin{bmatrix} 1 & \alpha & \cdots & \alpha^{M-1} \\ \alpha & 1 & \cdots & \alpha^{M-2} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha^{M-1} & \alpha^{M-2} & \cdots & 1 \end{bmatrix},$$

dapat juga dinyatakan sebagai

$$Corr(Y_{ij}, Y_{i,j+t}) = \alpha^t; \quad \text{untuk } t = 0,1,\cdots,M-1 \quad (10)$$

3. Hasil Dan Pembahasan

3.1 Estimasi Parameter

Untuk melakukan estimasi parameter terlebih dahulu dimisalkan $\mathbf{A}_i$ matriks diagonal berukuran $n \times n$ dengan $V(\mu_{ij})$ adalah elemen diagonalnya. Diberikan pula $R_i(\alpha)$ sebagai matriks korelasi berukuran $n \times n$, untuk *n* data berulang untuk satu individu *i*. Matriks varians-kovarians untuk y_i adalah:

$$V(\boldsymbol{\alpha}) = \phi \boldsymbol{A}_i^{1/2} \boldsymbol{R}_i(\boldsymbol{\alpha}) \boldsymbol{A}_i^{1/2}$$

Untuk kasus respon Gaussian dengan variansi homogen diperoleh:

$$V(\boldsymbol{\alpha}) = \phi \boldsymbol{R}_i(\boldsymbol{\alpha})$$

Estimasi GEE untuk β adalah solusi dari:

$$\sum_{i=1}^{m} \boldsymbol{D}_i^T [V(\widehat{\boldsymbol{\alpha}})]^{-1} (\boldsymbol{y}_i - \boldsymbol{\mu}_i) = 0$$

dengan $\widehat{\boldsymbol{\alpha}}$ merupakan estimator konsisten α dan $\boldsymbol{D}_i = \partial \boldsymbol{\mu}_i / \partial \boldsymbol{\beta}$. Matriks $\boldsymbol{D}_i$ disebut sebagai matriks derivatif, yaitu matriks yang berisi turunan dari $\boldsymbol{\mu_i}$ terhadap komponen $\boldsymbol{\beta}$. Matriks ini mentransformasikan unit asal Y_i (dan μ_i) menjadi unit pada skala $g(\mu_{ij})$. Skala pada unit fungsi penghubung $g(\mu_{ij})$ ini dapat digunakan untuk memberi interpretasi pada nilai $\boldsymbol{\beta}$. Misalnya bila Y_{ij} adalah variabel random biner, maka $g(\mu_{ij})$ berbentuk fungsi logit, bukan probabilitas.

Matriks derivatif $\boldsymbol{D}_i$ berukuran $n_i \times p$ dapat dijabarkan sebagai berikut:

$$\begin{bmatrix} \frac{\partial \mu_{i1}}{\partial \beta_1} & \frac{\partial \mu_{i1}}{\partial \beta_2} & \cdots & \frac{\partial \mu_{i1}}{\partial \beta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \mu_{in_i}}{\partial \beta_1} & \frac{\partial \mu_{in_i}}{\partial \beta_2} & \cdots & \frac{\partial \mu_{in_i}}{\partial \beta_p} \end{bmatrix}$$

dan merupakan fungsi dari β saja karena μ_{ij} merupakan fungsi yang bergantung pada β [8].

Dalam kasus Gaussian, nilai $\boldsymbol{\mu_i}, \boldsymbol{D_i}$, dan $V(\hat{\boldsymbol{\alpha}})$ berturut-turut adalah $\boldsymbol{X_i\beta}, \boldsymbol{X_i}$, dan $R_i(\hat{\boldsymbol{\alpha}})$, sehingga estimasi GEE untuk β menjadi solusi dari

$$\sum_{i=1}^{m} \boldsymbol{X}_i^T [\boldsymbol{R}_i(\hat{\boldsymbol{\alpha}})]^{-1} (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta}_i) = 0$$

yang menghasilkan estimator:

$$\hat{\boldsymbol{\beta}} = \left[\sum_{i=1}^{m} \boldsymbol{X}_i^T [\boldsymbol{R}_i(\hat{\boldsymbol{\alpha}})]^{-1} \boldsymbol{X}_i\right]^{-1} \left[\sum_{i=1}^{m} \boldsymbol{X}_i^T [\boldsymbol{R}_i(\hat{\boldsymbol{\alpha}})]^{-1} \boldsymbol{y}_i\right] \quad (11)$$

Setelah β diperoleh kemudian dihitung estimasi α dan ϕ. Terlebih dahulu dihitung residual Pearson terstandarisasi yakni

$$r_{ij} = \frac{(y_{ij} - \mu_{ij})}{\sqrt{[V(\hat{\boldsymbol{\alpha}})]_{jj}}},$$

yang kemudian r_{ij} digunakan untuk mengestimasi ϕ dan α. Langkah-langkah tersebut diulang sampai diperoleh nilai estimasi yang konvergen.

Parameter ϕ dapat diestimasi dengan

$$\hat{\phi} = \frac{\sum_{i=1}^{m} \sum_{i=1}^{m} r_{ij}^2}{\sum_{i=1}^{m} n_i}. \quad (12)$$

Untuk korelasi *unstructured*, α_{jk} diestimasi dengan

$$\hat{\alpha}_{jk} = \frac{1}{(m-p)\hat{\phi}} \sum_{i=1}^{m} r_{ij} r_{ik}. \quad (13)$$

Untuk korelasi *autoregressive*, α diestimasi dengan

$$\hat{\alpha} = \frac{1}{(K-p)\hat{\phi}} \sum_{i=1}^{m} \sum_{j \le n_i - 1} r_{ij} r_{i,j+1} \quad (14)$$

dengan $K = \sum_{i=1}^{m} (n_i - 1)$.

Untuk korelasi *exchangeable*, α diestimasi dengan

$$\hat{\alpha} = \frac{1}{(N-p)\hat{\phi}} \sum_{i=1}^{m} \sum_{j \ne k} r_{ij} r_{ik} \quad (15)$$

dengan $N = \sum_{i=1}^{m} n_i (n_i - 1)$ [7].

Setelah hasil estimasi konvergen, dapat diperoleh dua versi *standard error* untuk hasil estimasi β, yaitu *naïve standard error* (*model based*) yang berbentuk:

$$V(\beta) = \left[\sum_{i=1}^{m} \boldsymbol{D}_i^T \hat{\boldsymbol{V}}_i^{-1} \boldsymbol{D}_i\right]^{-1}$$

dan *robust standard error* (*empirical*) yaitu:

$$V(\hat{\beta}) = \boldsymbol{M}_0^{-1} \boldsymbol{M}_1 \boldsymbol{M}_0^{-1}$$

dengan

$$\boldsymbol{M}_0 = \sum_{\boldsymbol{i=1}}^{\boldsymbol{m}} \boldsymbol{D}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{D}_i, \quad \text{dan}$$

$$\boldsymbol{M}_1 = \sum_{\boldsymbol{i=1}}^{\boldsymbol{m}} \boldsymbol{D}_i^T \boldsymbol{V}_i^{-1} (\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i)(\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i)^T \boldsymbol{V}_i^{-1} \boldsymbol{D}_i$$

dengan $\hat{\boldsymbol{V}}_{\boldsymbol{i}}$ adalah $V_i(\hat{\boldsymbol{\alpha}})$. Dapat dilihat apabila $\hat{\boldsymbol{V}}_{\boldsymbol{i}} = (\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i)(\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i)^T$, kedua versi *standard error* tersebut akan sama [9].

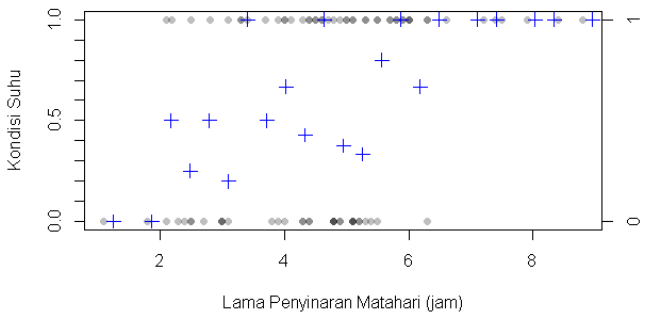
Kasus yang dianalisis pada penelitian ini mengacu pada Data Pengamatan Unsur Iklim dari 7 Stasiun Badan Meteorologi Klimatologi dan Geofisika (BMKG) Provinsi Sumatera Utara pada Tahun 2022 yang diakses melalui website https://sumut.bps.go.id. Tujuh stasiun tersebut adalah Stasiun Klimatologi Deli Serdang, Stasiun Meteorologi Maritim Belawan, Stasiun Meteorologi Kualanamu, Balai Besar Meteorologi Klimatologi dan Geofisika Wilayah I, Stasiun Meteorologi Aek Godang, Stasiun Meteorologi FL Tobing, dan Stasiun Meteorologi

Binaka. Data yang dimiliki merupakan data bulanan selama Tahun 2022 pada masing-masing stasiun sehingga terdapat 84 baris data. Variabel yang menjadi perhatian adalah rata-rata suhu, lama penyinaran matahari dan kecepatan angin. Model yang akan dibentuk adalah kajian perubahan kondisi suhu yang dapat terjadi akibat lamanya penyinaran matahari dan kecepatan angin [10].

Berdasarkan data dari 91 stasiun BMKG, suhu udara normal periode 1991-2020 di Indonesia adalah 26,8^{o}C [11]. Sehingga data suhu pada pengamatan unsur iklim di 7 stasiun BMKG Sumatera Utara di atas dapat diklasifikasikan ke dalam dua kategori yaitu suhu normal ke bawah dan suhu di atas normal.

Analisis data dilakukan dengan melihat bagaimana pengaruh variabel lama penyinaran matahari dalam satuan jam (X_1) dan kecepatan angin dalam satuan knot (X_2) terhadap probabilitas kondisi suhu (π_Y), dengan Y merupakan variabel respon biner (bernilai 0 atau 1). Dalam hal ini $\pi_{Y=0}$ adalah probabilitas kondisi suhu normal ke bawah dan $\pi_{Y=1}$ adalah probabilitas kondisi suhu di atas normal.

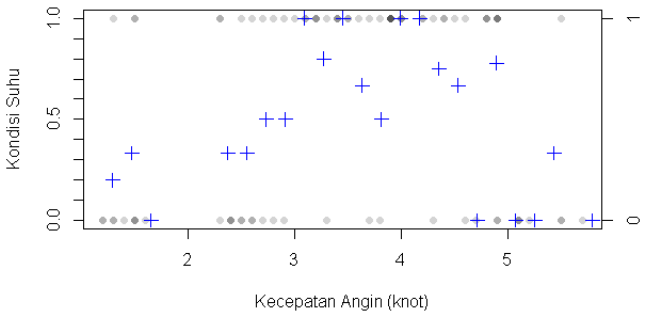
Terlebih dahulu dilihat bagaimana sebaran kondisi suhu berdasarkan lamanya penyinaran matahari sebagai berikut



Gambar 1. Sebaran kondisi suhu berdasarkan lama penyinaran matahari

Pada Gambar 1 terlihat titik-titik gelap pada bagian bawah dan atas yang merepresentasikan kategori suhu berdasarkan lamanya penyinaran matahari, posisinya berada pada nilai 0 atau 1. Kemudian terdapat juga simbol tambah (+) yang merupakan representasi banyaknya kondisi 1 pada nilai penyinaran tertentu. Sebagai contoh, untuk lama penyinaran matahari 4 jam terdapat satu titik berada pada nilai 0 dan dua titik berada pada nilai 1. Karena $2/3$ kondisi suhu pada nilai penyinaran 4 jam berada pada kondisi 1, berarti simbol tambah (+) yang sejajar nilai 4 jam berada pada posisi $66{,}6\overline{6}$. Dari Gambar 1 tersebut dapat dilihat adanya pola bahwa penyinaran matahari yang semakin lama akan memberikan kecenderungan kondisi suhu bernilai 1, demikian juga sebaliknya. Bahkan pola tersebut semakin jelas terlihat pada saat lama penyinaran matahari 7 jam lebih, semua kondisi suhunya berada pada kondisi 1.

Kemudian dilihat sebaran kondisi suhu berdasarkan kecepatan angin sebagai berikut

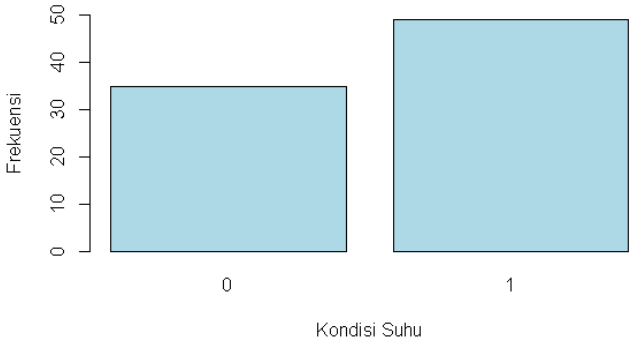


Gambar 2. Sebaran kondisi suhu berdasarkan kecepatan angin

Dari Gambar 2 dapat dilihat bahwa kondisi suhu punya kecenderungan bernilai 1 apabila kecepatan angin naik. Namun pola tersebut tidak

berlanjut seterusnya karena pada saat kecepatan angin di atas 5 knot yang terjadi justru kecenderungan kondisi suhu berada pada nilai 0.

Sebanyak 49 dari 84 baris data suhu tersebut berada pada kondisi 1. Sehingga bisa dikatakan bahwa Tahun 2022 merupakan tahun yang cukup panas sebagaimana disebutkan oleh BMKG [11]. Frekuensi kondisi suhu 0 dan 1 dapat dilihat sebagai berikut



Gambar 3. Perbandingan frekuensi kondisi suhu normal ke bawah (0) dan di atas normal (1)

Untuk melihat bagaimana kedua variabel penjelas, lama penyinaran matahari (X_1) dan kecepatan angin (X_2), mempengaruhi probabilitas variabel kondisi suhu (π_Y) dilakukan estimasi parameter menggunakan Persamaan (11) untuk β dan Persamaan (13), (14), dan (15) untuk α. Diperoleh hasilnya sebagai berikut

Tabel 1. Hasil Estimasi $\beta_0, \beta_1, \beta_2$ dengan *independence working correlation matrix*

	Estimasi	Robust S.E.	Robust z	p value
β_0	-0,066	0,403	-0,163	0,871
β_1	0,106	0,043	2,477	0,013
β_2	0,045	0,085	0,533	0,594

Tabel 2. Hasil Estimasi $\beta_0, \beta_1, \beta_2$ dengan *exchangeable working correlation matrix*

	Estimasi	Robust S.E.	Robust z	p value
β_0	-2,801	1,370	-2,044	0,041
β_1	0,555	0,199	2,791	0,005
β_2	0,111	0,129	0,862	0,389

Tabel 3. Hasil Estimasi $\beta_0, \beta_1, \beta_2$ dengan *autoregressive working correlation matrix*

	Estimasi	Robust S.E.	Robust z	p value
β_0	-2,410	1,793	-1,344	0,179
β_1	0,576	0,278	2,070	0,038
β_2	-0,021	0,279	-0,074	0,941

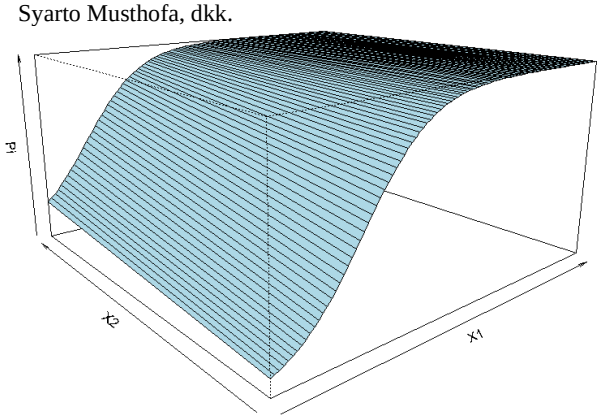
Dari estimasi yang dilakukan, *unstructured working correlation matrix* memberikan hasil yang divergen sehingga tidak memiliki hasil estimasi. Sedangkan hasil estimasi parameter untuk struktur korelasi *independence*, *exchangeable*, dan *autoregressive* dapat dilihat berturut-turut pada Tabel 1, Tabel 2, dan Tabel 3. Dari Tabel 1 diketahui hanya β_1 yang signifikan berdasarkan *p-value* (satu-satunya yang bernilai $< 0{,}05$), dengan struktur korelasinya sama dengan Persamaan (7). Dari Tabel 2 terdapat β_0 dan β_1 yang signifikan dengan struktur korelasinya sesuai Persamaan (8) dengan $\alpha = 0{,}599$. Kemudian dari Tabel 3 hanya β_1 yang signifikan dengan struktur korelasinya sesuai Persamaan (10) dengan $\alpha = 0{,}733$ untuk $t = 1,2,\cdots,11$. Dari ketiganya yang lebih baik adalah struktur korelasi *exchangeable* karena ada dua parameter yang signifikan. Bila dimodelkan dengan fungsi logit sesuai Persamaan (6) akan menjadi

$$logit(\pi_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$$
$$= -2{,}801 + 0{,}555 x_{1i} + 0{,}111 x_{2i} \quad (16)$$

dengan fungsi probabilitas untuk kondisi suhu adalah fungsi π_i sebagaimana Persamaan (5)

$$\pi_i = \frac{1}{1 + e^{-(-2{,}801 + 0{,}555 x_{1i} + 0{,}111 x_{2i})}}. \quad (17)$$

Grafik Persamaan (17) dapat dilihat sebagai berikut



Gambar 4. Probabilitas kondisi suhu berdasarkan lama penyinaran matahari (X_1) dan kecepatan angin (X_2)

Dari Persamaan (16) dan (17) serta Gambar 4 dapat dilihat bahwa penyinaran matahari dan kecepatan angin berkorelasi positif terhadap probabilitas kondisi suhu di atas normal. Namun, sebagaimana hasil yang diperoleh pada Tabel 2 kecepatan angin sebenarnya tidak signifikan berpengaruh dalam model tersebut. Hal itu terjadi karena perbandingan hasil estimasi parameter β_2 dengan simpangan bakunya (*robust standard error*) begitu rendah. Sehingga patut dipertimbangkan untuk mengeluarkan variabel tersebut dari analisis data dan melihat kembali hasilnya.

Berikutnya dilakukan estimasi parameter untuk model yang hanya memperhatikan pengaruh variabel lama penyinaran matahari (X_1) terhadap probabilitas kondisi suhu. Hasilnya diperoleh sebagai berikut

Tabel 4. Hasil Estimasi β_0, β_1 dengan *independence working correlation matrix*

	Estimate	Robust S.E.	Robust z	p-value
β_0	0.057	0.312	0.181	0.856
β_1	0.113	0.045	2.536	0.011

Tabel 5. Hasil Estimasi β_0, β_1 dengan *exchangeable working correlation matrix*

	Estimate	Robust S.E.	Robust z	p-value
β_0	-2.427	1.185	-2.049	0.040

β_1	0.553	0.193	2.863	0.004

Tabel 6. Hasil Estimasi β_0, β_1 dengan *autoregressive working correlation matrix*

	Estimate	Robust S.E.	Robust z	p-value
β_0	-2.462	1.549	-1.589	0.112
β_1	0.572	0.276	2.073	0.038

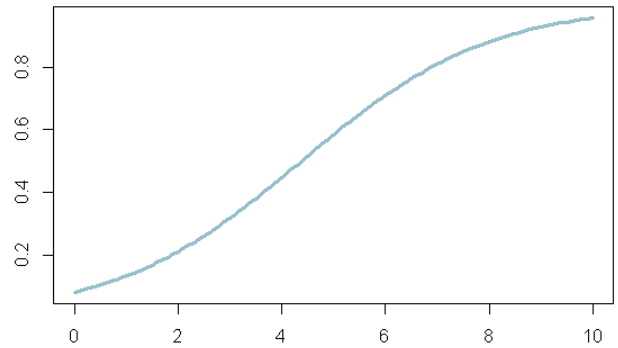
Sama seperti hasil estimasi untuk dua variabel independen yang dilakukan sebelumnya, estimasi untuk satu variabel independen pada *ustructured working correlation matrix* memberikan hasil yang divergen sehingga tidak memiliki nilai estimasi. Dari Tabel 4, Tabel 5, dan Tabel 6 diketahui bahwa struktur korelasi *exchangeable* signifikan pada kedua parameter yang diestimasi dengan nilai $\alpha = 0{,}607$. Sehingga model yang dibentuk adalah model dengan struktur korelasi *exchangeable* yang dinyatakan sebagai fungsi logit sebagai berikut

$$logit(\pi_i) = \beta_0 + \beta_1 x_{1i} = -2.427 + 0.553 x_{1i} \quad (18)$$

dengan fungsi probabilitas untuk kondisi suhu adalah fungsi π_i sebagai berikut

$$\pi_i = \frac{1}{1 + e^{-(-2.427 + 0.553 x_{1i})}}. \quad (19)$$

Grafik Persamaan (19) dapat dilihat berikut ini



Gambar 5. Probabilitas kondisi suhu berdasarkan lama penyinaran matahari (X_1)

Dari Persamaan (18) dan (19) dapat dilihat bahwa variabel lama penyinaran matahari

berkorelasi positif terhadap kondisi suhu di atas normal. Artinya semakin lama waktu penyinaran matahari akan memperbesar peluang kondisi suhu di atas normal. Visualisasi peluang kondisi tersebut dapat dilihat pada Gambar 5. Untuk lama penyinaran nol jam, berdasarkan Persamaan (19) probabilitas kondisi suhu di atas normal adalah 0.08112672 (sekitar 8%). Misalnya, apabila di penyinaran matahari berlangsung selama 7 jam maka probabilitas kondisi suhu di atas normal adalah 0.8095652 (sekitar 81%). Sedangkan jika penyinaran matahari lebih lama lagi maka peluang kondisi suhu di atas normal akan semakin mendekati 1.

4. Kesimpulan Dan Saran

Dari pembahasan dan hasil analis yang dilakukan dapat disimpulkan bahwa data longitudinal dengan respon biner mengasumsikan adanya kemungkinan korelasi antar pengamatan pada satu individu sehingga kurang sesuai apabila dianalisis dengan regresi logistik saja. Sedangkan model GEE memberi ruang pembahasan pada adanya kemungkinan korelasi antar pengamatan di satu individu. Dari analisis data diperoleh bahwa lama penyinaran matahari berkorelasi positif secara signifikan terhadap kondisi suhu di atas normal. Sedangkan kecepatan angin berkorelasi positif namun tidak signifikan terhadap kondisi suhu di atas normal. Sehingga model yang digunakan adalah model dengan variabel independen lama penyinaran matahari terhadap kondisi suhu dengan struktur korelasi *exchangeable*.

Dari eksplorasi data yang dilakukan tentang hubungan kecepatan angin terhadap kondisi suhu dapat dilihat bahwa keduanya tidak berkorelasi linear sehingga disarankan untuk menggunakan model lain untuk melihat bagaimana keterlibatan kecepatan angin terhadap perubahan kondisi suhu.

5. Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada semua pihak yang telah berkontribusi dalam pelaksanaan penelitian ini.

Daftar Pustaka

[1] A. Taufiq Hidayat, Z. Mujtahid, N. Elisyah, and H. Qausar, “Analisis Data Longitudinal dalam Mendeteksi Faktor Substansial yang Mempengaruhi Hasil Belajar Matematika Siswa MA Al Hikmah 2 Benda Brebes,” *Anal. Data Longitud.* , pp. 74–78, 2022.

[2] Danardono, *Analisis Data Longitudinal*. Yogyakarta: Gadjah Mada University Press, 2015.

[3] L. H. Hasibuan and S. Musthofa, “Penerapan Metode Regresi Linear Sederhana Untuk Prediksi Harga Beras di Kota Padang,” *JOSTECH J. Sci. Technol.*, vol. 2, no. 1, pp. 85–95, 2022.

[4] D. A. Fitri, “Estimasi Parameter Model Regresi Linier Berganda Dengan Teknik Bootstrap,” *J. Mat. UNAND*, vol. 3, no. 3, p. 41, 2014.

[5] S. Sperandei, “Understanding logistic regression analysis,” *Biochem. Medica*, vol. 24, no. 1, pp. 12–18, 2014.

[6] S. Winarni, “Analisis Data Longitudinal Dalam Desain Faktorial Menggunakan Linear Mixed Model Human Resources Development Program in EMR fields View project,” no. June, 2017.

[7] M. Wang, “Generalized Estimating

Equations in Longitudinal Data Analysis: A Review and Recent Developments," *Adv. Stat.*, vol. 2014, pp. 1–11, 2014.

[8] H. Geys, G. Molenberghs, and L. M. Ryan, "Generalized estimating equations," *Top. Model. Clust. Data*, pp. 77–87, 2002.

[9] I. Owusu-Darko, I. K. Adu, and N. K. Frempong, "Application of generalized estimating equation (GEE) model on students' academic performance," *Appl. Math. Sci.*, vol. 8, no. 65–68, pp. 3359–3374, 2014.

[10] T. B. Sitorus, F. H. Napitupulu, and H. Ambarita, "Korelasi Temperatur Udara dan Intensitas Radiasi Matahari Terhadap Performansi Mesin Pendingin Siklus Adsorpsi Tenaga Matahari," *J. Ilm. Tek. Mesin Cylind.*, vol. 1(1), no. 1, pp. 8–17, 2014.

[11] M. Sudirman, "Anomali Suhu Udara Rata-Rata Tahun 2022," 2023. https://www.bmkg.go.id/iklim/anomali-suhu-udara-tahunan.bmkg?p=anomali-suhu-udara-tahunan&tag=&lang=ID#.