



LAPORAN EAS

ANALISIS DATA KATEGORI

**PEMODELAN TINGKAT KASUS CAMPAK PADA BALITA DI KOTA SURABAYA
DENGAN PENDEKATAN REGRESI PROBIT DAN REGRESI SKEWED PROBIT**

Oleh:

Muhammad Adlansyah Muda (6003202020)

Dosen:

Dr. Purhadi, M.Sc.

Dr. Vita Ratnasari, M.Si.

PROGRAM STUDI MAGISTER DEPARTEMEN STATISTIKA

FAKULTAS SAINS DAN ANALITIKA DATA

INSTITUT TEKNOLOGI SEPULUH NOPEMBER

SURABAYA

2021

ABSTRAK

Kesehatan merupakan masalah serius karena tanpa hal tersebut manusia tidak dapat beraktivitas sehari-hari. Masalah kesehatan seperti penyakit campak hingga saat ini selalu menjadi perhatian dunia terutama Indonesia. Surabaya merupakan salah satu Kota di Indonesia yang memiliki jumlah kasus campak yang cukup besar, sehingga pada penelitian ini akan dilakukan analisis mengenai pemodelan tingkat kasus campak pada balita di Kota Surabaya. Faktor-faktor yang diduga mempengaruhi campak pada balita di penelitian ini ialah jumlah imunisasi campak, jumlah puskesmas, dan jumlah tenaga kesehatan pada tiap kecamatan di Kota Surabaya. Data penelitian yang digunakan imbalanced dimana sebanyak 12 Kecamatan di Kota Surabaya memiliki tingkat kasus campak dengan status darurat dan 19 Kecamatan lainnya dalam status normal. Hasil penelitian menggunakan model regresi Probit Biner menghasilkan akurasi sebesar 80,56% dan sensitivity sebesar 80,95%, sedangkan menggunakan model regresi Skewed Probit Biner memberikan nilai akurasi sebesar 80,56% dan sensitivity sebesar 84,21%. Hasil tersebut menunjukkan bahwa model regresi Skewed Probit Biner lebih baik dalam memprediksi tingkat kasus campak dengan status darurat pada tiap Kecamatan di Kota Surabaya.

Kata Kunci: *Campak, Imunisasi, Kesehatan, Probit, Skewed Probit.*

DAFTAR ISI

LAPORAN EAS	i
ABSTRAK	ii
DAFTAR ISI.....	iii
DAFTAR TABEL	v
DAFTAR GAMBAR.....	vi
DAFTAR LAMPIRAN	vii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan Penelitian.....	2
1.4 Manfaat Penelitian.....	2
BAB II TINJAUAN PUSTAKA	3
2.1 Statistika Deskriptif.....	3
2.2 Uji Multikolinieritas	4
2.3 Regresi Probit Biner	4
2.3.1 Model Regresi Probit Biner	5
2.3.2 Estimasi Parameter	6
2.3.3 Pengujian Estimasi Parameter	7
2.3.4 Uji Kesesuaian Model	8
2.4 Regresi Skewed Probit Biner	8
2.4.1 Model Regresi Skewed Probit Biner	8
2.4.2 Parameter Skewness	9
2.4.3 <i>Penalizing Complexity</i> Prior untuk Parameter Skewness.....	9
2.5 Ketepatan Klasifikasi Model.....	11
2.6 Campak.....	12
BAB III METODOLOGI PENELITIAN	13

3.1	Sumber Data.....	13
3.2	Variabel Penelitian	13
3.3	Langkah Analisis.....	13
BAB IV ANALISIS DAN PEMBAHASAN.....		15
4.1	Karakteristik Kasus Campak dan Faktor-faktor yang Mempengaruhinya.....	15
4.2	Regresi Probit Biner	17
4.2.1	Uji Signifikansi Parameter Secara Serentak	17
4.2.2	Uji Signifikansi Parameter Secara Parsial	17
4.2.3	Model Regresi Probit Biner	18
4.2.4	Uji Kesesuaian Model	19
4.2.5	Klasifikasi Model Regresi Probit Biner.....	19
4.3	Regresi Skewed Probit Biner	19
4.3.1	Model Regresi Skewed Probit Biner	19
4.3.2	Klasifikasi Model Regresi Skewed Probit Biner	21
4.4	Perbandingan Model Regresi Probit Biner dan Regresi Skewed Probit Biner	21
BAB V KESIMPULAN DAN SARAN.....		23
5.1	Kesimpulan.....	23
5.2	Saran.....	23
DAFTAR PUSTAKA.....		25
LAMPIRAN.....		26

DAFTAR TABEL

Tabel 3.1 Variabel Penelitian	13
Tabel 4.1 Karakteristik Faktor-faktor yang Mempengaruhi Tingkat Kasus Campak di Kota Surabaya.....	15
Tabel 4.2 Nilai VIF.	17
Tabel 4.3 Uji Signifikansi Parameter Secara Serentak.	17
Tabel 4.4 Uji Signifikansi Parameter Secara Parsial.....	17
Tabel 4.5 Parameter Model Regresi Probit Biner.	18
Tabel 4.6 Uji Kesesuaian Model Regresi Probit Biner.	19
Tabel 4.7 Hasil Klasifikasi Model Regresi Probit Biner.....	19
Tabel 4.8 Hasil Estimasi Parameter Model Skewed Probit Biner.....	20
Tabel 4.9 Hasil Klasifikasi Model Regresi Probit Biner.....	21
Tabel 4.10 Perbandingan Model Regresi Probit Biner dan Regresi Skewed Probit Biner.	21

DAFTAR GAMBAR

Gambar 2.1 PC Prior untuk $\theta = 5$ pada Skala α (Kiri) dan Skala γ_1 (Kanan)	10
Gambar 2.2 PC Prior untuk Berbagai θ pada Skala α (Kiri) dan Skala γ_1 (Kanan)	11
Gambar 3.1 Diagram Alir.....	14
Gambar 4.1 Perbandingan Tingkat Kasus Campak di Kota Surabaya.....	15
Gambar 4.2 Densitas Posterior dengan Estimasi Titik yang Sesuai untuk Variabel (a) Jumlah Imunisasi Campak, (b) Jumlah Puskesmas, dan (c) Jumlah Tenaga Kesehatan.....	20
Gambar 4.3 Densitas Posterior dari Skewness γ_1 (Kiri) dan <i>Quantile of the Intercept</i> q (Kanan) dengan Estimasi Titik yang Sesuai.	21
Gambar 4.4 Perbandingan Akurasi dan Sensitivity Kedua Model.....	22

DAFTAR LAMPIRAN

Lampiran 1 Data Penelitian.....	26
Lampiran 2 <i>Syntax</i> Regresi Probit Biner.....	27
Lampiran 3 <i>Syntax</i> Regresi Skewed Probit Biner.	28

BAB I

PENDAHULUAN

1.1 Latar Belakang

Masalah kesehatan seperti penyakit campak hingga saat ini selalu menjadi perhatian dunia terutama Indonesia. Campak merupakan salah satu dari lima penyakit penyebab kematian anak di Indonesia (Nurani & Ginanjar, 2012). Meskipun vaksin telah dikembangkan dan tersedia pada seluruh provinsi di Indonesia, tetapi virus campak ini masih menyerang 50 juta orang setiap tahun dan menyebabkan lebih dari satu juta kematian di Indonesia (Nawaafila, 2015). Pada tahun 2015, ada hampir 5000 kasus campak yang terjadi di Indonesia dan secara kumulatif pada tahun 2015 terjadi sebanyak 23.164 kasus.

Salah satu program yang pernah dilakukan pemerintah untuk menanggulangi kasus campak di Indonesia ialah dengan imunisasi campak. Imunisasi merupakan salah satu cara untuk menanggulangi masalah kesehatan yang terjadi karena dalam waktu 4 hingga 6 minggu setelah imunisasi akan timbul antibodi spesifik yang aktif mencegah penularan penyakit, sehingga anak tidak mudah tertular infeksi, tidak menderita sakit yang berat, dan tidak terjadi wabah ataupun kematian (Nawaafila, 2015). Setiap daerah memiliki karakteristik masing-masing sehingga tidak bisa digeneralisasikan dengan hasil penelitian yang dilakukan di tempat lain. Rendahnya kesadaran dan pemahaman dari masyarakat, lingkungan sosial yang kurang mendukung, latar belakang ekonomi rendah, pengorganisasian dan pemberdayaan masyarakat yang belum optimal, advokasi dan regulasi yang tidak berjalan secara baik menjadi faktor yang menyebabkan permasalahan kesehatan yang terjadi di masyarakat. Pencegahan secara dini terhadap penyakit campak perlu dilakukan agar tidak terjadi penambahan angka kematian akibat penyakit campak. Sehingga perlu mengetahui pemodelan tingkat kasus campak pada balita di tiap daerah untuk menentukan langkah penanganan dalam mengurangi angka kematian akibat campak.

Surabaya merupakan salah satu Kota di Indonesia yang memiliki jumlah kasus campak yang cukup besar, sehingga pada penelitian ini akan dilakukan analisis mengenai pemodelan tingkat kasus campak pada balita di Kota Surabaya. Metode yang digunakan untuk mengetahui tingkat kasus campak pada tiap kecamatan di Kota Surabaya ialah dengan menggunakan regresi Probit Biner dan Skewed Probit Biner. Perbedaan antara regresi Probit Biner dan Skewed Probit Biner ialah terdapat pada pendekatan distribusi dalam mendapatkan probabilitas dan efek marginal pada tiap kategori, dimana regresi Probit dengan pendekatan distribusi normal, sedangkan regresi Skewed Probit dengan pendekatan distribusi Skew Normal. Faktor-

faktor yang diduga mempengaruhi tingkat kasus campak pada balita pada penelitian ini ialah jumlah imunisasi campak, jumlah puskesmas, dan jumlah tenaga kesehatan pada tiap kecamatan di Kota Surabaya. Tujuan dari penelitian ini diharapkan dapat menjadi pertimbangan Pemerintah Kota Surabaya khususnya Dinas Kesehatan Kota Surabaya, tenaga kesehatan, peneliti, dan masyarakat dalam mengetahui pemodelan tingkat kasus campak pada balita di Kota Surabaya dan dapat melakukan tindakan dalam mencegah terjadinya penambahan angka kematian akibat penyakit campak.

1.2 Rumusan Masalah

Berdasarkan latar belakang, maka rumusan masalah pada penelitian ini ialah sebagai berikut.

1. Bagaimana karakteristik tingkat kasus campak di Kota Surabaya dan faktor-faktor yang diduga mempengaruhinya?
2. Bagaimana model regresi Probit Biner untuk memprediksi tingkat kasus campak di Kota Surabaya?
3. Bagaimana model regresi Skewed Probit Biner untuk memprediksi tingkat kasus campak di Kota Surabaya?
4. Bagaimana perbandingan model regresi Probit dan Skewed Probit Biner?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah, maka tujuan yang diharapkan pada penelitian ini ialah sebagai berikut.

1. Mengetahui karakteristik tingkat kasus campak di Kota Surabaya dan faktor-faktor yang diduga mempengaruhinya.
2. Mendapatkan model regresi Probit Biner untuk memprediksi tingkat kasus campak di Kota Surabaya.
3. Mendapatkan model regresi Skewed Probit Biner untuk memprediksi tingkat kasus campak di Kota Surabaya.
4. Mengetahui perbandingan model regresi Probit dan Skewed Probit Biner.

1.4 Manfaat Penelitian

Manfaat yang diharapkan pada penelitian ini ialah menjadi informasi tambahan bagi Pemerintah Kota Surabaya khususnya Dinas Kesehatan Kota Surabaya mengenai pemodelan tingkat kasus campak pada balita di Kota Surabaya, sehingga dapat membuat kebijakan dan bahan pertimbangan untuk perbaikan dan penanggulangan penyakit campak pada balita di Kota Surabaya.

BAB II

TINJAUAN PUSTAKA

Pada penelitian ini digunakan beberapa metode analisis untuk dapat mencapai tujuan yang telah dipaparkan pada bab sebelumnya. Beberapa metode tersebut ditulis dan dijelaskan secara sistematis sesuai dengan urutan langkah-langkah penyelesaian analisis. Berikut penjelasan metode-metode yang digunakan pada penelitian ini.

2.1 Statistika Deskriptif

Statistika deskriptif adalah metode-metode yang berkaitan dengan pengumpulan data penyajian suatu gugus data sehingga memberikan informasi yang lebih bermakna. Statistika deskriptif tidak mengambil kesimpulan tentang gugus data induknya (yang lebih banyak), hanya penyusunan tabel, diagram, grafik, indeks, dan besaran-besaran lainnya (Sumanto, 2014).

a. Mean

Mean adalah nilai rata-rata dari beberapa buah data. Definisi lain dari mean adalah jumlah seluruh data dibagi dengan banyaknya data, maka mean dari data tersebut dapat dituliskan sebagai berikut.

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (2.1)$$

Keterangan:

\bar{x} = rata-rata

x_i = nilai x pada pengamatan ke- i

n = banyaknya data dalam ruang sampel

b. Standar Deviasi dan Varians

Standar deviasi dan varians merupakan salah satu teknik statistika yang digunakan untuk menjelaskan homogenitas kelompok. Varians merupakan jumlah kuadrat semua deviasi nilai-nilai individual terhadap rata-rata kelompok. Sedangkan akar dari varians disebut dengan standar deviasi atau simpangan baku.

$$\text{Varians}(S^2) = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \quad (2.2)$$

$$\text{Standar deviasi}(S) = \sqrt{S^2}$$

c. Minimum dan Maksimum

Nilai minimum adalah nilai terkecil atau terendah dari sekelompok data pengamatan, sedangkan nilai maksimum adalah nilai terbesar atau tertinggi dari sekelompok data pengamatan (Walpole, Myers, Myers, & Ye, 2012).

2.2 Uji Multikolinieritas

Pengujian multikolinearitas dilakukan untuk memperkecil nilai *error* pada taksiran yang dihasilkan. Pengujian multikolinieritas ini merupakan asumsi yang harus dipenuhi dalam regresi. Jika terdapat korelasi yang tinggi antar variabel prediktor akan mengakibatkan *error* yang sangat besar pada taksiran yang dihasilkan. Pengujian multikolinieritas dapat dideteksi dengan nilai *Variance Inflation Factors* (VIF). Nilai VIF yang lebih besar dari 10 menunjukkan adanya korelasi antar variabel prediktor. Nilai VIF dapat dihitung dengan formula sebagai berikut (Hosmer & Lemeshow, 2000).

$$VIF_k = \frac{1}{1 - R_k^2} \quad (2.3)$$

Dimana

$$R_k^2 = \frac{SSR}{SST} = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2}$$

Keterangan:

k = jumlah variabel prediktor dalam model

R_k^2 = koefisien determinasi antara variabel prediktor ke- j dengan variabel prediktor lainnya.

SSR = jumlah kuadrat regresi

SST = jumlah kuadrat total

2.3 Regresi Probit Biner

Regresi Probit Biner merupakan suatu model regresi yang dapat digunakan untuk menjelaskan hubungan antara variabel respon dengan variabel prediktor. Variabel respon yang

digunakan merupakan data yang bertipe kategori biner sedangkan variabel prediktor yang dimaksud dapat berupa data kontinu dan/atau diskrit berskala nominal dan/atau biner.

2.3.1 Model Regresi Probit Biner

Menurut Greene (2008), pemodelan regresi Probit Biner diawali dengan memperhatikan model sebagai berikut.

$$Y^* = \beta^T x + \varepsilon \quad (2.4)$$

dengan,

Y^* = variabel respon diskrit

β = vektor parameter koefisien

x = matriks variabel prediktor

ε = vektor *error* yang diasumsikan berdistribusi $N(0,1)$

Pada regresi Probit Biner dilakukan pengkategorian terhadap Y^* secara biner dengan memberikan batasan atau *threshold* (γ), yaitu untuk $Y^* \leq \gamma$ dikategorikan dengan $Y = 0$, dan untuk $Y^* > \gamma$ dikategorikan dengan $Y = 1$. Berikut ini adalah probabilitas untuk $Y = 0$.

$$\begin{aligned} P(Y = 0) &= P(Y^* \leq \gamma) \\ &= P(\beta^T x + \varepsilon \leq \gamma) \\ &= P(\varepsilon \leq \gamma - \beta^T x) \\ &= \Phi(\gamma - \beta^T x) \end{aligned} \quad (2.5)$$

dan probabilitas untuk $Y = 1$ ialah

$$\begin{aligned} P(Y = 1) &= P(Y^* > \gamma) \\ &= 1 - P(Y^* \leq \gamma) \\ &= 1 - P(\beta^T x + \varepsilon \leq \gamma) \\ &= 1 - P(\varepsilon \leq \gamma - \beta^T x) \\ &= 1 - \Phi(\gamma - \beta^T x) \end{aligned} \quad (2.6)$$

dengan $\Phi(\gamma - \beta^T x) = \Phi(\cdot)$ Adalah fungsi distribusi kumulatif distribusi normal standar.

Interpretasi model regresi Probit Biner tidak berdasarkan nilai koefisien model akan tetapi menggunakan efek marginal (Greene, 2008). Efek marginal dihasilkan dari turunan pertama probabilitas setiap kategori pada persamaan (2.5) dan (2.6) sebagai berikut.

$$\frac{\partial P(Y = 0)}{\partial X_i} = -\Phi(\gamma - \beta^T x)\beta_i \quad (2.7)$$

$$\frac{\partial P(Y = 1)}{\partial X_i} = \Phi(\gamma - \beta^T x)\beta_i \quad (2.8)$$

Nilai efek marginal pada persamaan (2.7) dan (2.8) menyatakan bahwa besarnya pengaruh tiap variabel prediktor yang signifikan terhadap probabilitas tiap kategori pada variabel respon.

2.3.2 Estimasi Parameter

Penaksiran parameter yang digunakan pada regresi Probit Biner adalah metode *Maximum Likelihood Estimation* (MLE). MLE merupakan salah satu metode pendugaan parameter yang dapat digunakan untuk menduga parameter suatu model yang sudah diketahui distribusinya. Untuk mendapatkan dugaan parameter dengan metode MLE dapat diperoleh dengan cara memaksimumkan fungsi *ln-likelihood*.

Menurut Ratnasari (2012), tahapan untuk mendapatkan penaksir parameter model regresi Probit Biner dengan menggunakan MLE diawali dengan mengambil n buah sampel *random*, yaitu Y_1, Y_2, \dots, Y_n dimana variabel respon Y memiliki dua kategori atau biner yang berdistribusi Bernoulli $(1, p)$. Untuk mendapatkan koefisien parameter (β) diawali dengan membentuk fungsi *likelihood* sebagai berikut.

$$L(\beta) = \prod_{i=1}^n \{ [1 - \phi(\gamma - \beta^T x)]^{y_i} [\phi(\gamma - \beta^T x)]^{1-y_i} \}$$

Kemudian memaksimumkan fungsi *ln likelihood* dengan cara melakukan turunan pertama fungsi $\ln L(\beta)$ terhadap β .

$$\frac{\partial \ln L(\beta)}{\partial \beta} = \sum_{i=1}^n x_i \phi(\beta^T x) \left[\frac{y_i}{1 - \phi(\gamma - \beta^T x)} + \frac{y_i - 1}{\phi(\gamma - \beta^T x)} \right] \quad (2.9)$$

Pada persamaan (2.9) diperoleh fungsi implisit sehingga penaksir β tidak langsung diperoleh atau disebut tidak *close form*. Cara untuk mendapatkan penaksir parameter β dapat menggunakan prosedur iterasi Newton Raphson. Komponen yang diperlukan dalam proses iterasi Newton Raphson adalah menentukan $g(\beta)$ merupakan vektor turunan pertama dari fungsi *ln likelihood* terhadap β yang merupakan persamaan (2.9) dan matrik Hessian $H(\beta)$ merupakan matriks turunan kedua dari fungsi *ln likelihood* terhadap parameter β dengan persamaan berikut.

$$\begin{aligned} \frac{\partial^2 \ln L(\beta)}{\partial \beta \partial \beta^T} = & - \sum_{i=1}^n x_i x_i^T y_i \frac{[1 - \phi(-\beta^T x)](-\beta^T x) \phi(-\beta^T x) + \phi(-\beta^T x) \phi(-\beta^T x)}{[1 - \phi(-\beta^T x)]^2} \\ & + \sum_{i=1}^n (1 - y_i) x_i x_i^T \frac{\phi(-\beta^T x) (-\beta^T x) \phi(-\beta^T x) - \phi(-\beta^T x) \phi(-\beta^T x)}{[\phi(-\beta^T x)]^2} \end{aligned} \quad (2.10)$$

Secara umum, iterasi dengan metode Newton Raphson untuk menaksir β dengan komponen yang digunakan pada persamaan (2.9) dan (2.10) adalah sebagai berikut.

$$\beta^{(m)} = \beta^{(m-1)} - \left(\frac{\partial^2 \ln L(\beta)}{\partial \beta^{(m-1)} \partial \beta^{(m-1)}} \right)^{-1} \frac{\partial \ln L(\beta)}{\partial \beta^{(m-1)}}$$

Proses iterasi akan berhenti jika terpenuhi kondisi konvergen, yaitu $\|\beta^{(m)} - \beta^{(m-1)}\| \leq \varepsilon$, dimana ε adalah bilangan yang sangat kecil.

2.3.3 Pengujian Estimasi Parameter

a. Uji Serentak

Setelah parameter hasil estimasi diperoleh, maka kemudian dilakukan pengujian signifikansi terhadap koefisien β secara serentak dengan hipotesis pengujian serentak sebagai berikut.

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

$$H_1 : \text{Paling tidak terdapat satu } \beta_j \neq 0 ; j = 1, 2, \dots, p$$

Statistik uji:

$$\begin{aligned} G^2 &= -2 \ln \left[\frac{L(\omega)}{L(\Omega)} \right] \\ &= 2 \ln L(\Omega) - 2 \ln L(\omega) \\ &= 2 [\ln L(\Omega) - \ln L(\omega)] \end{aligned} \tag{2.11}$$

Keputusan H_0 ditolak jika $G > \chi^2_{(\alpha/2, v)}$ atau $P - \text{value} < \alpha$ dengan v derajat bebas adalah banyaknya parameter dalam model dibawah populasi dikurangi dengan banyaknya parameter model dibawah H_0 .

b. Uji Parsial

Ketika dari uji serentak tidak mengeluarkan hasil yang signifikan maka kemudian dilakukan pengujian signifikansi terhadap koefisien β secara parsial terhadap variabel respon yaitu dengan membandingkan parameter hasil maksimum likelihood, dugaan β dengan standard error parameter tersebut. Hipotesis pengujian parsial ialah sebagai berikut.

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0 \text{ dengan } j = 1, 2, \dots, p$$

Statistik uji:

$$W = \left[\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \right]^2 = Z^2 \tag{2.12}$$

Statistik uji W (Wald) tersebut mengikuti distribusi *Chi-Squared* sehingga H_0 ditolak jika $W > \chi^2_{(\alpha/2, v)}$ dengan v merupakan derajat bebas atau banyaknya prediktor.

2.3.4 Uji Kesesuaian Model

Uji Kesesuaian model (*goodness of fit test*) merupakan pengujian yang digunakan untuk mengetahui apakah terdapat perbedaan antara hasil observasi dengan kemungkinan hasil prediksi *wald* (Hosmer & Lemeshow, 2000). Hipotesis yang digunakan dalam pengujian kesesuaian model ialah sebagai berikut.

H_0 : Tidak terdapat perbedaan antara hasil observasi dengan kemungkinan hasil prediksi model (model sesuai)

H_1 : Terdapat perbedaan antara hasil observasi dengan kemungkinan hasil prediksi model (model tidak sesuai)

Statistik uji yang digunakan dalam pengujian kesesuaian model sebagai berikut.

$$D = -2 \sum_{i=1}^n \left[y_i \ln \left(\frac{P_i}{y_i} \right) + (1 - y_i) \ln \left(\frac{1 - P_i}{1 - y_i} \right) \right] \quad (2.13)$$

Keputusan H_0 ditolak jika $D > \chi^2_{(\alpha/2, v)}$ dengan derajat bebas v sebesar $n - p - 1$ atau $P\text{-value} < \alpha$.

2.4 Regresi Skewed Probit Biner

Regresi Skewed Probit Biner merupakan perpanjangan dari regresi Probit Biner ketika fungsi link adalah CDF Skew-Normal. Regresi Skewed Probit dapat mencakup efek random seperti fungsi spline dari kovariat, efek spasial, dan temporal (Niekerk, 2021).

2.4.1 Model Regresi Skewed Probit Biner

Menurut Niekerk (2021), pemodelan regresi Skewed Probit Biner diawali dengan memperhatikan model sebagai berikut.

$$Y_i = \beta_0 + \beta' X_i + \sum_{i=1}^n f^k(Z_i) \quad (2.14)$$

dimana

Y_i = variabel respon diskrit

β = vektor parameter koefisien

X = matriks variabel prediktor

$f^k(Z_i)$ = efek random seperti spatial, spline, dan temporal

Mempertimbangkan regresi Skewed Probit dari Persamaan (2.14), probabilitas untuk tiap kategori ditunjukkan pada Persamaan (2.15) dan (2.16).

$$\text{Prob}[Y = 0] = p = F(\beta_0 + \beta_1 X | \alpha) \quad (2.15)$$

$$\text{Prob}[Y = 1] = 1 - p = 1 - F(\beta_0 + \beta_1 X | \alpha) \quad (2.16)$$

2.4.2 Parameter Skewness

Dalam mendapatkan PC (*Penalizing Complexity*) prior dari skewness, didapatkan Kullback-Leibler Divergence (KLD) dari link Skew-Normal untuk link Probit. KLD dari α ialah sebagai berikut:

$$\begin{aligned} KLD(\alpha) &= \int f(x | \alpha) \log \frac{f(x | \alpha)}{f(x | \alpha = 0)} \\ &= \frac{\pi^2 + 16 - 8\pi}{6\pi^3} \alpha^6 - \frac{144\pi + 3\pi^3 - 38\pi^2 - 168}{6\pi^4} \alpha^8 \\ &\quad + \frac{-42240\pi - 2560\pi^3 + 16176\pi^2 + 129\pi^4 + 39936}{120\pi^5} \alpha^{10} + o(\alpha^{12}) \\ &\approx c_1 \alpha^6 + c_2 \alpha^8 + c_3 \alpha^{10} \end{aligned} \quad (2.17)$$

dimana $\gamma = \alpha^3$ dan $\alpha = \text{sign}(\gamma) \sqrt[3]{|\gamma|}$.

Meskipun γ memiliki asimtotik biasa, estimasi darinya sulit diartikan sehingga dapat berfokus pada interpretasi lebih (standarisasi) skewness dari distribusi Skew-Normal, γ_1 yang merupakan fungsi monoton dari γ .

$$\gamma_1 = \frac{(4 - \pi) \left(\sqrt{\frac{2\delta^2}{\pi}} \right)^3}{2 \left(1 - \frac{2\delta^2}{\pi} \right)^{\frac{3}{2}}} \quad (2.18)$$

dengan $\delta = \frac{\alpha}{\sqrt{1 + \alpha^2}}$ dan interval skewness ialah $-0.99527 < \gamma_1 < 0.99527$.

2.4.3 Penalizing Complexity Prior untuk Parameter Skewness

PC (*Penalizing Complexity*) prior sangat membantu dan sangat dibutuhkan dalam kasus dimana prior secara tradisional dipilih karena kenyamanan matematis atau konvensi (Niekerk, 2021). PC prior telah digunakan di berbagai bidang penelitian, misalnya menurunkan PC prior untuk model autoregresif dan menurunkan PC prior untuk bidang random Gaussian. Pada penelitian ini PC prior diturunkan untuk α karena invarians dari PC prior dibawah

reparameterization dari parameter skewness. Penurunan PC prior untuk γ dan γ_1 mengikuti langsung dari perubahan variabel latihan. Menggunakan Persamaan (2.17), menentukan jarak satu arah dari Skew-Normal ke densitas Gaussian berikut:

$$d(\alpha) = \sqrt{2\text{KLD}(\alpha)} \approx \sqrt{2(c_1\alpha^6 + c_2\alpha^8 + c_3\alpha^{10})} \quad (2.19)$$

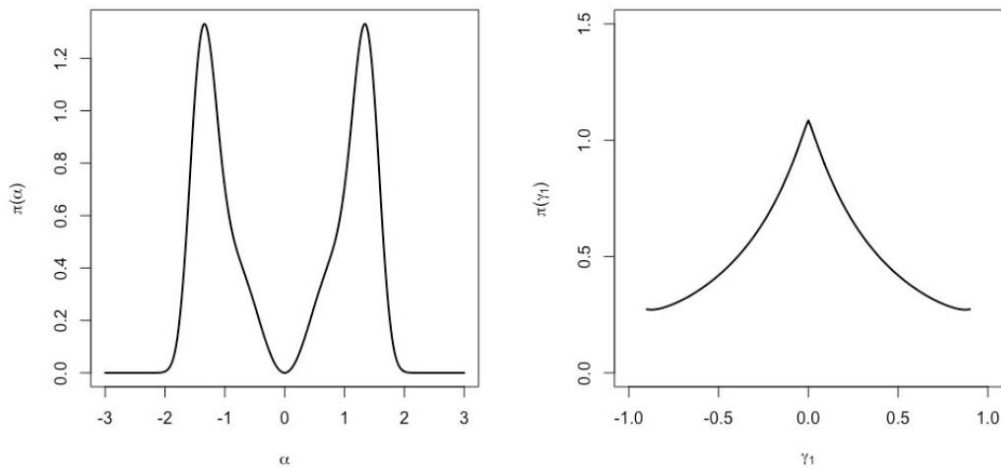
PC prior untuk parameter skewness α kemudian dibentuk dengan menetapkan prior eksponensial dengan parameter θ ke jarak. Parameter θ menggabungkan informasi dari pengguna untuk mengontrol perilaku ekor (*tail behavior*) dan dengan demikian laju kontraksi menuju fungsi probit link. PC prior mengikuti kemudian secara langsung, sebagai berikut:

$$\begin{aligned} \pi(\alpha) &= \frac{1}{2} \theta \exp[-\theta d(\alpha)] \left| \frac{\partial d(\alpha)}{\partial \alpha} \right| \\ &\approx \frac{\theta}{2\sqrt{2(c_1\alpha^6 + c_2\alpha^8 + c_3\alpha^{10})}} \left| 2(6c_1\alpha^5 + 8c_2\alpha^7 + 10c_3\alpha^9) \right| \\ &\quad \times \exp\left[-\theta|\alpha^3| \sqrt{2(c_1 + c_2\alpha^2 + c_3\alpha^4)}\right] \end{aligned} \quad (2.20)$$

Untuk nilai $|\alpha|$ yang kecil. Parameter yang ditentukan pengguna θ digunakan untuk mengatur kontraksi menuju regresi probit, misalnya untuk pu kecil dimana $pu > 0$ sehingga

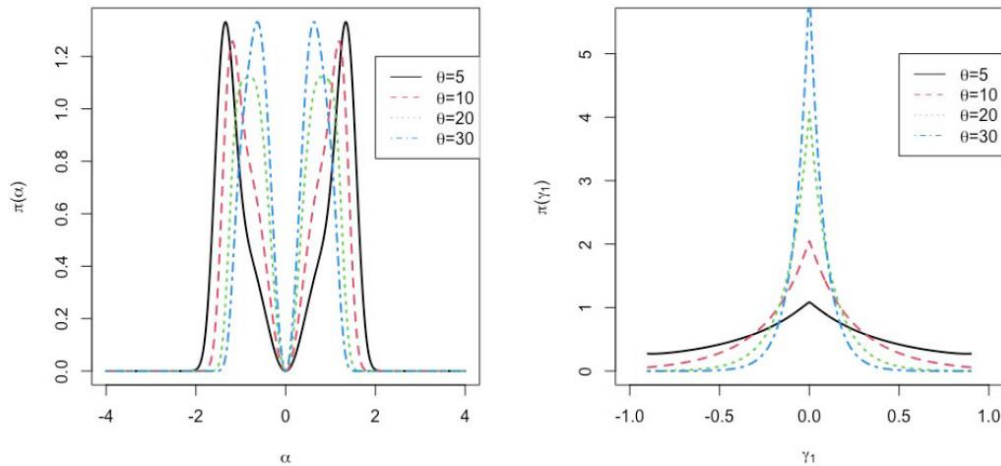
$$\text{Prob}(d(\alpha) > U) = pu = \exp(-\theta U)$$

yang memberikan $\theta = -\log pu/U$. Tidak ada ekspresi eksplisit untuk PC prior α secara umum, tetapi prior dapat dihitung secara numerik. Prior untuk γ_1 tersedia dalam paket R-INLA dengan prior = “pc.sn” dan parameter param = θ . Penulis menggunakan reparameterisasi γ_1 , karena γ_1 mengkuantifikasi skewness sebagai properti dengan interpretasi yang baik.



Gambar 2.1 PC Prior untuk $\theta = 5$ pada Skala α (Kiri) dan Skala γ_1 (Kanan)

PC prior α dan γ_1 diilustrasikan pada Gambar 1 untuk $\theta = 5$, pada skala α dan γ_1 . Pada Gambar 2 berbagai nilai untuk θ dianggap memberikan intuisi tentang pengaruh θ . Dari gambar ini dijelaskan bahwa nilai θ yang lebih besar menghasilkan tingkat kontraksi yang lebih tinggi dengan massa yang sedikit menjauh dari 0. Inferensi posterior kemiringan tidak sensitif terhadap nilai θ untuk sampel sedang dan besar. Dalam kasus sampel kecil, nilai θ yang sangat besar akan membuat estimator Bayes menuju 0 dengan kecepatan tinggi.



Gambar 2.2 PC Prior untuk Berbagai θ pada Skala α (Kiri) dan Skala γ_1 (Kanan)

2.5 Ketepatan Klasifikasi Model

Menurut Johnson dan Winchern (1992) evaluasi prosedur klasifikasi adalah suatu evaluasi yang melihat peluang kesalahan klasifikasi yang dilakukan oleh suatu fungsi klasifikasi. Ukuran yang dipakai adalah akurasi untuk data *balance* dan *sensitivity* untuk data *imbalanced*. Penentuan klasifikasi dapat dilihat dari Tabel 2.1 sebagai berikut.

Tabel 2.1 Tabel Klasifikasi

Hasil Observasi	Taksiran	
	π_1	π_2
y_1	n_{11}	n_{12}
y_2	n_{21}	n_{22}

Keterangan:

n_{11} = jumlah subjek dari y_1 tepat diklasifikasikan sebagai π_1

n_{12} = jumlah subjek dari y_1 salah diklasifikasikan sebagai π_2

n_{21} = jumlah subjek dari y_2 salah diklasifikasikan sebagai π_1

n_{22} = jumlah subjek dari y_2 tepat diklasifikasikan sebagai π_2

Rumus untuk penentuan kebaikan klasifikasi adalah sebagai berikut.

$$\text{Akurasi}(\%) = \frac{n_{11} + n_{22}}{n_{11} + n_{12} + n_{21} + n_{22}} \times 100\% \quad (2.21)$$

$$\text{Sensitivity}(\%) = \frac{n_{11}}{n_{11} + n_{12}} \times 100\% \quad (2.22)$$

2.6 Campak

Campak merupakan penyakit sangat menular yang disebabkan oleh virus dan dapat mengakibatkan kematian. Kematian pada campak sebagian besar disebabkan oleh komplikasi diantaranya diare, peumonia, dan ensefalitis (Kementerian Kesehatan Republik Indonesia, 2020). Indonesia termasuk ke dalam 10 negara dengan jumlah kasus campak terbesar di dunia. Manusia diperkirakan satu-satunya reservoir, walaupun monyet dapat terinfeksi tetapi tidak berperan dalam penyebaran.

BAB III

METODOLOGI PENELITIAN

Sumber data, variabel penelitian, dan langkah analisis yang digunakan pada penelitian ini adalah sebagai berikut.

3.1 Sumber Data

Sumber data yang digunakan dalam penelitian ini adalah data sekunder dari publikasi Dinas Kesehatan Kota Surabaya dengan judul Profil Kesehatan Kota Surabaya tahun 2019 dan publikasi dari Badan Pusat Statistik (BPS) dengan judul Surabaya Dalam Angka tahun 2020. dengan unit penelitian sebesar 31 data. Variabel respon pada data tersebut ialah tingkat kasus campak pada tiap kecamatan di Kota Surabaya dengan kategori 1 untuk tingkat darurat (di atas rata-rata) dan 0 untuk tingkat normal (di bawah rata-rata). Sedangkan variabel prediktor yang digunakan pada penelitian ini ialah jumlah imunisasi campak, jumlah puskesmas, dan jumlah tenaga kesehatan pada tiap kecamatan di Kota Surabaya.

3.2 Variabel Penelitian

Variabel penelitian yang digunakan dalam penelitian ini adalah sebagai berikut.

Tabel 3.1 Variabel Penelitian

Variabel	Nama Variabel	Skala
Y	Tingkat Kasus Campak	Nominal
	1 = Darurat	
	0 = Normal	
X ₁	Jumlah Imunisasi Campak	Rasio
X ₂	Jumlah Puskesmas	Rasio
X ₃	Jumlah Tenaga Kesehatan	Rasio

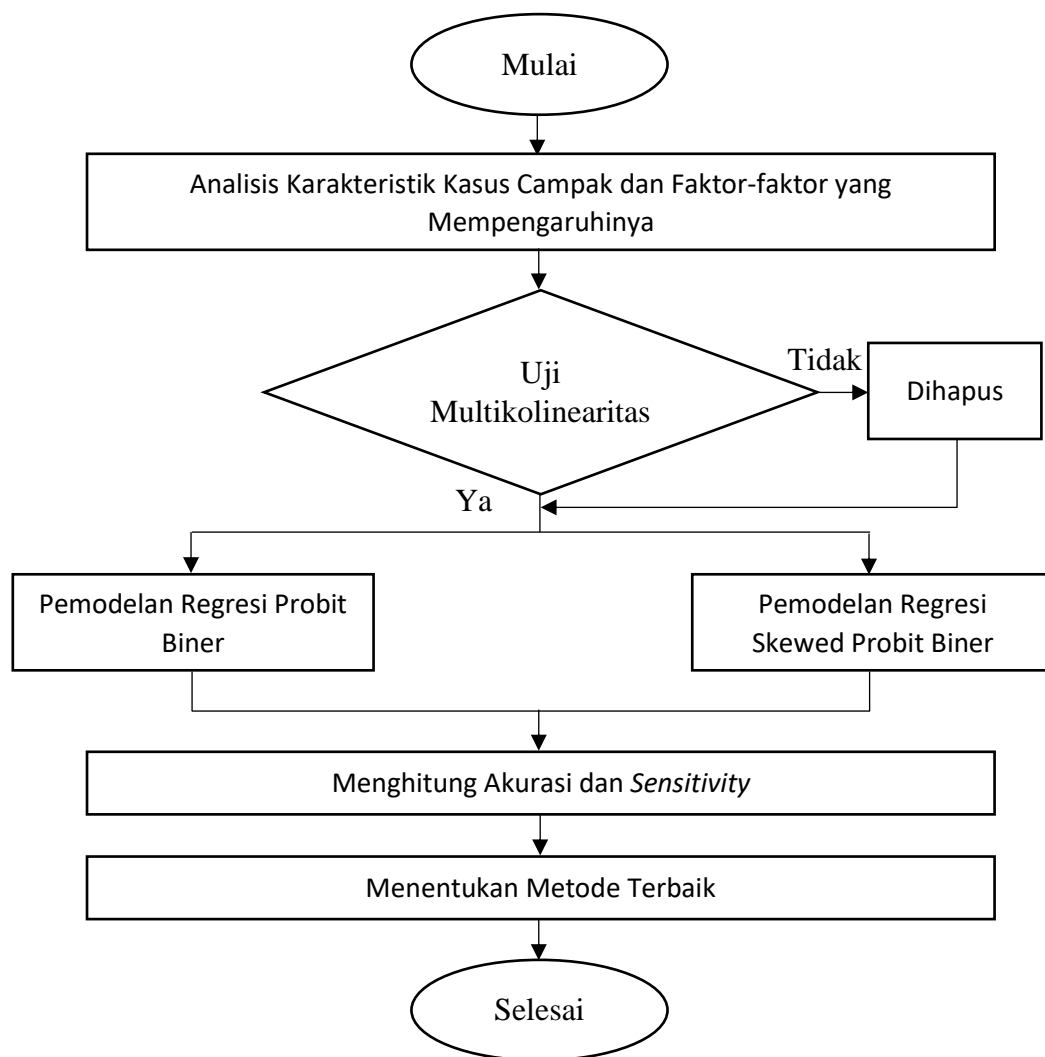
3.3 Langkah Analisis

Langkah analisis yang dilakukan dalam penelitian ini adalah sebagai berikut.

1. Mengetahui karakteristik tingkat kasus campak di Kota Surabaya dan faktor - faktor yang mempengaruhinya menggunakan statistika deskriptif, serta melakukan uji multikolinearitas.
2. Membentuk model regresi Probit Biner tingkat kasus campak pada balita di Kota Surabaya dengan uraian sebagai berikut:
 - a. Melakukan uji signifikansi parameter secara serentak.
 - b. Melakukan uji signifikansi parameter secara parsial.

- c. Menentukan model regresi Probit Biner.
 - d. Uji kesesuaian model.
 - e. Mengukur ketepatan klasifikasi model.
3. Membentuk model regresi Skewed Probit Biner tingkat kasus campak pada balita di Kota Surabaya dengan uraian sebagai berikut:
 - a. Menentukan model regresi Skewed Probit Biner.
 - b. Mengukur ketepatan klasifikasi model.
4. Menghitung nilai akurasi dan *sensitivity* pada tiap model untuk mengetahui model terbaik.

Berdasarkan langkah analisis yang telah dijelaskan sebelumnya, maka visualisasi diagram alirnya ditampilkan pada Gambar 3.1.



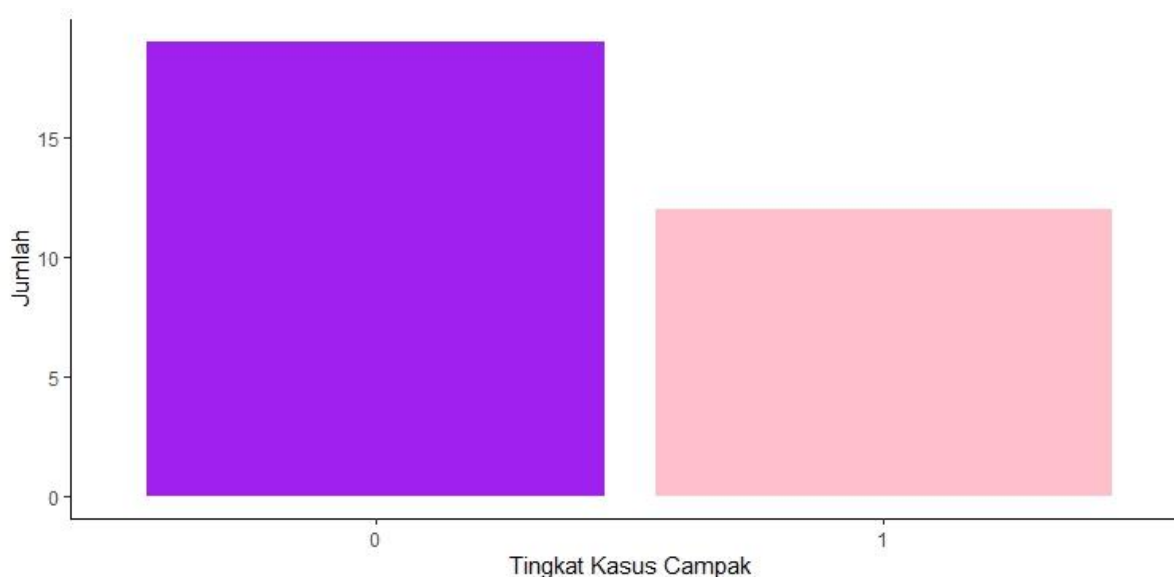
Gambar 3.1 Diagram Alir.

BAB IV

ANALISIS DAN PEMBAHASAN

4.1 Karakteristik Kasus Campak dan Faktor-faktor yang Mempengaruhinya

Pada setiap kecamatan di Kota Surabaya mempunyai tingkat kasus campak yang berbeda-beda. Hal ini mungkin disebabkan oleh ketidakmerataannya sarana kesehatan dan dapat pula disebabkan karena kebersihan serta pola hidup masyarakat di masing-masing kecamatan. Gambar 4.1 menunjukkan perbandingan tingkat kasus campak darurat dan normal di Kota Surabaya.



Gambar 4.1 Perbandingan Tingkat Kasus Campak di Kota Surabaya

Gambar 4.1 menunjukkan bahwa sebanyak 12 Kecamatan di Kota Surabaya memiliki tingkat kasus campak dengan status darurat dan 19 Kecamatan lainnya dalam status normal. Hal tersebut menunjukkan bahwa pada penelitian ini memiliki data yang *imbalanced*. Karakteristik faktor-faktor yang mempengaruhi tingkat kasus campak pada tiap kecamatan di Kota Surabaya ditampilkan pada Tabel 4.1 berikut.

Tabel 4.1 Karakteristik Faktor-faktor yang Mempengaruhi Tingkat Kasus Campak di Kota Surabaya.

Variabel	Mean	Variance	Minimum	Median	Maximum
Jumlah Imunisasi	1315,00	428838,00	599,00	1142,00	3031,00
Jumlah Puskesmas	2,03	0,97	1,00	2,00	4,00
Jumlah Tenaga Kesehatan	49,90	473,56	21,00	43,00	100,00

Imunisasi memberikan kekebalan tubuh sehingga dapat terhindar dari penyakit campak. Pada Tabel 4.1 menjelaskan bahwa rata-rata pemberian imunisasi pada tiap kecamatan ialah sebesar 1315 balita dengan keragamannya yang sangat besar yaitu sebesar 428838, artinya

pemberian imunisasi di setiap kecamatan memiliki perbedaan yang jauh antar kecamatan satu sama lain dimana 50% kecamatan melakukan pemberian imunisasi campak ke lebih dari 1142 balita, sedangkan 50% lainnya di bawah 1142 balita. Adapun pemberian imunisasi terbanyak ialah sebesar 3031 balita yaitu pada Kecamatan Tambaksari, sedangkan pemberian imunisasi paling sedikit sebesar 599 balita yaitu pada Kecamatan Gayungan.

Puskesmas dengan program posyandu memberikan kemudahan kepada masyarakat dalam memperoleh pelayanan kesehatan dasar yang diperlukan. Pada Tabel 4.1 menjelaskan bahwa rata-rata jumlah puskesmas pada tiap kecamatan di Kota Surabaya ialah sebesar 2 puskesmas dengan keragamannya yang cukup kecil yaitu 0,97 artinya jumlah puskesmas pada tiap kecamatan cukup homogen dimana 50% kecamatan memiliki puskesmas lebih dari 2 dan 50% lainnya memiliki puskesmas kurang dari 2. Adapun jumlah puskesmas terbanyak terdapat pada Kecamatan Semampir, Kecamatan Kenjeran, dan Kecamatan Sawahan sebesar 4 puskesmas, sedangkan jumlah puskesmas paling sedikit terdapat pada Kecamatan Asemrowo, Kecamatan Benowo, Kecamatan Pakal, Kecamatan Pabean Cantikan, Kecamatan Bulak, Kecamatan Tenggilis, Kecamatan Gunung Anyar, Kecamatan Karang Pilang, Kecamatan Dukuh Pakis, Kecamatan Gayungan, dan Kecamatan Jambangan.

Tenaga kesehatan merupakan salah satu faktor penting yang diduga mempengaruhi penanganan campak pada tiap kecamatan. Semakin tinggi jumlah tenaga kesehatan diharapkan semakin mempercepat penanganan campak pada tiap kecamatan di Kota Surabaya. Berdasarkan Tabel 4.1 diketahui bahwa rata-rata jumlah tenaga kesehatan di setiap kecamatan ialah sebesar 50 orang dengan keragamannya yang cukup besar yaitu 473,56 yang berarti jumlah tenaga kesehatan di setiap kecamatan memiliki perbedaan yang cukup besar dimana 50% kecamatan memiliki jumlah tenaga kesehatan lebih dari 43 orang dan 50% lainnya memiliki jumlah tenaga kesehatan dibawah 43 orang. Adapun jumlah tenaga kesehatan terbanyak terdapat pada Kecamatan Sawahan sebesar 100 orang, sedangkan jumlah tenaga kesehatan paling sedikit terdapat pada Kecamatan Asemrowo sebesar 21 orang.

Sebelum melanjutkan analisis dengan menggunakan regresi Probit dan regresi Skewed Probit Biner, salah satu asumsi yang harus dipenuhi dalam pemodelan regresi adalah tidak terjadi multikolinearitas. Salah satu pendeteksian yang dapat dilakukan yakni dengan memperhatikan nilai VIF (*Variance Inflation Factors*). Apabila nilai VIF lebih dari 10 maka terdapat indikasi terjadinya multikolinearitas pada variabel tersebut. Sebelum mengetahui nilai VIF, dilakukan pemodelan regresi antar variabel prediktor terlebih dahulu dan kemudian menggunakan Persamaan (2.3) untuk mendapatkan nilai VIF. Hasil perhitungan nilai VIF ditampilkan pada Tabel 4.2 berikut:

Tabel 4.2 Nilai VIF.

Variabel	VIF
Jumlah Imunisasi	3,044
Jumlah Puskesmas	5,905
Jumlah Tenaga Kesehatan	4,824

Berdasarkan Tabel 4.2 dapat dilihat bahwa nilai VIF dari seluruh variabel memiliki nilai VIF yang kurang dari 10. Sehingga dapat dikatakan bahwa tidak terdeteksi adanya multikolinearitas pada variabel yang akan digunakan dalam analisis ini dan dapat dikatakan bahwa variabel tersebut telah memenuhi asumsi dalam regresi, maka analisis dapat dilanjutkan ke tahap selanjutnya.

4.2 Regresi Probit Biner

Regresi Probit Biner ini dilakukan untuk mendapatkan model dengan variabel respon berbentuk biner (2 kategori). Pemodelan dilakukan dengan meregresikan semua kemungkinan kombinasi variabel yang ada yaitu sebanyak 3 variabel.

4.2.1 Uji Signifikansi Parameter Secara Serentak

Langkah pertama yang dilakukan ialah pengujian signifikansi parameter secara serentak. Hal ini dilakukan untuk mengetahui apakah terdapat pengaruh yang signifikan antara semua parameter dalam model dan hasilnya ditunjukkan pada Tabel 4.3 berikut.

Tabel 4.3 Uji Signifikansi Parameter Secara Serentak.

G^2	<i>P-Value</i>	Keputusan
13,530	0.004	Tolak H_0

Tabel 4.3 menunjukkan bahwa nilai *p-value* kurang dari nilai α sebesar 0,05, sehingga didapatkan keputusan tolak H_0 atau dapat dikatakan bahwa minimal terdapat satu parameter yang berpengaruh signifikan dalam model.

4.2.2 Uji Signifikansi Parameter Secara Parsial

Pada penelitian ini uji signifikansi parameter secara parsial menggunakan uji *Z* dan *Wald*. Hasil pengujian signifikansi parameter secara parsial ditampilkan pada Tabel 4.4 berikut.

Tabel 4.4 Uji Signifikansi Parameter Secara Parsial.

Variabel	<i>Z-Value</i>	<i>P-Value</i>	<i>Wald</i>	<i>P-Value</i>
<i>Intercept</i>	-2,623	0,009	6,878	0,014
Jumlah Imunisasi	2,115	0,034	4,473	0,044
Jumlah Puskesmas	0,932	0,351	0,869	0,360
Jumlah Tenaga Kesehatan	-1,487	0,137	2,210	0,149

Berdasarkan *p-value* yang ditampilkan Tabel 4.4, dapat diketahui bahwa hanya variabel jumlah imunisasi saja yang berpengaruh signifikan dalam model, sedangkan variabel jumlah puskesmas dan jumlah tenaga kesehatan tidak berpengaruh signifikan dalam model dengan menggunakan nilai α sebesar 0,05. Sehingga variabel jumlah puskesmas dan jumlah tenaga

kesehatan dapat dikeluarkan dari model dan hanya variabel jumlah imunisasi saja yang dapat digunakan untuk mendapatkan model tingkat kasus campak di Kota Surabaya menggunakan regresi Probit Biner.

4.2.3 Model Regresi Probit Biner

Setelah mengetahui variabel yang signifikan dalam model regresi Probit Biner, maka didapatkan hasil estimasi parameter yang ditampilkan pada Tabel 4.5 berikut.

Tabel 4.5 Parameter Model Regresi Probit Biner.

Variabel	Estimasi
<i>Intercept</i>	-2,232
Jumlah Imunisasi	0,001

Tabel 4.5 menunjukkan hasil estimasi parameter untuk variabel prediktor dalam model tingkat kasus campak di Kota Surabaya. Sehingga model regresi Probit Biner yang dapat dibentuk ialah sebagai berikut.

$$Y^* = -2,232 + 0,001 X_1$$

Probabilitas model regresi Probit Biner ditunjukkan pada persamaan berikut ini:

$$\hat{P}(Y=0) = \Phi(-2,232 + 0,001X_1)$$

$$\hat{P}(Y=1) = 1 - \Phi(-2,232 + 0,001X_1)$$

Interpretasi dari model regresi Probit Biner tersebut menggunakan efek marginal yang digunakan untuk mengetahui besarnya pengaruh dari variabel prediktor yang signifikan terhadap probabilitas setiap kategori pada variabel respon. Efek marginal dibentuk berdasarkan Persamaan (2.7) dan (2.8). Besarnya efek marginal dari variabel jumlah imunisasi ialah sebagai berikut:

$$\frac{\partial \hat{P}(Y=0)}{\partial X_1} = -0,001\Phi(-2,232 + 0,001X_1)$$

$$\frac{\partial \hat{P}(Y=1)}{\partial X_1} = 0,001\Phi(-2,232 + 0,001X_1)$$

Berdasarkan persamaan efek marginal jumlah imunisasi tersebut dapat diketahui besarnya pengaruh jumlah imunisasi terhadap tingkat kasus campak di Kota Surabaya. Sebagai contoh misalkan jumlah imunisasi campak pada balita di Kecamatan Sukolilo sebesar 1562, maka efek marginal jumlah imunisasi terhadap tingkat kasus campak di Kota Surabaya berturut-turut sebesar -0,00025 dan 0,00025. Berdasarkan nilai tersebut dapat diketahui bahwa nilai efek jumlah imunisasi menurunkan probabilitas sebesar 0,00025 untuk Kecamatan Sukolilo masuk dalam tingkat kasus campak normal. Sedangkan nilai efek marginal jumlah

imunisasi menaikkan probabilitas sebesar 0,00025 untuk Kecamatan Sukolilo masuk dalam tingkat kasus campak darurat.

4.2.4 Uji Kesesuaian Model

Setelah mendapatkan model regresi Probit Biner yang optimal, kemudian dilakukan uji kesesuaian model untuk mengetahui apakah model yang terbentuk telah sesuai dan hasilnya ditampilkan pada Tabel 4.6.

Tabel 4.6 Uji Kesesuaian Model Regresi Probit Biner.

<i>Deviance</i>	<i>Df</i>	<i>P-Value</i>
30,394	29	0.395

Berdasarkan Tabel 4.6 diketahui bahwa nilai *p-value* besar dari nilai α sebesar 0,05, sehingga keputusannya ialah gagal tolak H_0 atau dapat dikatakan bahwa model yang terbentuk telah sesuai dalam memprediksi tingkat kasus campak pada balita di Kota Surabaya.

4.2.5 Klasifikasi Model Regresi Probit Biner

Menggunakan model regresi Probit Biner yang optimal, selanjutnya dilakukan klasifikasi untuk mengukur kebaikan model dalam memprediksi tingkat kasus campak pada balita tiap Kecamatan di Kota Surabaya dengan hasil ditampilkan pada Tabel 4.7.

Tabel 4.7 Hasil Klasifikasi Model Regresi Probit Biner.

Observasi		Prediksi	
		Tingkat Kasus Campak	
		0	1
Tingkat Kasus	0	17	2
Campak	1	4	8

Berdasarkan hasil klasifikasi yang ditampilkan pada Tabel 4.7, maka dapat dilakukan perhitungan nilai akurasi dan *sensitivity* untuk melihat kebaikan model regresi Probit. Nilai *sensitivity* digunakan karena lebih baik dalam melihat kebaikan hasil klasifikasi untuk data yang *imbalanced*. Nilai akurasi untuk model regresi Probit Biner ialah sebesar 80,65%, sedangkan untuk nilai *sensitivity* sebesar 80,95%.

4.3 Regresi Skewed Probit Biner

Regresi Skewed Probit Biner dilakukan untuk mendapatkan model dengan variabel respon berbentuk biner yang *imbalanced* untuk perbandingan kedua kategorinya. Pemodelan regresi Skewed Probit Biner tersedia dalam *software* R menggunakan *package* R-INLA.

4.3.1 Model Regresi Skewed Probit Biner

Hasil estimasi parameter untuk model Regresi Skewed Probit Biner ditampilkan pada Tabel 4.8 berikut.

Tabel 4.8 Hasil Estimasi Parameter Model Skewed Probit Biner.

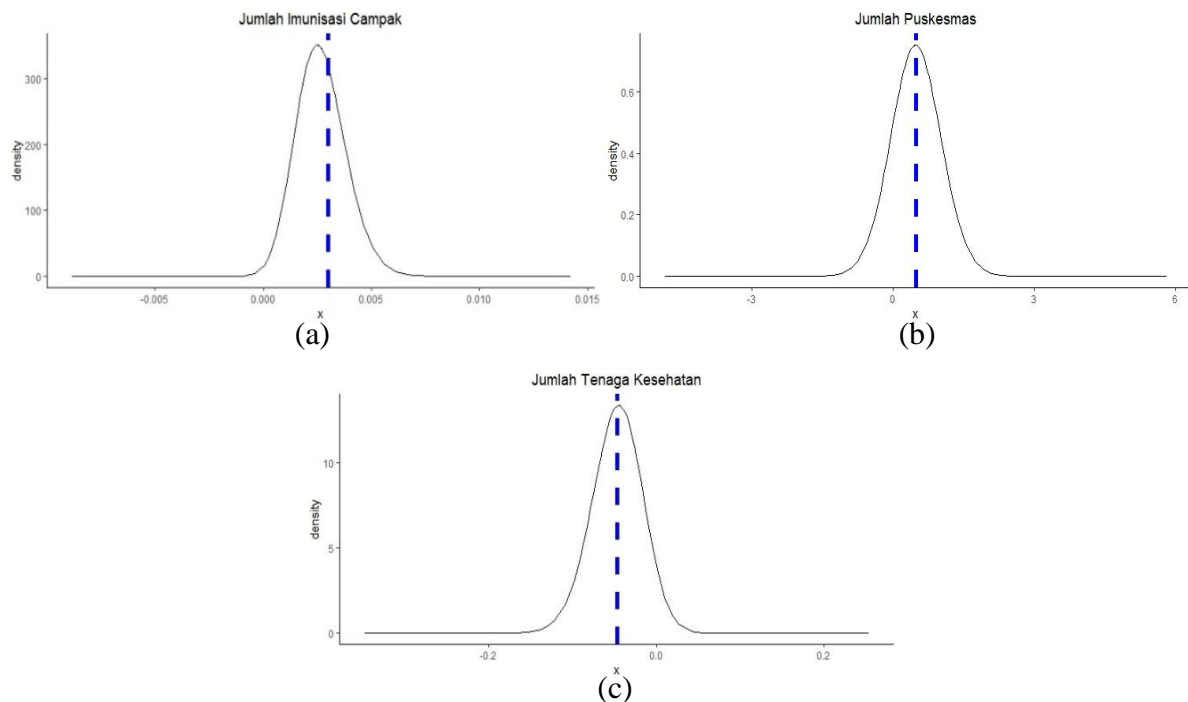
	<i>Posterior mean</i>	<i>95% Credible Interval</i>
<i>Quantile of the intercept (q)</i>	0,025	(0,003 ; 0,087)
X_1	0,003	(0,001 ; 0,005)
X_2	0,482	(-0,561 ; 1,538)
X_3	-0,047	(-0,108 ; 0,010)
<i>Skewness (γ_1)</i>	-0,017	(-0,289 ; 0,241)

Berdasarkan Tabel 4.8 dapat dilihat bahwa hasil estimasi parameter dari model Skew Probit Biner memiliki kesamaan dengan hasil estimasi parameter dari model Probit Biner, yaitu koefisien bernilai positif untuk variabel jumlah imunisasi dan koefisien bernilai negatif untuk variabel jumlah tenaga kesehatan. Probabilitas model Skew Probit Biner ditunjukkan pada persamaan berikut ini:

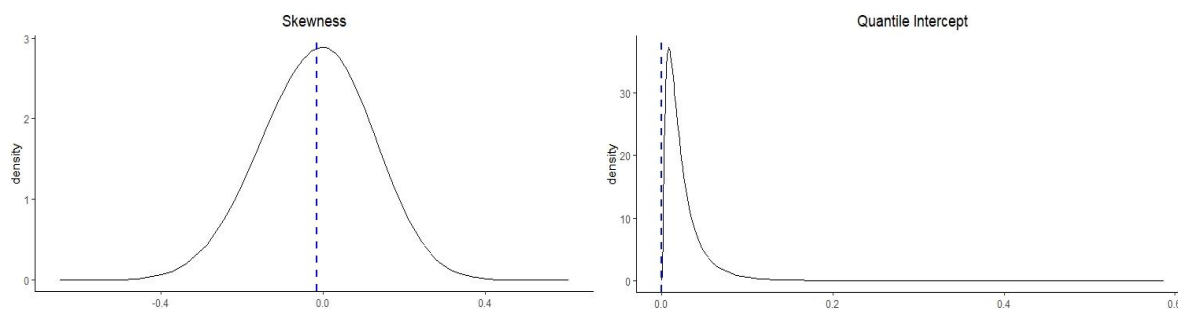
$$\hat{P}(Y=0) = F(0,025 + 0,003X_1 + 0,482X_2 - 0,047X_3 | -0,017)$$

$$\hat{P}(Y=1) = 1 - F(0,025 + 0,003X_1 + 0,482X_2 - 0,047X_3 | -0,017)$$

Densitas posterior dari efek tetap semua variabel prediktor ditampilkan pada Gambar 3 dimana nilai koefisien ditampilkan pada garis vertikal berwarna biru dan densitas posterior dari skewness γ_1 dan q (*quantile of the intercept*) dengan estimasi titik yang sesuai yang ditampilkan dengan garis vertikal berwarna biru ditampilkan pada Gambar 4. Dikarenakan regresi Skewed Probit Biner merupakan perpanjangan dari regresi Probit Biner ketika fungsi link adalah CDF Skew-Normal, maka nilai *density* akan berbentuk distribusi Skew-Normal.



Gambar 4.2 Densitas Posterior dengan Estimasi Titik yang Sesuai untuk Variabel (a) Jumlah Imunisasi Campak, (b) Jumlah Puskesmas, dan (c) Jumlah Tenaga Kesehatan.



Gambar 4.3 Densitas Posterior dari Skewness γ_1 (Kiri) dan *Quantile of the Intercept* q (Kanan) dengan Estimasi Titik yang Sesuai.

4.3.2 Klasifikasi Model Regresi Skewed Probit Biner

Setelah mendapatkan model regresi Skewed Probit Biner, selanjutnya dilakukan klasifikasi untuk mengukur kebaikan model dalam memprediksi tingkat kasus campak pada balita tiap Kecamatan di Kota Surabaya dengan hasil ditampilkan pada Tabel 4.9.

Tabel 4.9 Hasil Klasifikasi Model Regresi Probit Biner.

Observasi		Prediksi	
		Tingkat Kasus Campak	
		0	1
Tingkat Kasus Campak	0	16	3
	1	3	9

Berdasarkan hasil klasifikasi yang ditampilkan Tabel 4.9, maka didapatkan nilai akurasi untuk model regresi Skewed Probit Biner sebesar 80,65% dan nilai *sensitivity* sebesar 84,21%.

4.4 Perbandingan Model Regresi Probit Biner dan Regresi Skewed Probit Biner

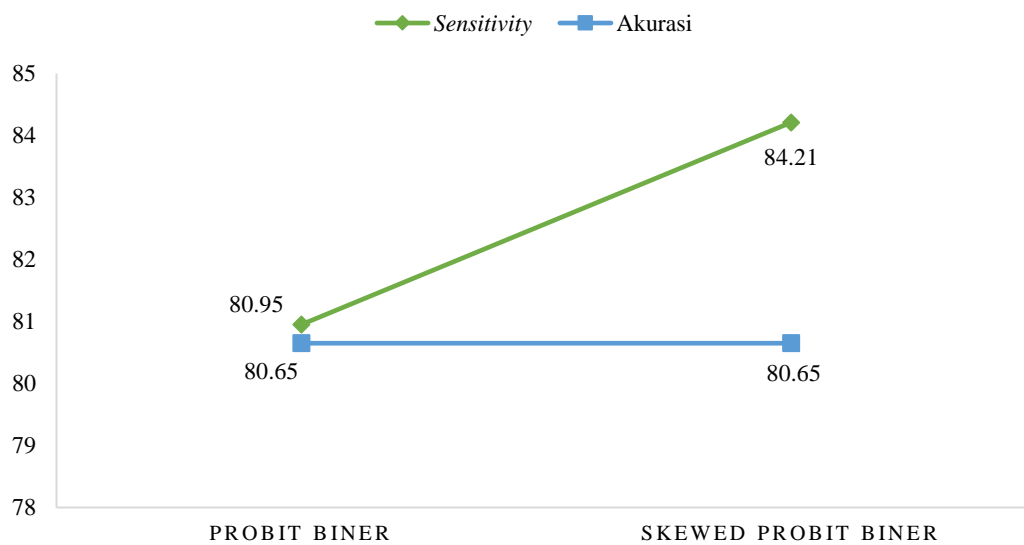
Setelah mendapatkan model regresi Probit Biner dan regresi Skewed Probit Biner, selanjutnya dilakukan perbandingan kebaikan model berdasarkan nilai akurasi dan *sensitivity*. Hasil perbandingan nilai akurasi dan *sensitivity* untuk kedua model ditampilkan pada Tabel 4.10.

Tabel 4.10 Perbandingan Model Regresi Probit Biner dan Regresi Skewed Probit Biner.

Model	Akurasi	<i>Sensitivity</i>
Regresi Probit Biner	80,65%	80,95%
Regresi Skewed Probit Biner	80,65%	84,21%

Tabel 4.10 menunjukkan bahwa berdasarkan nilai akurasi kedua model memberikan ketepatan klasifikasi yang sama. *Sensitivity* sebaiknya digunakan sebagai ukuran kebaikan model dikarenakan cukup baik dalam memberikan informasi kebaikan model untuk data *imbalanced*. *Sensitivity* merupakan proporsi kecamatan tepat diprediksi memiliki tingkat kasus campak darurat dari seluruh kecamatan yang memang memiliki tingkat kasus campak darurat.

Secara visual perbandingan kebaikan model regresi Probit Biner dan regresi Skewed Probit Biner ditampilkan pada Gambar 4.4.



Gambar 4.4 Perbandingan Akurasi dan *Sensitivity* Kedua Model.

Berdasarkan nilai *sensitivity* yang ditampilkan pada Gambar 4.4 diketahui bahwa regresi Skewed Probit Biner lebih baik dibandingkan model regresi Probit Biner. Hal tersebut menunjukkan bahwa model regresi Skewed Probit Biner lebih baik dalam memprediksi tingkat kasus campak pada tiap Kecamatan di Kota Surabaya dengan nilai *sensitivity* sebesar 84,21%.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Kesimpulan yang dapat diambil dari penelitian ini ialah sebagai berikut.

1. Sebanyak 12 Kecamatan di Kota Surabaya memiliki tingkat kasus campak dengan status darurat dan 19 Kecamatan lainnya dalam status normal. Hal tersebut menunjukkan bahwa pada penelitian ini memiliki data yang *imbalanced*. Berdasarkan hasil pengujian multikolinearitas didapatkan bahwa seluruh variabel prediktor yang digunakan pada penelitian ini memiliki nilai VIF yang kurang dari 10, sehingga dapat disimpulkan tidak terdeteksi adanya multikolinearitas.
2. Berdasarkan analisis regresi Probit Biner, faktor yang berpengaruh signifikan terhadap tingkat kasus campak di Kota Surabaya ialah jumlah imunisasi campak dengan probabilitas model Probit biner ditunjukkan pada persamaan berikut:

$$\hat{P}(Y=0) = \Phi(-2,232 + 0,001X_1)$$

$$\hat{P}(Y=1) = 1 - \Phi(-2,232 + 0,001X_1)$$

3. Berdasarkan analisis regresi Skewed Probit Biner dengan menggunakan faktor jumlah imunisasi campak, jumlah puskesmas, dan jumlah tenaga kesehatan didapatkan probabilitas model Skewed Probit Biner untuk memprediksi tingkat kasus campak pada tiap Kecamatan di Kota Surabaya yang ditunjukkan pada persamaan berikut:

$$\hat{P}(Y=0) = F(0,025 + 0,003X_1 + 0,482X_2 - 0,047X_3 | -0,017)$$

$$\hat{P}(Y=1) = 1 - F(0,025 + 0,003X_1 + 0,482X_2 - 0,047X_3 | -0,017)$$

4. Nilai akurasi dari hasil klasifikasi model regresi Probit Biner dan regresi Skewed Probit Biner sama yaitu sebesar 80,65%, sedangkan berdasarkan nilai *sensitivity* diketahui bahwa regresi Skewed Probit Biner lebih baik dibandingkan model regresi Probit Biner. Hal tersebut menunjukkan bahwa model regresi Skewed Probit Biner lebih baik dalam memprediksi tingkat kasus campak pada tiap Kecamatan di Kota Surabaya dengan nilai *sensitivity* sebesar 84,21%.

5.2 Saran

Dalam penelitian ini menggunakan 3 variabel prediktor yang diduga mempengaruhi tingkat kasus campak pada balita di Kota Surabaya, sehingga untuk penelitian selanjutnya dapat menambahkan variabel lainnya. Kemudian pada penelitian ini data *imbalanced* tidak diatasi, sehingga pada penelitian selanjutnya dapat melakukan sampling seperti SMOTE

(*Synthetic Minority Oversampling Technique*) untuk menangani data *imbalanced*. Selain itu penelitian selanjutnya dapat melakukan metode lain yang cukup baik dalam menangani data *imbalanced* seperti Random Forest, Neural Network, dan sebagainya.

DAFTAR PUSTAKA

- Greene, W. (2008). *Econometrics Analysis (6th ed.)*. New Jersey: Prentice Hall, Inc.
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression*. New York: John Wiley & Sons, Inc.
- Kementerian Kesehatan Republik Indonesia. (2020). *Pedoman Surveilans Campak-Rubela*. Indonesia: Kementerian Kesehatan Republik Indonesia.
- Nawaafila. (2015). *Pemodelan Jumlah Penderita Campak di Indonesia dengan Pendekatan Regresi Nonparametrik Spline*. Surabaya: Institut Teknologi Sepuluh Nopember.
- Niekerk, J. (2021). Skewed Probit Regression-Identifiability, Contraction and Reformulation. *REVSTAT-Statistical Journal*, 19, 1-22.
- Nurani, D., & Ginanjar, P. (2012). Gambaran Epidemiologi Kasus Campak di Kota Cirebon Tahun 2004-2011 (Studi Kasus Data Surveilans Epidemiologi Campak di Dinas Kesehatan Kota Cirebon). *Jurnal Kesehatan Masyarakat (JKM)*, 293-304.
- Ratnasari, V. (2012). *Estimasi Parameter dan Uji Signifikansi Model Probit Bivariat*. Surabaya: Jurusan Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Institut Teknologi Sepuluh Nopember.
- Sumanto. (2014). *Statistika Deskriptif Edisi ke-1*. Yogyakarta: CAPS (Centre of Academic Publishing Service).
- Walpole, R., Myers, R., Myers, S., & Ye, K. (2012). *Probability & Statistics for Engineers & Scientists Ninth Edition*. United States of America: Prentice Hall.

LAMPIRAN

Lampiran 1 Data Penelitian.

Kecamatan	Tingkat Kasus Campak	Jumlah Imunisasi Campak	Jumlah Puskesmas	Jumlah Tenaga Kesehatan
Sukomanunggal	1	1407	2	68
Tandes	0	1355	2	67
Asemrowo	0	699	1	21
Benowo	0	935	1	36
Pakal	0	805	1	28
Lakarsantri	1	843	3	50
Sambikerep	0	937	2	36
Genteng	0	729	2	40
Tegalsari	1	1236	2	37
Bubutan	0	1307	2	36
Simokerto	0	1186	2	50
Pabean Cantikan	1	1046	1	25
Semampir	1	2731	4	75
Krembangan	0	1597	3	82
Bulak	0	653	1	23
Kenjeran	1	2530	4	95
Tambaksari	1	3031	3	56
Gubeng	1	1706	2	43
Rungkut	1	1588	2	58
Tenggilis	0	817	1	34
Gunung Anyar	1	843	1	35
Sukolilo	1	1562	3	69
Mulyorejo	0	1369	2	53
Sawahan	1	2674	4	100
Wonokromo	0	1918	3	84
Karang Pilang	0	977	1	35
Dukuh Pakis	0	812	1	29
Wiyung	0	965	2	52
Gayungan	0	599	1	25
Wonocolo	0	1142	3	67
Jambangan	0	751	1	38

Lampiran 2 Syntax Regresi Probit Biner.

```
library(caret)
library(readxl)
data1=read_excel("C:/Users/asus/Downloads/data_campak.xlsx")
data=na.omit(data1)
data=data[,-c(1,2,5)]
dim(data)
str(data)

#regresi Probit
gl1= glm(Kode ~data$`Jumlah Imunisasi Campak`+data$`Jumlah
Puskesmas`+data$`Jumlah Tenaga Kesehatan`, family = binomial(link = "probit"), data
= data)
summary(gl1)

#Uji Wald
library(survey)
regTermTest(gl1, ~1)
regTermTest(gl1, ~data$`Jumlah Imunisasi Campak`)
regTermTest(gl1, ~data$`Jumlah Puskesmas`)
regTermTest(gl1, ~data$`Jumlah Tenaga Kesehatan`)

#Uji Serentak
library(pscl)
pR2(gl1)
pchisq(13.5303255, 3, lower.tail = FALSE)

gl2= glm(Kode ~data$`Jumlah Imunisasi Campak`, family = binomial(link = "probit"),
data = data)
summary(gl2)

#P-value uji Deviance (Residual Deviance)
pchisq(gl2$deviance, df=gl2$df.residual, lower.tail=FALSE)
predicted_scores=predict(gl1,data,type="response")
pred <- rep('no',length(predicted_scores))
pred[predicted_scores>=0.5] <- "yes"
table(data$Kode,pred)
b<-as.factor(pred)
a<-as.factor(data$Kode)

levels(b) <- list("0" = "no", "1" = "yes")
confusionMatrix(a,b)
```

Lampiran 3 Syntax Regresi Skewed Probit Biner.

```
set.seed(123)
r = inla(Kode ~ 1 + data$`Jumlah Imunisasi Campak` + data$`Jumlah
Puskesmas` + data$`Jumlah Tenaga Kesehatan`, data = data,
        family = "binomial", control.fixed = list(remove.names = "(Intercept)", prec = 1),
        control.predictor = list(compute = TRUE), control.compute = list(dic = TRUE,
waic = TRUE, return.marginals.predictor = TRUE), control.family = list(control.link =
list(model = "sn", hyper = list(
        skew = list(param = 10))))))
summary(r)

pred.sn = r$summary.fitted.values$mean
pred <- rep(0, length(pred.sn))
pred[pred.sn >= 0.5] <- 1
confusionMatrix(as.factor(data$Kode), as.factor(pred))

# Posterior of coefficient of x1
ggplot(as.data.frame(r$marginals.fixed[[1]])) +
  geom_line(aes(x = x, y = y)) + geom_vline(xintercept = 0.003, linetype =
"dashed", col = "blue", size = 2) +
  ylab("density") + labs(title = "Jumlah Imunisasi Campak") + theme_classic() +
  theme(plot.title = element_text(hjust = 0.5))
# Posterior of coefficient of x2
ggplot(as.data.frame(r$marginals.fixed[[2]])) +
  geom_line(aes(x = x, y = y)) + geom_vline(xintercept = 0.482, linetype =
"dashed", col = "blue", size = 2) +
  ylab("density") + labs(title = "Jumlah Puskesmas") + theme_classic() + theme(plot.title
= element_text(hjust = 0.5))
# Posterior of coefficient of x3
ggplot(as.data.frame(r$marginals.fixed[[3]])) +
  geom_line(aes(x = x, y = y)) + geom_vline(xintercept = -0.047, linetype =
"dashed", col = "blue", size = 2) +
  ylab("density") + labs(title = "Jumlah Tenaga Kesehatan") + theme_classic() +
  theme(plot.title = element_text(hjust = 0.5))

# Posterior of skew
ggplot(as.data.frame(r$marginals.hyperpar[[1]])) +
  geom_line(aes(x = x, y = y)) +
  geom_vline(xintercept = -0.017, linetype = "dashed", col = "blue", size = 1) +
  ylab("density") + labs(title = "Skewness", x = "") + theme_classic() + theme(plot.title =
element_text(hjust = 0.5))
# Posterior of q
ggplot(as.data.frame(r$marginals.hyperpar[[2]])) +
  geom_line(aes(x = x, y = y)) +
  geom_vline(xintercept = 0.025, linetype = "dashed", col = "blue", size = 1) +
  ylab("density") + labs(title = "Quantile Intercept", x = "") + theme_classic() +
  theme(plot.title = element_text(hjust = 0.5))
```