

Model Regresi Gamma untuk Menganalisis Indeks Pengeluaran Kabupaten/Kota di Pulau Sumatra

Bambang Widjanarko Otok^{1*}, Dyah Setyo Rini², Rahmi Fadhilah³

^{1,2,3}Institut Teknologi Sepuluh Nopember, Sukolilo, Surabaya, 60111

^{1,2,3}Departemen Statistika ITS Surabaya Indonesia

e-mail: bambang_wo@its.ac.id

Diajukan: 29 Nopember 2024, Diperbaiki: 17 Maret 2025, Diterima: 18 Maret 2025

Abstrak

Regresi Gamma merupakan salah satu bagian dari *Generalized Linear Models* (GLMs) yang mampu memodelkan data yang bernilai positif dan tidak simetris. Ketidaksimetrisan data seperti ini sering terjadi dalam kehidupan sehari-hari. Ini terlihat pada data Indeks Pembangunan Manusia (IPM). IPM memiliki indikator yang disebut dengan Indeks Dimensi Pembangunan Manusia (IDPM), meliputi Indeks Pengeluaran, Indeks Pendidikan, dan Indeks Umur Harapan Hidup. Penelitian ini bertujuan memodelkan Indeks Pengeluaran kabupaten/kota di Sumatra menggunakan regresi Gamma, karena data Indeks Pengeluaran bersifat positif dan tidak simetris. Pada pemodelan Indeks Pengeluaran, variabel prediktor yang digunakan adalah persentase penduduk miskin, kepadatan penduduk, persentase penduduk yang menggunakan toilet sendiri, dan tingkat pengangguran terbuka pada setiap kabupaten/kota di Sumatra tahun 2023. Data yang digunakan diperoleh melalui *website* BPS Provinsi yang bersesuaian dengan kabupaten/kota di Sumatra. Berdasarkan hasil analisis, semua variabel prediktor yang digunakan memiliki dampak signifikan terhadap Indeks Pengeluaran pada taraf signifikansi 1% dan 5%. Selain itu, setiap penaksir parameter memiliki nilai standar kesalahan. Selain itu, MSE model juga dikategorikan kecil, yaitu 0.00163. Hal tersebut mampu membuktikan bahwa model didukung oleh data walaupun nilai koefisien determinasi model (R^2) pada penelitian ini hanya sebesar 47,59%.

Kata Kunci: Regresi Gamma, MSE, IDPM, Indeks Pengeluaran

Abstract

Gamma regression is part of Generalised Linear Models (GLMs) that can model data that is positive and asymmetric. The occurrence of data asymmetry is common in everyday life, for example in Human Development Index (HDI) data. The HDI has indicators called the Human Development Dimension Index, including the expenditure index, the education index and the life expectancy index. This study aims to model the expenditure index of districts/cities in Sumatra using Gamma regression because the expenditure index data is positive and non-symmetric. In modelling the Expenditure Index, the predictor variables used are the percentage of poor population, population density, percentage of population using their own toilet, and open unemployment rate in each district/city in Sumatra in 2023. The data used were obtained from the BPS website of the province corresponding to the regency/city in Sumatra. Based on the results of the analysis, all the predictor variables used had a significant effect on the expenditure index at the 1% and 5% significance levels, and the standard error value of each parameter estimate was small. In addition, the MSE of the model is also classified as small, which is 0.00163. This can prove that the model is supported by the data, although the coefficient of determination of the model (R^2) in this study is only 47.59%.

Keywords: Gamma regression, MSE, IDPM, Expenditure Index

1 Pendahuluan

Sumatra adalah pulau dengan populasi terbesar kedua di Indonesia. Berdasarkan data Badan Pusat Statistik Sumatera Barat [1], menunjukkan bahwa sekitar 21,68% dari total populasi Indonesia, yang setara dengan 60 juta jiwa, tinggal di pulau ini. Provinsi Sumatra Utara merupakan provinsi dengan populasi tertinggi, mencakup 14,8% dari total penduduk Sumatra. Pemerintahan menghadapi tantangan dalam meningkatkan kualitas sumber daya manusia, yang diukur dengan Indeks Pembangunan Manusia (IPM). IPM menilai pencapaian dalam tiga dimensi utama: pengetahuan, standar hidup yang layak, serta umur panjang dan kesehatan [2], [3].

Berdasarkan data BPS, rata-rata IPM di Pulau Sumatra tahun 2023 dari yang terendah hingga tertinggi adalah Aceh (72,11), Bengkulu (72,45), Lampung (72,89), Sumatra Selatan (73,18), Jambi (73,45), Sumatra Barat (75,64), Riau (75,89), Kepulauan Riau (76,23), dan Sumatra Utara (75,13). Beberapa provinsi menunjukkan peningkatan IPM dibandingkan tahun sebelumnya, namun laju pertumbuhannya bervariasi. Misalnya, Sumatra Barat mengalami peningkatan sebesar 0,48 poin atau 0,64% [1], sementara Sumatra Selatan mencatat kenaikan sebesar 0,70 poin atau 0,97% [4]. Meski secara keseluruhan terjadi tren peningkatan, data IPM tetap menunjukkan pola distribusi yang tidak normal, dengan kecenderungan skewness positif, terutama pada dimensi pendidikan dan pengeluaran [5], [6].

IPM terdiri dari tiga dimensi utama. Dimensi pertama adalah umur panjang dan kehidupan yang sehat, yang diukur melalui indikator Umur Harapan Hidup (UHH). Dimensi kedua mencakup tingkat pendidikan, yang direpresentasikan oleh Harapan Lama Sekolah (HLS) dan Rata-rata Lama Sekolah (RLS). Dimensi ketiga adalah standar hidup layak, yang dinilai berdasarkan pengeluaran per kapita riil yang telah disesuaikan [2]. Karena dimensi standar hidup layak merupakan representasi penting dari kemampuan ekonomi masyarakat, fokus analisis penelitian ini terletak pada dimensi ini. Metode univariat digunakan dalam analisis ini untuk menyelidiki hubungan antara indeks pengeluaran per kapita dan IPM. Penelitian sebelumnya oleh [7] menunjukkan bahwa belanja modal memengaruhi IPM di Indonesia secara signifikan. Namun, penelitian ini mengasumsikan distribusi normal untuk variabel respons melalui penggunaan regresi linear [7], [8]. Ini menjadi salah satu keterbatasan penelitian ini.

Untuk menggambarkan bagaimana variabel dependen dan independen berinteraksi satu sama lain, metode statistik yang umum digunakan adalah model regresi linier. Model ini mengasumsikan variabel dependen berdistribusi normal, hubungan antarvariabel bersifat linear, dan varians data homogen [8]. Namun, asumsi distribusi normal seringkali tidak terpenuhi, terutama pada data seperti IPM yang memiliki distribusi asimetris dengan domain positif. Oleh karena itu, distribusi Gamma menjadi alternatif yang relevan dalam kondisi ini. Distribusi Gamma

adalah distribusi kontinu yang digunakan untuk memodelkan data positif dengan karakteristik skewness positif. Distribusi ini memiliki parameter bentuk (*shape*) dan parameter skala (*scale*), sehingga memberikan fleksibilitas dalam menangkap pola distribusi data yang tidak simetris [5], [9], [10]. Regresi Gamma juga merupakan salah satu jenis *Generalized Linear Models* (GLMs). Istilah ini digunakan untuk menggambarkan pemodelan regresi dengan variabel respon berdistribusi Gamma [11].

Dalam penelitian ini, regresi Gamma digunakan untuk memodelkan hubungan antara dimensi IPM, yaitu indeks pengeluaran. Penggunaan regresi Gamma didasarkan pada karakteristik data IPM di Pulau Sumatra yang menunjukkan pola distribusi asimetris dengan kemencengan positif (*positive skewness*) pada indeks pengeluaran dan pendidikan, serta kemencengan negatif (*negative skewness*) pada indeks UHH. Diharapkan bahwa penelitian ini akan menghasilkan hasil yang lebih akurat dan representatif jika dibandingkan dengan metode regresi linear yang telah digunakan dalam penelitian sebelumnya. Metode regresi Gamma digunakan dalam penelitian ini [8], [12]. Penelitian ini juga memiliki keunikan dalam pendekatan metodologi, karena belum banyak studi yang mengaplikasikan regresi Gamma pada analisis Indeks Pengeluaran, khususnya di Pulau Sumatra [13], [14].

Penelitian ini bertujuan untuk menemukan komponen yang memengaruhi Indeks Pengeluaran Pulau Sumatra dengan menggunakan metode regresi Gamma. Selain itu, penelitian ini mencakup pengembangan algoritma dan komputasi untuk pemodelan regresi Gamma untuk menawarkan solusi yang lebih adaptif untuk distribusi data yang tidak biasa [9], [15].

2 Metode Penelitian

Data sekunder yang digunakan dalam penelitian ini berasal dari 155 kabupaten/kota di Sumatra dari Badan Pusat Statistik (BPS). Variabel yang dianalisis mencakup variabel respon, yaitu Indeks Pengeluaran kabupaten/kota di Sumatra tahun 2023 (Y), dan variabel prediktor, yaitu persentase penduduk miskin kabupaten/kota di Sumatra tahun 2023 (X_1), kepadatan penduduk kabupaten/kota di Sumatra tahun 2023 (X_2), persentase penduduk kabupaten/kota di Sumatra yang menggunakan toilet sendiri tahun 2023 (X_3), dan tingkat pengangguran terbuka kabupaten/kota di Sumatra tahun 2023 (X_4).

2.1 Penaksiran Parameter pada Model Regresi Gamma

Variabel random Y berdistribusi Gamma dua parameter, yaitu *shape parameter* (ω) dan *scale parameter* (θ), $Y \sim \text{Gamma}(\omega, \theta)$, memiliki fungsi densitas peluang sebagai berikut [16]:

$$f(y) = \begin{cases} \frac{1}{\theta^\omega \Gamma(\omega)} y^{\omega-1} \exp\left(-\frac{y}{\theta}\right), & \omega, \theta > 0, 0 < y < \infty \\ 0 & , \text{untuk } y \text{ lainnya} \end{cases} \quad (1)$$

dengan $\Gamma(\omega)$ merupakan fungsi Gamma, $\Gamma(\omega) = \int_0^\infty t^{\omega-1} e^{-t} dt$ dan $\Gamma(\omega) = (\omega-1)!$.

Jika $Y \sim \text{Gamma}(\omega, \theta)$, maka sifat-sifat dari Y adalah sebagai berikut [17]:

- Ekspektasi, $E(Y) = \omega\theta$
- Varian, $\text{Var}(Y) = \omega\theta^2$
- Fungsi pembangkit momen, $M_Y(t) = (1 - \theta t)^{-\omega}$, $t < \frac{1}{\theta}$
- Skewness, $\theta_1 = \frac{2}{\sqrt{\omega}}$
- Kurtosis, $\theta_2 = 3 + \frac{6}{\omega}$

Regresi Gamma dipilih karena variabel respon berdistribusi Gamma, bersifat kontinu non-negatif, dan memiliki koefisien variasi yang konstan. Model *log-linear* digunakan untuk memastikan ekspektasi positif dari variabel respon [18]. Jika variabel yang mempengaruhi variabel respon Y adalah X_1, X_2, \dots, X_p , maka ekspektasi dari Y pada pengamatan ke- i berdasarkan *log-linear model* adalah sebagai berikut [11].

$$\begin{aligned} g(\mu_i) &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} \\ \mu_i &= \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}) \end{aligned} \quad (2)$$

dengan $\mu_i = E(Y_i)$ dan $g(\mu_i)$ adalah fungsi transformasi.

Karena $\mu_i = E(Y_i)$, maka diperoleh:

$$\begin{aligned} \mu_i &= \omega\theta = \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}) \\ \theta &= \frac{\exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip})}{\omega} \\ &= \frac{\exp(\mathbf{x}^T \boldsymbol{\beta})}{\omega} \end{aligned} \quad (3)$$

dengan $\mathbf{x} = [1 \ x_1 \ x_2 \ \dots \ x_p]^T$ dan $\boldsymbol{\beta} = [\beta_0 \ \beta_1 \ \beta_2 \ \dots \ \beta_p]^T$.

Dengan mensubstitusi rumus θ pada Persamaan (3) ke Persamaan (1), diperoleh fungsi kepadatan peluang untuk y_i adalah sebagai berikut:

$$f(y_i) = \begin{cases} \frac{1}{\left(\frac{\exp(\mathbf{x}_i^T \boldsymbol{\beta})}{\omega}\right)^\omega} \frac{y_i^{\omega-1} \exp\left(-\frac{y_i}{\left(\frac{\exp(\mathbf{x}_i^T \boldsymbol{\beta})}{\omega}\right)}\right)}{\Gamma(\omega)}, & \omega > 0, y_i > 0, \boldsymbol{\beta} \in \square \\ 0, & \text{untuk } y_i \text{ lainnya} \end{cases} \quad (4)$$

dengan $\mathbf{x}_i = [1 \quad x_{i1} \quad x_{i2} \quad \cdots \quad x_{ip}]^T$.

Berdasarkan Persamaan (4), parameter regresi yang ditaksir adalah $\boldsymbol{\beta}$ dan ω . Kedua parameter ditaksir dengan metode kemungkinan maksimum, yang dikenal sebagai estimasi kemungkinan maksimum atau *Maximum Likelihood Estimation* (MLE). Metode ini merupakan suatu metode penaksir parameter dari suatu model dengan cara memaksimalkan *likelihood function* [19]. *Likelihood function* dari Persamaan (4) dituliskan sebagai berikut:

$$L(\omega, \boldsymbol{\beta}) = \prod_{i=1}^n f(y_i) = \left[\left(\frac{1}{\Gamma(\omega)}\right)^n \left(\frac{\exp\left(\sum_{i=1}^n \mathbf{x}_i^T \boldsymbol{\beta}\right)}{\omega}\right)^{-\omega} \left(\prod_{i=1}^n (y_i^{\omega-1})\right) \exp\left(\sum_{i=1}^n \frac{-\omega y_i}{\exp(\mathbf{x}_i^T \boldsymbol{\beta})}\right) \right] \quad (5)$$

dan logaritma natural dari *likelihood function* adalah:

$$\ln L(\omega, \boldsymbol{\beta}) = -n \ln \Gamma(\omega) - \omega \sum_{i=1}^n \mathbf{x}_i^T \boldsymbol{\beta} + \omega n \ln \omega + (\omega - 1) \sum_{i=1}^n \ln y_i + \sum_{i=1}^n \frac{-\omega y_i}{\exp(\mathbf{x}_i^T \boldsymbol{\beta})} \quad (6)$$

Oleh karena memaksimalkan *likelihood function* sama dengan memaksimalkan *ln likelihood function*, maka langkah selanjutnya menentukan derivatif parsial pertama dari Persamaan (6) terhadap parameter ω dan terhadap parameter $\boldsymbol{\beta}$, kemudian hasil derivatif parsial pertama tersebut disamadengankan nol. Derivatif parsial pertama *ln likelihood function* terhadap ω adalah sebagai berikut:

$$\frac{\partial \ln L(\omega, \boldsymbol{\beta})}{\partial \omega} = -n \frac{\Gamma'(\omega)}{\Gamma(\omega)} - \sum_{i=1}^n \mathbf{x}_i^T \boldsymbol{\beta} + n \ln(\omega) + n + \sum_{i=1}^n \frac{-y_i}{\exp(\mathbf{x}_i^T \boldsymbol{\beta})} = 0 \quad (7)$$

sedangkan derivatif parsial pertama dari *likelihood function* terhadap $\boldsymbol{\beta}$ adalah:

$$\frac{\partial \ln L(\omega, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -\omega \sum_{i=1}^n \mathbf{x}_i^T + \omega \sum_{i=1}^n \frac{\mathbf{x}_i^T y_i}{\exp(\mathbf{x}_i^T \boldsymbol{\beta})} = 0 \quad (8)$$

Karena Persamaan (7) dan (8) memiliki bentuk yang tidak tertutup, penyelesaian kedua persamaan tersebut memerlukan penggunaan metode numerik. Algoritma *Fisher Scoring* adalah

salah satu metode numerik yang dapat digunakan. Algoritma ini digunakan untuk regresi gamma di bawah populasi sebagai berikut:

- 1) Menentukan nilai awal $\boldsymbol{\gamma}_0 = [\beta_{0(0)} \quad \beta_{1(0)} \quad \cdots \quad \beta_{p(0)} \quad \omega_0]^T$ dan menentukan toleransi *error* (ε), dengan $\boldsymbol{\beta}_{(0)}^T = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$ dan $\alpha_{(0)}$ adalah sembarang bilangan positif.
- 2) Menghitung $\mathbf{V}(\boldsymbol{\gamma}_m)$ dan $\mathbf{H}(\boldsymbol{\gamma}_m)$.

$$\mathbf{V}(\boldsymbol{\gamma}_m) = \begin{bmatrix} \frac{\partial \ln L(\omega, \boldsymbol{\beta} | y_1, y_2, \dots, y_n)}{\partial \boldsymbol{\beta}} \\ \frac{\partial \ln L(\omega, \boldsymbol{\beta} | y_1, y_2, \dots, y_n)}{\partial \omega} \end{bmatrix}$$

$$\mathbf{H}(\boldsymbol{\gamma}_m) = \begin{bmatrix} E\left(\frac{\partial^2 \ln L(\omega, \boldsymbol{\beta} | y_1, y_2, \dots, y_n)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T}\right) & E\left(\frac{\partial^2 \ln L(\omega, \boldsymbol{\beta} | y_1, y_2, \dots, y_n)}{\partial \boldsymbol{\beta}^T \partial \omega}\right) \\ E\left(\frac{\partial^2 \ln L(\omega, \boldsymbol{\beta} | y_1, y_2, \dots, y_n)}{\partial \omega \partial \boldsymbol{\beta}^T}\right) & E\left(\frac{\partial^2 \ln L(\omega, \boldsymbol{\beta} | y_1, y_2, \dots, y_n)}{\partial \omega^2}\right) \end{bmatrix}$$

- 3) Menghitung $\boldsymbol{\gamma}_{m+1} = \boldsymbol{\gamma}_m + \mathbf{H}^{-1}(\boldsymbol{\gamma}_m) \mathbf{V}(\boldsymbol{\gamma}_m)$, $m = 0, 1, 2, \dots$
- 4) Menghitung $\|\boldsymbol{\gamma}_{m+1} - \boldsymbol{\gamma}_m\|$, iterasi akan konvergen jika $\|\boldsymbol{\gamma}_{m+1} - \boldsymbol{\gamma}_m\| < \varepsilon$.
- 5) Ulangi Langkah (3) dan seterusnya dengan $m = m + 1$ hingga didapatkan $\|\boldsymbol{\gamma}_{m+1} - \boldsymbol{\gamma}_m\| < \varepsilon$.

Pengujian Signifikansi Parameter

Pada regresi Gamma, pengujian signifikansi parameter termasuk pengujian parameter baik secara serentak maupun secara parsial. Pengujian serentak menggunakan uji rasio kemungkinan maksimum (MLRT), sedangkan pengujian parsial menggunakan uji-Z.

Untuk pengujian parameter yang dilakukan secara bersamaan, rumusan hipotesis, statistik uji, dan kriteria penolakan hipotesis adalah sebagai berikut:

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

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

Statistik uji:

$$G^2 = -2 \ln \Lambda = -2 \ln L(\varpi) + 2 \ln L(\Omega) \quad (9)$$

dengan $\varpi = \{\beta_0, \omega_0\}$ merupakan himpunan parameter dibawah hipotesis nol dan $\Omega = \{\boldsymbol{\beta}, \omega\}$ merupakan himpunan parameter dibawah populasi.

Kriteria penolakan hipotesis:

Tolak H_0 pada taraf signifikansi α , jika $G^2 > \chi_{\alpha, p}^2$, p merupakan banyaknya variabel prediktor.

Sedangkan rumusan hipotesis, statistik uji, dan kriteria penolakan hipotesis pada pengujian parameter secara parsial adalah:

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0, j = 1, 2, \dots, p$$

Statistik uji:

$$Z = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \quad (10)$$

dengan $SE(\hat{\beta}_j) = \sqrt{Var(\hat{\beta}_j)}$.

$Var(\hat{\beta}_j)$ adalah entri-entri pada diagonal utama ke- $j+1$ dari matriks $-\mathbf{H}^{-1}(\hat{\gamma})$.

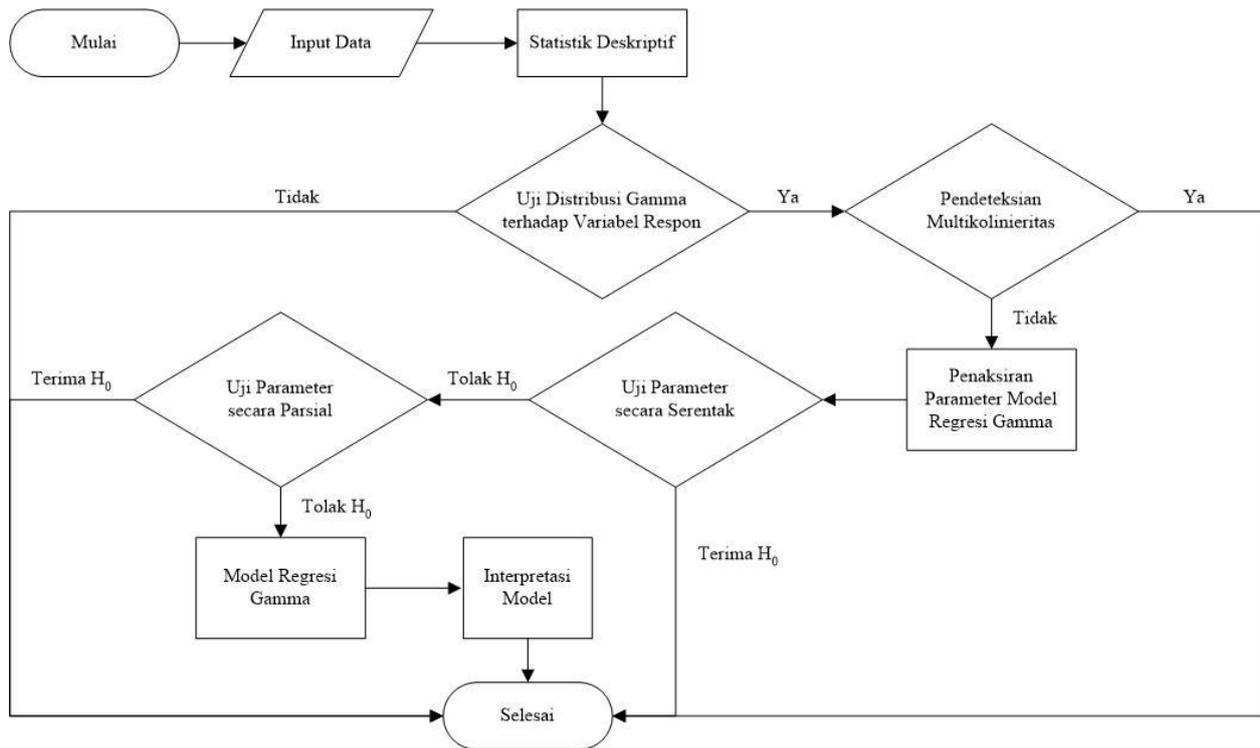
Kriteria penolakan hipotesis:

Tolak H_0 pada taraf signifikansi α , jika $|Z_{hitung}| > Z_{\alpha/2}$ [20].

2.2 Tahapan Analisis Data

Adapun tahapan analisis data pada penelitian ini adalah:

- 1) *Descriptive statistics* terhadap variabel respon dan variabel prediktor.
- 2) Pengujian *goodness of fit* terhadap variabel respon menggunakan uji Kolmogorov-Smirnov.
- 3) Pendeteksian multikolinieritas pada variabel prediktor menggunakan *Variance Inflation Factor* (VIF).
- 4) Pemodelan regresi Gamma, meliputi:
 - a. Penaksiran parameter menggunakan metode MLE dan algoritma *Fisher's Scoring*.
 - b. Pengujian parameter secara serentak menggunakan statistik uji G^2 seperti pada Persamaan (9).
 - c. Pengujian parameter secara parsial menggunakan statistik uji Z seperti pada Persamaan (10)
- 6) Interpretasi model.



Gambar 1. Flowchart Analisis Data

3 Hasil dan Pembahasan

3.1 Descriptive Statistics

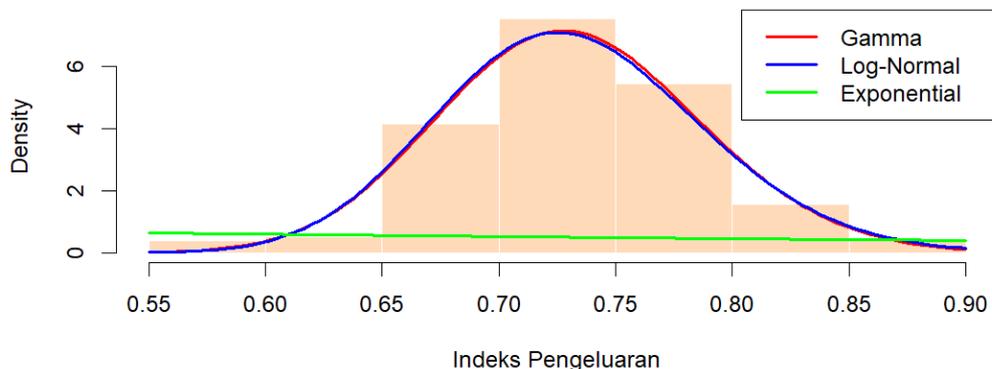
Tabel 1 menampilkan deskriptif statistik untuk data variabel respon dan prediktor.

Tabel 1. Descriptive Statistics Variabel Penelitian

Variabel	Min	Median	Max	Rata-rata	Varian	Std. Deviasi
Y	0,56	0,72	0,89	0,73	0,003	0,055
X1	2,27	9,11	22,98	10,01	20,25	4,49
X2	1,09	111,50	8859,00	599,25	1901841,1	1379,07
X3	51,03	90,80	99,49	87,70	74,42	8,62
X4	0,80	4,68	10,86	4,61	3,46	1,86

Indeks Pengeluaran terbagi menjadi tiga kategori: kategori rendah mencakup nilai 0,00–0,59, kategori sedang mencakup nilai 0,60–0,69, kategori tinggi mencakup nilai 0,70–0,79, dan kategori sangat tinggi mencakup nilai 0,80–1,00. Berdasarkan Tabel 1, terlihat bahwa rata-rata Indeks Pengeluaran di Sumatra termasuk dalam kategori tinggi, yaitu 0,7317. Nilai varian dari data Indeks Pengeluaran mengindikasikan bahwa data cukup beragam, dimana nilai minimum dari data berada pada kategori rendah (0,5641) dan nilai maksimum data termasuk dalam kategori sangat tinggi (0,8973).

Gambar 2 menyajikan histogram dan plot fungsi densitas peluang dari tiga distribusi terhadap variabel respon, yaitu distribusi Gamma, distribusi Log-Normal, dan distribusi Eksponensial. Grafik yang dihasilkan oleh program R adalah sebagai berikut:



Gambar 2. Histogram dan Plot Densitas dari Variabel Indeks Pengeluaran

Berdasarkan Gambar 1, distribusi Gamma memberikan model yang tampaknya paling mendekati histogram data Indeks Pengeluaran.

3.2 Pengujian *Goodness of Fit* terhadap Variabel Respon

Salah satu statistik uji yang dapat diandalkan untuk menguji distribusi data adalah Anderson-Darling. Uji ini digunakan untuk menentukan distribusi data [21]. Dalam penelitian ini, hipotesis berikut dibuat:

$$H_0 : F_Y = F_Y^0 \text{ (distribusi data mengikuti distribusi tertentu)}$$

$$H_1 : F_Y \neq F_Y^0 \text{ (distribusi data tidak mengikuti distribusi tertentu)}$$

Distribusi eksponensial, log-normal, dan gamma diuji pada variabel respons (Indeks Pengeluaran) dalam penelitian ini. Tabel 2 menguraikan hasil perhitungan statistik uji Anderson-Darling menggunakan program R:

Tabel 2. Uji *Goodness of Fit* pada Variabel Respon

Distribusi	Statistik Uji	<i>P-value</i>
Eksponensial	60,774	<0,003
Log Normal	0,809	0,036
Gamma	0,7	0,072

Karena nilai *p-value* (0,072) yang dihasilkan lebih besar dari taraf nyata pengujian 5%, distribusi Gamma adalah opsi terbaik untuk data Indeks Pengeluaran pada taraf signifikansi 5%, seperti yang ditunjukkan dalam Tabel 2. Oleh karena itu, langkah selanjutnya adalah memodelkan data menggunakan regresi Gamma.

3.3 Pendeteksian Multikolinieritas pada Variabel Prediktor

Asumsi penting dalam pemodelan regresi adalah asumsi non-multikolinieritas pada variabel prediktor. Dalam penelitian ini, *Variance Inflation Factor* (VIF) digunakan untuk mendeteksi multikolinieritas dari variabel prediktor penelitian [20]. Nilai VIF dihitung menggunakan program R dan disajikan pada Tabel 3 sebagai berikut:

Tabel 3. Nilai VIF Variabel Prediktor

Prediktor	X1	X2	X3	X4
VIF	1,10936	1,2302	1,10705	1,20911

Berdasarkan hasil pada Tabel 3, diperoleh bahwa nilai VIF dari masing-masing variabel prediktor lebih kecil dari 10. Hal ini mengindikasikan bahwa tidak terdapat korelasi antar variabel prediktor. Dengan kata lain, asumsi non-multikolinieritas terpenuhi untuk data penelitian ini.

3.4 Pemodelan Regresi Gamma

Pemodelan menggunakan regresi Gamma didasari oleh hasil pengujian distribusi dari variabel respon, yaitu distribusi Gamma. Tahapan pada bagian ini terdiri dari penaksiran dan pengujian parameter model. Kedua tahapan tersebut menggunakan *Gammareg Package* yang ada pada program R [22].

Karena bentuk turunan parsial pertama untuk setiap parameter tidak dapat dinyatakan dalam bentuk tertutup, metode iteratif digunakan untuk menyelesaikan proses penaksiran parameter dalam pemodelan regresi Gamma. Hasil taksiran parameter dan uji parsial disajikan pada Tabel 4, sementara hasil uji kesesuaian model dan pengujian parameter secara bersamaan disajikan pada Tabel 5.

Tabel 4. Taksiran Parameter Model Regresi Gamma

Parameter	Estimasi	Std. Error	Z hitung	P-Value	Odd
β_0	-0,4642	0,0547	-8,4900	0,0000***	0,628638
β_1	-0,0075	0,0011	-6,9970	0,0000***	0,992528
β_2	0,0000	0,0000	2,9620	0,0036**	1
β_3	0,0021	0,0006	3,8420	0,0002***	1,002102
β_4	0,0067	0,0027	2,4740	0,0145*	1,006722

Keterangan: (***) signifikan pada taraf nyata 1% ($p - value \leq 0,01$), (**) signifikan pada taraf nyata 5% ($p - value \leq 0,05$), (*) signifikan pada taraf nyata 10% ($p - value \leq 0,10$).

Tabel 5. Statistik Uji Model Regresi Gamma

Statistik Uji	Output
<i>Null Deviance</i>	0,905368
<i>Residual Deviance</i>	0,478874
G ²	0,426494
<i>P-value</i>	0,980251
AIC	-537,071
MSE	0,00163
R ²	0,475965
<i>Adjusted-R²</i>	0,461897

Hasil taksiran parameter di Tabel 4 menunjukkan nilai error standar penaksir yang rendah, serta *p-value* yang lebih kecil dari taraf nyata pengujian 1% dan 5% untuk masing-masing variabel. Artinya, setiap variabel prediktor memiliki dampak yang signifikan terhadap model secara parsial. Nilai *Null Deviance* dan *Mean Square Error* (MSE) sangat kecil, seperti yang ditunjukkan dalam Tabel 5. Ini menunjukkan bahwa data mendukung model; nilai menunjukkan bahwa 47,59% dari keragaman variabel respons dapat dijelaskan oleh variabel prediktor yang digunakan dalam penelitian ini, sedangkan 52,41% lainnya dapat dijelaskan oleh variabel tambahan yang tidak ada dalam model.

3.5 Interpretasi Model

Berdasarkan hasil analisis, model regresi Gamma yang dihasilkan adalah sebagai berikut:

$$\hat{y} = \exp(-0,4642 - 0,0075x_1 + 0,0000x_2 + 0,0021x_3 + 0,0067x_4) \quad (11)$$

Menurut model regresi Gamma yang ditemukan pada Persamaan (11), jika variabel persentase penduduk miskin kabupaten/kota di Sumatra meningkat sebesar 1% dan variabel lain dianggap konstan, Indeks Pengeluaran kabupaten/kota di Sumatra akan menurun sebesar $\exp(-0,0075)$ kali atau sebesar 0,99 kali. Sebaliknya, jika variabel kepadatan penduduk kabupaten/kota di Sumatra meningkat sebesar 1% dan variabel lain dianggap konstan, Indeks Pengeluaran kabupaten/kota di Sumatra akan meningkat sebesar 1,006 kali jika variabel tingkat pengangguran terbuka meningkat 1% dan variabel lain dianggap konstan.

4 Simpulan

Berdasarkan hasil dan pembahasan, Indeks Pengeluaran yang merupakan salah satu indeks dimensi pembangunan manusia sangat cocok sangat cocok dimodelkan dengan regresi Gamma karena ketidaksimetrisan data. Selain itu, persentase penduduk miskin, kepadatan penduduk, persentase penduduk yang menggunakan toilet sendiri, dan tingkat pengangguran terbuka, memengaruhi Indeks Pengeluaran secara signifikan atau secara tidak langsung memengaruhi IPM.

Dikarenakan nilai koefisien determinasi (R^2) pada penelitian ini masih kurang dari 50%, penelitian selanjutnya dapat menggunakan variabel prediktor lainnya yang mampu menjelaskan keragaman Indeks Pengeluaran lebih baik lagi.

6 Daftar Pustaka

- [1] Badan Pusat Statistik Sumatera Barat, “Indeks Pembangunan Manusia Sumatera Barat Tahun 2023.” Accessed: Nov. 25, 2024. [Online]. Available: <https://sumbar.bps.go.id>
- [2] Badan Pusat Statistik, “Tanya Jawab Seputar Indeks Pembangunan Manusia (IPM).” Accessed: Nov. 25, 2024. [Online]. Available: <https://pamekasankab.bps.go.id>
- [3] K. I. Subramanian SV, “Income inequality and health: what have we learned so far?,” *Epidemiologi Review*, vol. 26, no. 1, pp. 78–91, 2004.
- [4] Badan Pusat Statistik (BPS) Sumatera Selatan, “Indeks Pembangunan Manusia Sumatera Selatan Tahun 2023. BPS Sumsel,” 2023. Accessed: Nov. 25, 2024. [Online]. Available: <https://sumsel.bps.go.id>
- [5] B. Jørgensen, “Exponential Dispersion Models,” *Journal of the Royal Statistical Society*, vol. 49, no. 2, pp. 127–162, 1987.
- [6] M. Corrales and E. Cepeda-Cuervo, “A Bayesian Approach to Mixed Gamma Regression Models,” *Rev Colomb Estad*, vol. 42, pp. 81–99, Mar. 2019, doi: 10.15446/rce.v42n1.69334.
- [7] D. Retnowati and Z. Zumaeroh, “Faktor-Faktor Yang Mempengaruhi Indeks Pembangunan Manusia di Indonesia,” *Jurnal Nusantara*, vol. 11, no. 3, pp. 45–57, 2023.
- [8] P. McCullagh and J. A. Nelder, *Generalized Linear Models (2nd edition)*. New York: Chapman & Hall, 1989. doi: <https://doi.org/10.1201/9780203753736>.
- [9] M. L. Corrales and E. C. Cuervo, “A Bayesian Approach to Mixed Gamma Regression Models,” *Colombian Journal of Statistics/Revista Colombiana de Estadística*, vol. 42, no. 1, 2019.
- [10] M. Schmid and T. Hothorn, “Flexible Boosting of Accelerated Failure Time Models,” *BMC Bioinformatics*, vol. 9, no. 1, p. 269, 2008, doi: <https://doi.org/10.1186/1471-2105-9-269>.
- [11] A. J. Dobson and A. G. Barnett, *An Introduction to Generalized Linear Models Fourth Edition*, 4th ed. 2018. doi: <https://doi.org/10.1201/9781315182780>.
- [12] W. H. Greene, *Econometric Analysis (7th ed.)*. Boston: Pearson Education, 2012.
- [13] A. Widodo, W. Waridin, and J. M. Kodoatie, “Analisis Pengaruh Pengeluaran Pemerintah di Sektor Pendidikan dan Kesehatan terhadap Pengentasan Kemiskinan Melalui

-
- Peningkatan Pembangunan Manusia di Provinsi Jawa Tengah,” *Jurnal Dinamika ekonomi pembangunan*, vol. 1, no. 1, pp. 25–42, 2011.
- [14] T. Tambunan, “Economic Growth, Appropriate Policies and Poverty Reduction in a Developing Country: Some Experience from Indonesia,” *South Asia Economic Journal*, vol. 6, no. 1, pp. 59–78, 2005, doi: 10.1177/139156140500600104.
- [15] P. Hall, *The Jackknife, the Bootstrap and Other Resampling Plans*. New York, NY: Springer New York, 2008. doi: 10.1007/978-0-387-75692-9_11.
- [16] B. Bhattacharya, “Tests of Parameters of Several Gamma Distributions with Inequality Restrictions,” *Ann Inst Stat Math*, vol. 54, pp. 565–576, 2002, doi: <https://doi.org/10.1023/A:1022411127154>.
- [17] M. Lavine, *Introduction to Statistical Thought*. JASA, 2009.
- [18] P. McCullagh, “Quasi-Likelihood Functions,” *the annals of statistics*, vol. 11, no. 1, pp. 59–67, 1983, doi: DOI: 10.1214/aos/1176346056.
- [19] D. Roxbee, *Theoretical Statistics*, 1 st Edition. New York: Chapman and Hall, 1974. doi: <https://doi.org/10.1201/b14832>.
- [20] H. Yasin, S. Inayati, and S. Setiawan, “3-PARAMETER GAMMA REGRESSION MODEL FOR ANALYZING HUMAN DEVELOPMENT INDEX OF CENTRAL JAVA PROVINCE,” *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 16, no. 1, pp. 171–180, Mar. 2022, doi: 10.30598/barekengvol16iss1pp171-180.
- [21] D. E. Putri, Purhadi, and D. D. Prastyo, “Parameter estimation and hypothesis testing on geographically weighted gamma regression,” in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Oct. 2017. doi: 10.1088/1742-6596/893/1/012025.
- [22] M. F. Zarate, R. Duplat, C. Elias, P. Maintainer, and M. Corrales, “Package ‘Gammareg’ Type Package Title Classic Gamma Regression: Joint Modeling of Mean and Shape Parameters License GPL (≥ 2),” 2023. [Online]. Available: <http://www.docentes.unal.edu.co/ecepedac/docs/MODELAGEM20DA20VARIABILIDADE.pdf>.