

PEMODELAN REGRESI GAMMA MENGGUNAKAN METODE OPTIMASI *BROYDEN-FLETCHER-GOLDFARB-SHANNO* (BFGS) (Studi Kasus: Pencemaran Sungai di Kota Semarang)

Efifah Nur Safitri^{1*}, Arief Rachman Hakim², Suparti³

^{1,2,3} Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro

*e-mail: ervifa563@gmail.com

DOI: 10.14710/j.gauss.14.2.245-256

Article Info:

Received: 2024-07-14

Accepted: 2025-09-12

Available Online: 2025-09-15

Keywords:

River Pollution; Gamma

Regression; COD; MLE; BFGS.

Abstract: Semarang City is one of the major industrial areas in Central Java and is located not far from residential areas. These large industries utilize a variety of chemicals to meet their needs, producing various wastes that are the main cause of high concentrations of Chemical Oxygen Demand (COD) in waters. Measurements on the COD value obtained are continuous data with Gamma distribution. In this study, Gamma regression is used to model the relationship between one or more predictor variables and a positive continuous response variable following a Gamma distribution. Parameter estimation in the Gamma regression model uses the Maximum Likelihood Estimation (MLE) method because it does not produce an analytical solution, an optimization method will be carried out with the Broyden Fletcher Goldfarb Shanno (BFGS) algorithm. This study aims to determine the Gamma regression model in the case of river pollution in Semarang City, determine the form of parameter estimation in the Gamma regression model and get what factors affect river pollution in Semarang City. Based on the R^2 value of 0.4431498 where the ability of the predictor variables to explain the response variable is 44.32%, the remaining 55.68% of the response variable is explained by other factors not contained in the model.

1. PENDAHULUAN

Pencemaran sungai yang ada di Kota Semarang terjadi karena terletak di kawasan industri besar di Jawa Tengah yang memanfaatkan bahan kimia sehingga menghasilkan beraneka macam limbah. Limbah industri, limbah rumah tangga dan limbah peternakan yang menjadi sumber utama limbah organik dan menyebabkan tingginya konsentrasi COD (Lumaela dkk, 2013). Apabila konsentrasi COD pada sungai melebihi ukuran batas standar baku mutu air dapat disimpulkan bahwa sungai tersebut masuk dalam kategori tercemar. Pada penelitian ini akan digunakan indikator COD sebagai variabel respon dan variabel prediktor yang berpengaruh yaitu debit, temperatur, TSS, pH, BOD, dan DO.

Analisis regresi diterapkan untuk memodelkan hubungan antara variabel respon dengan variabel prediktor. Regresi Gamma digunakan untuk pemodelan hubungan antara satu atau lebih variabel prediktor dengan variabel respon kontinu positif yang berdistribusi Gamma. Pada penelitian ini pengukuran pada nilai COD yang didapat merupakan data kontinu yang berdistribusi Gamma. Regresi Gamma menghasilkan parameter yang bersifat global yang berarti parameter tersebut dianggap valid di lokasi mana data itu diambil. *Maximum Likelihood Estimation* (MLE) digunakan untuk mengestimasi parameter pada regresi Gamma, karena tidak mendapatkan solusi secara analitik dilakukan metode optimasi dengan algoritma *Broyden-Fletcher-Goldfarb-Shanno* (BFGS). Beberapa penelitian terkait regresi Gamma yang sudah dilakukan sebelumnya antara lain Luthfiana (2021) yang menerapkan pemodelan regresi Gamma dengan metode optimasi *Newton Raphson*, Faizah (2018) yang membandingkan model regresi Gamma

dengan regresi Lognormal serta Nasution (2016) yang menerapkan regresi Gamma pada pencemaran sungai di Kota Surabaya dengan metode optimasi *Newton Raphson*.

Berdasarkan latar belakang tersebut, rumusan masalah dalam penelitian ini adalah menentukan pemodelan regresi Gamma pada kasus pencemaran sungai Kota Semarang. Selain itu, menentukan estimasi parameter model regresi Gamma dan faktor-faktor yang mempengaruhi kasus pencemaran Sungai di Kota Semarang.

2. TINJAUAN PUSTAKA

Pada Peraturan Pemerintah No 82 Tahun 2001 mengenai Pengelolaan Kualitas Air dan Pengendalian Pencemaran Air pada pasal 1 angka 11 dijelaskan bahwa pencemaran air disebabkan karena adanya suatu zat, energi, makhluk hidup maupun komponen lain yang bersumber dari aktivitas manusia yang menyebabkan penurunan kualitas pada air sehingga air tidak bisa digunakan sesuai dengan kegunaannya. Menurut Putri & Prastyo (2016) indikator pada pencemaran air merupakan suatu unsur yang menandai jika terjadi perubahan kualitas air atau dapat dikatakan bahwa air tersebut telah tercemar. Perubahan kualitas air tersebut secara umum dapat diamati dari 3 parameter yaitu parameter kimia, parameter fisika dan parameter biologi. Adapun indikator pencemaran air yang umum diketahui seperti cod, debit, temperatur, TSS, pH, BOD, dan DO. Distribusi Gamma merupakan distribusi kontinu yang memiliki dua parameter α dan θ yang berpengaruh terhadap bentuk distribusinya. Dikatakan berdistribusi Gamma apabila suatu peubah acak Y pada parameter α dan θ bernilai positif, bentuk fungsi densitas peluang (fdp) dari peubah acak kontinu Y dengan dua parameter dinyatakan dengan:

$$f(y; \alpha, \theta) = \begin{cases} \frac{1}{\theta^\alpha \Gamma(\alpha)} y^{\alpha-1} e^{-\frac{y}{\theta}}; & 0 < y < \infty \\ 0; & y \text{ lainnya} \end{cases} \quad (1)$$

Dengan $\Gamma(\alpha)$ adalah fungsi Gamma yang memiliki nilai positif. Fungsi Gamma dengan parameter α dan θ sebagai berikut:

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy, \text{ untuk } \alpha > 0 \quad (2)$$

Bentuk fungsi densitas kumulatif (fdk) pada distribusi Gamma yaitu sebagai berikut:

$$f(y; \alpha, \theta) = \int_0^y \frac{1}{\theta^\alpha \Gamma(\alpha)} t^{\alpha-1} e^{-\frac{t}{\theta}} dt \quad (3)$$

Menurut Casella dan Berger (2002), nilai harapan (μ) dan varians pada distribusi Gamma (α, θ) adalah sebagai berikut:

Nilai harapan pada distribusi Gamma:

$$\mu = E(Y) = \alpha\theta \quad (4)$$

Nilai varians pada distribusi Gamma:

$$Var(Y) = \alpha\theta^2 \quad (5)$$

Menurut Davidian (2005), asumsi dasar bahwa koefisien variasi pada distribusi Gamma bernilai konstan. Koefisien variasi dari distribusi Gamma dinyatakan dengan:

$$CV (\text{Coefficient of Variation}) = \frac{1}{\sqrt{\alpha}} \quad (6)$$

Anderson Darling yaitu metode dalam statistika yang berguna untuk menguji kesesuaian distribusi pada peubah respon yang mengandung variabel kontinu. Dengan hipotesis:

H_0 : variabel respon sesuai dengan distribusi tertentu

H_1 : variabel respon tidak sesuai dengan distribusi tertentu

Statistik Uji:

$$A = -n - \frac{1}{n} \sum_{i=1}^n (2i-1) (\ln F_{(y_i)}^0 + \ln(1 - F_{(y_{n+1-i})}^0)) \quad (7)$$

dengan:

A = Anderson Darling

$F_{(y_i)}^0$ = fungsi distribusi kumulatif dari distribusi tertentu

n = ukuran sampel

Kriteria Keputusan:

Tolak H_0 jika $A_{hitung} > A_p$.

Uji non heteroskedastisitas merupakan uji asumsi umum yang digunakan pada pengujian ketidaksamaan variansi antara residual observasi dalam model regresi. (Ghozali, 2018). Pada penelitian ini dilakukan pengujian heteroskedastisitas dengan *Breusch-Pagan* (BP test) dengan hipotesis:

Hipotesis:

$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 = \sigma^2$ (tidak terdapat heteroskedastisitas)

H_1 : paling sedikit ada satu i dengan $\sigma_i^2 \neq \sigma^2$ (terdapat heteroskedastisitas)

Diasumsikan σ_i^2 atau variansi erornya sebagai berikut:

$$\sigma_i^2 = f(\alpha_0 + a_1 Z_{1i} + \dots + a_k Z_{ki})$$

Diartikan bahwa σ_i^2 adalah suatu fungsi dari variabel Z yang nonstokastik. Secara spesifik, diasumsikan bahwa:

$$\sigma_i^2 = \alpha_0 + a_1 Z_{1i} + \dots + a_k Z_{ki}$$

Dengan σ_i^2 adalah fungsi linier dari Z , jika $a_1 = a_2 = \dots = a_k = 0$, maka $\sigma_i^2 = \alpha_0$ yang merupakan konstan. Untuk menguji apakah σ_i^2 homoskedastik, dapat diuji hipotesis bahwa $a_1 = a_2 = \dots = a_k = 0$. Hal tersebut merupakan dasar dari uji Breusch-Pagan.

Statistik Uji:

$$BP = \frac{1}{2} (JKG) \sim \chi^2_{(k-1)} \quad (8)$$

Kriteria Keputusan:

H_0 ditolak apabila statistik $BP_{hitung} > \chi^2_{(k-1),p}$ (Gujarati, 2004).

Uji non multikolinearitas digunakan untuk mengetahui apakah dalam variabel prediktor memiliki hubungan dengan variabel prediktor yang lain. Nilai VIF (*Variance Inflation Factor*) dapat diterapkan untuk melihat dalam model regresi tersebut apakah terdapat multikolinearitas atau tidak (Gujarati, 2004). Apabila perhitungan $VIF < 10$ maka tidak terdapat multikolinieritas pada regresi. Menghitung VIF dapat dirumuskan dengan:

$$VIF = \frac{1}{1-R^2} \quad (9)$$

R^2 merupakan koefisien determinasi antar variabel prediktor.

Dalam penelitian ini, model yang diterapkan yaitu *generalized linear model* dengan menggunakan fungsi *link* $g(\mu_i)$ yaitu log. Menurut Dobson & Barnett (2018). Apabila variabel yang berpengaruh pada variabel respons Y adalah $x_1, x_2, \dots, x_k$, dan $\beta = [\beta_0, \beta_1, \dots, \beta_k]^T$ merupakan koefisien regresi, maka model log linier pada nilai μ untuk respon Y pada pengamatan ke- i adalah:

$$\begin{aligned} g(\mu_i) &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \\ \mu_i &= E(Y) = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k) \\ \mu_i &= \exp(x_i^T \beta) \end{aligned} \quad (10)$$

Oleh karena itu, model yang diaplikasikan dalam observasi ini yaitu $y_i = \exp(x_i^T \beta)$. Pada model yang diterapkan, bagi variabel dengan skala interval maupun rasio, pada saat penambahan skor variabel x_1 sebesar satu satuan akan mengakibatkan variabel respon bertambah besar atau lebih kecil sebesar $\exp(\beta_1)$ satu satuan. Apabila koefisien β_1 memiliki nilai positif, variabel respon akan bertambah besar dan apabila koefisien β_1

memiliki nilai negative maka nilai variabel respons akan lebih kecil. Hal tersebut berlaku bagi $\beta_1, \beta_2, \dots, \beta_k$.

Berdasarkan persamaan 10 diperoleh:

$$\begin{aligned} E(Y_i) &= \mu_i = \alpha \theta_i \\ \alpha \theta_i &= \exp(\mathbf{x}_i^T \boldsymbol{\beta}) \\ \theta_i &= \frac{\exp(\mathbf{x}_i^T \boldsymbol{\beta})}{\alpha} \end{aligned} \quad (11)$$

Oleh karena itu, fungsi kepadatan peluang dari model regresi Gamma adalah:

$$f(y_i) = \frac{1}{\Gamma(\alpha) \left(\frac{\exp(\mathbf{x}_i^T \boldsymbol{\beta})}{\alpha} \right)^\alpha} y_i^{\alpha-1} \exp\left(\frac{-y_i}{\left(\frac{\exp(\mathbf{x}_i^T \boldsymbol{\beta})}{\alpha} \right)} \right) \quad (12)$$

Maximum Likelihood Estimation (MLE) bertujuan untuk memaksimumkan nilai fungsi *likelihood* dari estimasi parameter model (Casella & Berger, 2002). Apabila $y_1, \dots, y_n$ merupakan sampel acak dari fungsi kepadatan $f(y; \theta)$, sehingga fungsi *likelihood*nya adalah $f(y_1; \theta)f(y_2; \theta) \dots f(y_n; \theta)$ (Mood et al., 2007). Fungsi *likelihood* pada model regresi Gamma berdasarkan persamaan 12 sebagai berikut:

$$\begin{aligned} L(\alpha, \beta) &= \prod_{i=1}^n f(y_i) \\ &= \prod_{i=1}^n \left[\frac{1}{\Gamma(\alpha) \left(\frac{\exp(\mathbf{x}_i^T \boldsymbol{\beta})}{\alpha} \right)^\alpha} y_i^{\alpha-1} \exp\left(\frac{-y_i}{\left(\frac{\exp(\mathbf{x}_i^T \boldsymbol{\beta})}{\alpha} \right)} \right) \right] \\ &= \left[\left(\frac{1}{\Gamma(\alpha)} \right)^n \left(\frac{\exp \sum_{i=1}^n \mathbf{x}_i^T \boldsymbol{\beta}}{\alpha^n} \right)^{-\alpha} \prod_{i=1}^n y_i^{\alpha-1} \exp \sum_{i=1}^n \left(\frac{-y_i}{\left(\frac{\exp(\mathbf{x}_i^T \boldsymbol{\beta})}{\alpha} \right)} \right) \right] \end{aligned} \quad (13)$$

Oleh karena itu, diperoleh persamaan fungsi *ln likelihood* sebagai berikut:

$$\begin{aligned} \ln L(\alpha, \beta) &= \ln \left[\left(\frac{1}{\Gamma(\alpha)} \right)^n \left(\frac{\exp \sum_{i=1}^n \mathbf{x}_i^T \boldsymbol{\beta}}{\alpha^n} \right)^{-\alpha} \left(\prod_{i=1}^n y_i^{\alpha-1} \right) \exp \sum_{i=1}^n \left(\frac{-y_i}{\left(\frac{\exp(\mathbf{x}_i^T \boldsymbol{\beta})}{\alpha} \right)} \right) \right] \\ &= -n \ln \Gamma(\alpha) - \alpha \sum_{i=1}^n \mathbf{x}_i^T \boldsymbol{\beta} + \alpha(n \ln \alpha) + (\alpha - 1) \sum_{i=1}^n \ln y_i - \sum_{i=1}^n \frac{y_i}{\left(\frac{\exp(\mathbf{x}_i^T \boldsymbol{\beta})}{\alpha} \right)} \end{aligned} \quad (14)$$

Setiap parameter yang diestimasi pada regresi Gamma didapat dengan memperoleh turunan parsial pertama dan dari fungsi *ln likelihood* kemudian disamadengankan nol. Turunan parsial pada parameter $\boldsymbol{\beta}^T$ sebagai berikut:

$$\frac{\partial \ln L(\alpha, \beta)}{\partial \boldsymbol{\beta}^T} = \alpha \sum_{i=1}^n \mathbf{x}_i y_i (\exp(\mathbf{x}_i^T \boldsymbol{\beta}))^{-1} - \alpha \sum_{i=1}^n \mathbf{x}_i = 0 \quad (15)$$

Sedangkan untuk parameter α apabila turunan parsial fungsi *ln likelihood* disamadengankan nol didapat persamaan sebagai berikut:

$$\frac{\partial \ln L(\alpha, \beta)}{\partial \alpha} = -n \frac{1}{\Gamma(\alpha)} \Gamma'(\alpha) - \sum_{i=1}^n \mathbf{x}_i^T \boldsymbol{\beta} + (n \ln \alpha + n) - \sum_{i=1}^n \frac{y_i}{\exp(\mathbf{x}_i^T \boldsymbol{\beta})} = 0 \quad (16)$$

Karena pada persamaan turunan parsial setiap parameter tidak bisa terselesaikan secara analitik, maka diterapkan metode optimasi numerik yaitu dengan algoritma *Broyden-Fletcher-Goldfarb-Shanno* (BFGS) yaitu salah satu metode *Quasi Newton* yang diterapkan dalam meminimalkan suatu masalah optimasi nonlinier tanpa kendala yang berasal dari *gradient descent* melalui pendekatan matriks *Hessian* dengan tidak menghitung invers dari matriks *Hessian* sehingga cara ini dianggap efektif dan memiliki tingkat konvergensinya lebih cepat (Vandenberghe, 2011). Estimasi parameter di bawah

populasi pada regresi Gamma dengan metode algoritma *Metode-Broyden-Fletcher-Goldfarb-Shanno* (BFGS) adalah:

- a. Inisialisasi $\gamma_0, \mathbf{H}(\gamma_0)$, dengan $m = 0$ dan $\varepsilon > 0$ sebagai nilai toleransi konvergensi.

$$\gamma_0 = [\beta_{0(0)} \beta_{1(0)} \dots \beta_{k(0)} \alpha_{(0)}]^T$$

Menurut Rencher dan Shaaljae (2008), $\beta^T_{(0)}$ didapatkan dari metode OLS.

Estimasi parameter $\beta^T_{(0)}$ dengan metode OLS adalah:

$$\hat{\beta}_{(0)} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y};$$

$$\mathbf{X} = \begin{bmatrix} 1 & x_{1.1} & x_{2.1} & \dots & x_{k.1} \\ 1 & x_{1.2} & x_{2.2} & \dots & x_{k.2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1.n} & x_{2.n} & \dots & x_{k.n} \end{bmatrix}; \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Sesuai dengan persamaan (1) bahwa nilai $\alpha_{(0)}$ bernilai positif ($\alpha > 0$). $\mathbf{H}(\gamma_0)$ merupakan pendekatan matriks *Hessian* berupa matriks identitas $\mathbf{I}(\gamma_0)$ dengan $m = 0$ dan nilai $\varepsilon > 0$ sebagai toleransi konvergensi.

- b. Menghitung nilai $\mathbf{g}(\gamma_m) = \begin{bmatrix} \frac{\partial \ln L(\beta, \alpha | y_1, y_2, \dots, y_n)}{\partial \beta^T} \\ \frac{\partial \ln L(\beta, \alpha | y_1, y_2, \dots, y_n)}{\partial \alpha} \end{bmatrix}$

Proses iterasi akan dilanjutkan ke langkah berikutnya apabila $\|\mathbf{g}(\gamma_m)\| > \varepsilon$. Tetapi jika $\|\mathbf{g}(\gamma_m)\| < \varepsilon$ iterasi akan berhenti.

- c. Menghitung nilai $\mathbf{k}_m = -\mathbf{H}^{-1}(\gamma_m) \mathbf{g}(\gamma_m)$:

$\mathbf{k}_m$ merupakan vektor *descent direction* pada iterasi ke- m dan

$$\mathbf{H}(\gamma_m) = \begin{bmatrix} \frac{\partial^2 \ln L(\beta, \alpha | y_1, y_2, \dots, y_n)}{\partial \beta^T \partial \beta} & \frac{\partial^2 \ln L(\beta, \alpha | y_1, y_2, \dots, y_n)}{\partial \beta^T \partial \alpha} \\ \frac{\partial^2 \ln L(\beta, \alpha | y_1, y_2, \dots, y_n)}{\partial \alpha \partial \beta^T} & \frac{\partial^2 \ln L(\beta, \alpha | y_1, y_2, \dots, y_n)}{\partial \alpha^2} \end{bmatrix}$$

- d. Menentukan nilai $\lambda_m = \frac{\mathbf{g}^T(\gamma_m) \mathbf{k}_m}{\mathbf{k}_m^T \mathbf{k}_m}$

- e. Menghitung $\gamma_{m+1} = \gamma_m + \lambda_m \mathbf{k}_m$

- f. Menghitung $\mathbf{s}_m = \gamma_{m+1} - \gamma_m$

- g. Menghitung $\mathbf{d}_m = \mathbf{g}(\gamma_{m+1}) - \mathbf{g}(\gamma_m)$

- h. Menghitung matriks $\mathbf{H}(\gamma_{m+1})$ menggunakan persamaan:

- i. $\mathbf{H}(\gamma_{m+1}) = \left(1 - \frac{\mathbf{s}_m \mathbf{d}_m^T}{\mathbf{d}_m^T \mathbf{s}_m}\right) \mathbf{H}(\gamma_m) \left(1 - \frac{\mathbf{s}_m \mathbf{d}_m^T}{\mathbf{d}_m^T \mathbf{s}_m}\right) + \frac{\mathbf{s}_m \mathbf{s}_m^T}{\mathbf{d}_m^T \mathbf{s}_m}$

- j. Mengulangi lagi langkah (2) hingga seterusnya dengan $m = m+1$

Pada pemeriksaan parameter model regresi Gamma dilakukan secara simultan maupun parsial. Uji simultan bertujuan mendeteksi apakah terdapat pengaruh atau tidak pada variabel prediktor secara serentak (simultan) terhadap variabel respon menggunakan *Maximum Likelihood Ratio Test* (MLRT) berdasarkan hipotesis sebagai berikut:

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

$$H_1 : \text{minimal terdapat satu } \beta_j \neq 0 ; j = 1, 2, \dots, k$$

Statistik Uji:

$$G = -2 \ln \left(L(\hat{\omega}) - L(\hat{\Omega}) \right) \quad (17)$$

Kriteria keputusan :

Jika $G_{hitung} > \chi^2_{k,p}$ maka H_0 ditolak (Nasution, 2017).

Uji parsial dapat dilakukan karena uji simultan yang sudah diuji sebelumnya memiliki keputusan tolak H_0 sehingga uji parsial dilakukan dengan tujuan untuk melihat pengaruh

secara individu variabel prediktor terhadap variabel respon. Berdasarkan hipotesis sebagai berikut:

$$H_0 : \beta_j = 0$$

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

Statistik Uji:

$$Z = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \quad (18)$$

$$\text{dengan } SE(\hat{\beta}_j) = \sqrt{\text{var}(\hat{\beta}_j)}$$

Kriteria Uji:

Jika $|Z_{hit}| > Z_{\rho/2}$ maka H_0 ditolak.

Pada model regresi, Menurut Ghazali (2016) penentuan ukuran kebaikan model dalam memprediksi variabel terikat dapat dijelaskan dengan mengetahui nilai koefisien determinasi yaitu suatu skor yang mendeskripsikan seberapa besar variabel respon dapat dijelaskan oleh variabel prediktor yang dapat dirumuskan dengan:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (19)$$

3. METODE PENELITIAN

Data pada penelitian ini diperoleh dari Dinas Lingkungan Hidup Kota Semarang pada tahun 2022 periode II dengan 60 titik lokasi pengamatan yang berada di Kota Semarang. Dengan variabel prediktor (Y) yaitu COD dan variabel prediktor (X) yaitu debit, temperatur, TSS, pH, BOD, dan DO.

Pada penelitian ini digunakan *software* R Studio 4.2.2 dengan tahapan analisis data sebagai berikut:

- Membuat analisis deskriptif variabel respon dan variabel prediktor.
- Melakukan uji distribusi Gamma untuk variabel respon menggunakan metode Anderson Darling.
- Menetapkan model regresi Gamma dengan estimasi parameter metode *Maximum Likelihood Estimation* (MLE) berdasarkan langkah berikut:
 - Memaksimumkan fungsi *likelihood* pada model regresi Gamma.
 - Mendapatkan turunan parsial pertama fungsi *ln likelihood* terhadap masing-masing parameter kemudian disamadengankan nol.
 - Jika tidak mendapatkan hasil yang *close-form*, dilakukan estimasi dengan metode optimasi BFGS (*Broyden-Fletcher-Goldfarb-Shanno*).
- Melakukan pengujian asumsi non heteroskedastisitas dengan Uji *Breusch Pagan*.
- Melakukan pengujian asumsi non multikolinieritas pada variabel prediktor.
- Melakukan uji signifikansi parameter secara simultan dan parsial pada setiap model regresi.
- Melakukan interpretasi pada model yang terbentuk dan menetapkan faktor yang memiliki pengaruh terhadap variabel respon.

4. HASIL DAN PEMBAHASAN

Statistik deskriptif variabel COD (*Chemical Oxygen Demand*) pada data pencemaran sungai di Kota Semarang tahun 2022 dapat dilihat pada Tabel 1.

Tabel 1. Statistik deskriptif variabel COD

Variabel	Min	Max	Mean
----------	-----	-----	------

Y	4,967	77,793	27,65278
---	-------	--------	----------

Pada Tabel 1 hasil analisis deskriptif diperoleh hasil bahwa total COD (*Chemical Oxygen Demand*) paling banyak adalah 77,793 dan paling sedikit adalah 4,967. Rata-rata dari total *Chemical Oxygen Demand* (COD) pada setiap titik pengamatan Sungai di Kota Semarang yaitu 27,65278. Hal tersebut menunjukkan bahwa rata-rata sungai di Kota Semarang termasuk dalam golongan tercemar.

Pengujian asumsi pada regresi Gamma dilakukan terhadap variabel respon *Chemical Oxygen Demand* (COD) menggunakan uji Anderson-Darling dan didapat hasil pada Tabel 2.

Tabel 2. Uji Distribusi Gamma

Distribusi	AD	p-value
Gamma	1,8758	07,57e-05

Pada Tabel 2 diperoleh hasil nilai $A (1,8758) < A_p (2,5044)$ sehingga gagal tolak H_0 yang berarti indikator *Chemical Oxygen Demand* (COD) sesuai dengan distribusi tertentu yaitu berdistribusi Gamma.

Pada regresi Gamma menghasilkan parameter yang bersifat global yang artinya sama untuk masing-masing lokasi pengamatan. Pada penelitian ini terdapat 60 titik lokasi pengamatan dengan faktor yang mempengaruhi antar titik lokasi pengamatan adalah sama. Variabel prediktor yang digunakan pada pemodelan regresi Gamma ini yaitu X_1 (debit), X_2 (temperatur), X_3 (TSS), X_4 (pH), X_5 (BOD), X_6 (DO).

Berdasarkan hasil pengujian menggunakan *software* RStudio diperoleh model pada regresi Gamma yang disajikan pada Tabel 3.

Tabel 3. Model Regresi Gamma

Parameter	Estimasi	SE	Nilai Z	p-value
Konstan	3,127000	0,889700	3,5140	0,0009114
X_1	0,010470	0,018380	0,5697	0,5713000
X_2	-0,021750	0,018870	-1,1530	0,2541000
X_3	-0,002846	0,001842	-1,5450	0,1282000
X_4	0,012100	0,079320	0,1525	0,8793000
X_5	0,074690	0,004207	17,7500	< 2,2e-16
X_6	-0,082560	0,048730	-1,6940	0,0961000

Berdasarkan Tabel 3 model yang terbentuk dari 6 variabel prediktor terhadap indikator *Chemical Oxygen Demand* (COD) pada pencemaran Sungai di Kota Semarang dengan tingkat signifikansi (ρ) 5% yaitu:

$$\hat{y} = \exp (3,127 + 0,01047X_1 - 0,02175X_2 - 0,002846X_3 + 0,0121X_4 + 0,07469X_5 - 0,08256X_6)$$

Estimasi parameter pada regresi Gamma diterapkan metode *Maximum Likelihood Estimation* (MLE) dengan metode optimasi numerik yaitu *Broyden Fletcher Goldfarb Shanno* (BFGS). Berdasarkan hasil pengujian didapatkan hasil estimasi parameter pada regresi Gamma pada Tabel 4.

Tabel 4. Estimasi Parameter Model Regresi Gamma

Parameter	Estimasi	SE
Konstan	3,127000	0,889700
X_1	0,010470	0,018380

X_2	-0,021750	0,018870
X_3	-0,002846	0,001842
X_4	0,012100	0,079320
X_5	0,074690	0,004207
X_6	-0,082560	0,048730

Berdasarkan Tabel 4 hasil estimasi diperoleh nilai $\hat{\alpha} = 16,91036$ dan $\hat{\theta} = 0,05913534$. Dari hasil estimasi parameter regresi Gamma dapat dilihat bahwa parameter *shape* (α) memiliki nilai lebih dari 1 yang artinya peluang dari kadar COD yang melewati batas ambang baku mutu air yang sudah ditetapkan akan mengalami peningkatan dari waktu ke waktu.

Pengujian asumsi non heteroskedastisitas pada regresi Gamma dilakukan dengan *Breusch Pagan test* dan didapatkan hasil pada Tabel 5.

Tabel 5. Uji Heteroskedastisitas

Breusch Pagan	p-value	df
4,1715	0,6535	6

Pada Tabel 5 diperoleh nilai Breusch-Pagan yaitu $4,1715 < \text{nilai } \chi^2$ yaitu 12,592 dengan nilai p-value = 0,6535 pada taraf signifikansi 5% memiliki nilai $>$ taraf signifikansi (ρ), sehingga bisa disimpulkan terima H_0 , yang berarti tidak terjadi heteroskedastisitas.

Pengujian asumsi untuk mendeteksi kasus multikolinearitas dapat dilakukan menggunakan nilai VIF dari masing-masing variabel prediktor dan diperoleh hasil pada Tabel 6.

Tabel 6. Uji Multikolinearitas

Variabel	VIF
X_1	1,083857
X_2	1,207232
X_3	1,101104
X_4	1,125823
X_5	1,153213
X_6	1,169352

Pada Tabel 6 pengujian multikolinearitas diperoleh hasil bahwa nilai VIF pada setiap variabel prediktor lebih kecil dari 10, sehingga dapat disimpulkan bahwa antar variabel prediktor tidak terjadi multikolinearitas dan semua variabel dapat digunakan dalam pemodelan regresi Gamma.

Pengujian secara simultan dilakukan pada seluruh variabel prediktor untuk mendeteksi ada atau tidak pengaruh variabel prediktor secara serentak terhadap model dan didapatkan hasil pada Tabel 7.

Tabel 7. Uji Serentak Regresi Gamma

G	df	$\chi^2_{k,\rho}$	Keputusan
21,608	6	30,144	Tolak H_0

Pada Tabel 7 uji serentak (simultan) pada model regresi didapatkan nilai G yaitu 21,608 yang memiliki hasil lebih dari nilai $\chi^2_{k,\rho}$ yaitu 30,144. Oleh karena itu dapat ditarik kesimpulan bahwa keputusan yang didapat menolak H_0 atau dapat dikatakan secara bersama-sama parameter berpengaruh terhadap model.

Pengujian secara parsial pada regresi Gamma untuk mendeteksi pengaruh variabel predictor secara individual terhadap variabel respon dan diperoleh hasil pada Tabel 8.

Tabel 8. Uji Parsial Regresi Gamma

Parameter	Estimasi	SE	Nilai Z	p-value	Keputusan
Konstan	3,127000	0,889700	3,5140	0,0009114	Tolak H_0
X_1	0,010470	0,018380	0,5697	0,5713000	Gagal Tolak H_0
X_2	-0,021750	0,018870	-1,1530	0,2541000	Gagal Tolak H_0
X_3	-0,002846	0,001842	-1,5450	0,1282000	Gagal Tolak H_0
X_4	0,012100	0,079320	0,1525	0,8793000	Gagal Tolak H_0
X_5	0,074690	0,004207	17,7500	< 2,2e-16	Tolak H_0
X_6	-0,082560	0,048730	-1,6940	0,0961000	Gagal Tolak H_0

Pada Tabel 8 variabel yang memiliki pengaruh terhadap indikator *Chemical Oxygen Demand* (COD) pada pencemaran sungai di Kota Semarang dengan tingkat signifikansi 5% yaitu variabel X_5 atau BOD (*Biochemical Oxygen Demand*) dengan $Z_{\alpha/2} = 1,96$.

Meregresikan kembali variabel yang signifikan yaitu variabel X_5 atau BOD (*Biochemical Oxygen Demand*) dan diperoleh hasil pada Tabel 9.

Tabel 9. Uji Parsial Regresi Gamma Variabel Signifikan

Parameter	Estimasi	SE	Nilai Z	p-value	Keputusan
Konstan	2,188	0,060710	36,03	< 2,2e-16	Tolak H_0
X_5	0,076690	0,004101	18,70	< 2,2e-16	Tolak H_0

Berdasarkan hasil analisis pada Tabel 9 dapat disimpulkan bahwa variabel X_5 atau *Biochemical Oxygen Demand* (BOD) berpengaruh terhadap *Chemical Oxygen Demand* (COD) dan diperoleh model regresi Gamma pada pencemaran sungai di Kota Semarang yaitu:

$$\hat{y} = \exp(2,188 + 0,076690X_5)$$

Pada model yang terbentuk menjelaskan bahwa setiap BOD yang terlarut dalam air sungai nilai variabel *Chemical Oxygen Demand* (COD) akan dilipat gandakan sebanyak $\exp(0,076690) = 1,079707$ mg/l.

Koefisien determinasi (R^2) yang digunakan untuk menggambarkan seberapa besar variabel respon bisa dijelaskan oleh variabel prediktor dan diperoleh hasil sebesar 0,4431498 atau dalam kategori moderat yang berarti bahwa kemampuan variabel prediktor dalam menjelaskan variabel respon sebesar 44,32%, sisanya 55,68% variabel respon dijelaskan oleh faktor lain yang tidak terdapat dalam model tersebut.

5. KESIMPULAN

Berdasarkan hasil analisis yang dilakukan, diperoleh kesimpulan sebagai berikut:

1. Pada model regresi Gamma bentuk estimasi parameter yang digunakan yaitu dengan metode *Maximum Likelihood Estimation* (MLE). Karena persamaan yang didapatkan menghasilkan bentuk yang tidak *close form* sehingga dilakukan metode optimasi dengan *Broyden-Fletcher-Goldfarb-Shanno* (BFGS) dalam memperoleh nilai parameter.
2. Berdasarkan hasil analisis pemodelan regresi Gamma menggunakan metode *Maximum Likelihood Estimation* (MLE) dengan metode optimasi *Broyden-Fletcher-Goldfarb-Shanno* (BFGS) diperoleh model yaitu

$$\hat{y} = \exp(2,188 + 0,076690X_5)$$
3. Berdasarkan hasil analisis yang diperoleh faktor yang berpengaruh terhadap tingginya nilai *Chemical Oxygen Demand* (COD) pada sungai di Kota Semarang tahun 2022 yaitu variabel X_5 atau *Biochemical Oxygen Demand* (BOD). Pada model yang terbentuk menjelaskan bahwa setiap *Biochemical Oxygen Demand* (BOD) yang

terlarut dalam air sungai nilai variabel *Chemical Oxygen Demand* (COD) akan dilipatgandakan sebanyak $0,076690 = 1,079707$ mg/l. Pada koefisien regresi untuk variabel X_5 atau BOD (*Biochemical Oxygen Demand*) dalam air sungai menunjukkan apabila konsentrasi BOD semakin tinggi akan berpengaruh terhadap kualitas air sungai yang dapat meningkatkan konsentrasi *Chemical Oxygen Demand* (COD) dalam perairan. Peningkatan *Biochemical Oxygen Demand* (BOD) sebesar satu satuan akan memperbesar konsentrasi *Chemical Oxygen Demand* (COD) sebesar $1,079707$ mg/l.

DAFTAR PUSTAKA

- Bossio, M. C., & Cuervo, E. C. 2015. *Gamma regression models with the Gammareg R package*. *Comunicaciones en estadística*, 8(2), 211-223.
- Dobson, A. J., & Barnett, A. G. 2018. *An introduction to generalized linear models*. CRC press.
- Putri, E. D., & Prastyo, D. 2017. *Parameter estimation and hypothesis testing on geographically weighted gamma regression*. In *Journal of Physics Conference Series* (Vol. 893, No. 1, p. 012025).
- Faizah, K. 2018. *Regresi Gamma dan Regresi Lognormal untuk Memodelkan Biomassa Klorofil-a di Sungai Danube Eropa*. Tesis. Institut Teknologi Sepuluh November Surabaya.
- Ghozali, I. 2018. *Aplikasi Analisis Multivariate dengan Program IBM SPSS 25*. Badan Penerbit Universitas Diponegoro: Semarang
- Gujarati, D. N. 2004. *Basic Econometrics (4 ed.)*. New York: The McGraw-Hill.
- Lumaela, A. K., Otok, B. W., & Sutikno, S. 2013. *Pemodelan chemical oxygen demand (cod) sungai di Surabaya dengan metode mixed geographically weighted regression*. *Jurnal Sains dan Seni ITS*, 2(1), D100-D105.
- Luthfiana, U. 2021. *Model Regresi Gamma Pada Indikator Pencemaran Sungai di Kota Semarang*. Skripsi. Universitas Islam Negeri Walisongo Semarang.
- Misbahussurur, A. 2009. *Estimasi Parameter Distribusi Gamma dengan Metode Maksimum Likelihood*. Skripsi. Universitas Islam Negeri Maulana Malik Ibrahim.
- Nasution, A.S. 2017. *Estimasi Parameter dan Pengujian Hipotesis pada Model Regresi Gamma (Studi Kasus Pemodelan Pencemaran Air Sungai di Surabaya)*. Tesis. Institute Teknologi Sepuluh Nopember Surabaya.
- Saputro, D. R., & Purnami, W. 2016. *Algoritme Pendugaan Parameter Model Regresi Logistic Biler (RLB) dengan Maksimum Likelihood dan Broyden-Fletcher-Goldfarb-Shanno (BFGS)*. Seminar Nasional Matematika Dan Pendidikan Matematika.