Jurnal Riset Statistika (JRS)
e-ISSN 2798-6578 | p-ISSN 2808-3148
https://journals.unisba.ac.id/index.php/JRS

Tersedia secara online di
Unisba Press
https://publikasi.unisba.ac.id/

Kepercayaan *The Bootstrap Percentile Confidence Interval* untuk Koefisien Variasi

Alya Salsabilla Suherman, Abdul Kudus*

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

ARTICLE INFO

Article history :
Received : 9/5/2024
Revised : 30/5/2024
Published : 31/7/2024

Volume : 4
No. : 1
Halaman : 21 - 30
Terbitan : **Juli 2024**

ABSTRAK

Distribusi Invers Gaussian (IG) digunakan untuk memodelkan data yang memiliki distribusi kemiringan positif atau condong kekanan. Koefisien variasi (KV) merupakan perbandingan antara standar deviasi dengan nilai rata-rata. KV mencerminkan ukuran penyebaran yang baik digunakan ketika membandingkan dua atau lebih kelompok data. Salah satu metode untuk menaksir selang kepercayaan koefisien variasi (KV) yaitu metode The Bootstrap Percentile Confidence Interval (BPCI). Dalam skripsi ini akan dilakukan penerapan metode BPCI untuk koefisien variasi dari distribusi IG pada data kualitas udara PM 2.5. Dalam proses penelitian tahapan analisis yang dilakukan meliputi menghitung taksiran parameter dari distribusi IG menggunakan metode maksimum likelihood, melakukan uji kecocokan Kolmogorov-Smirnov, membangkitkan data dengan parameter sebagaimana taksirannya dan menaksir parameternya sebanyak 1000 kali, kemudian menghitung selang kepercayaan 95% dengan metode BPCI. Berdasarkan hasil perhitungan diperoleh bahwa data kualitas udara PM 2.5 di Kota Semarang tahun 2023 cocok dimodelkan dengan distribusi IG dan selang kepercayaan BPCI dengan tingkat kepercayaan 95% untuk KV dari data tersebut berada diantara rentang [0,31 ;0,42].

Kata Kunci : Distribusi Invers Gaussian*; Bootstrap Percentile Confidence Interval;* Kualitas Udara.

ABSTRACT

The Inverse Gaussian (IG) distribution is used to model data that has a positively skewed or right-skewed distribution. The coefficient of variation (CV) is the ratio of the standard deviation to the mean value. It reflects a good measure of dispersion used when comparing two or more groups of data. One method for estimating the coefficient of variation (CV) confidence interval is The Bootstrap Percentile Confidence Interval (BPCI) method. In this thesis, the BPCI method will be applied to the coefficient of variation of the IG distribution on PM 2.5 air quality data. In the research process, the stages of analysis carried out include calculating the estimated parameters of the IG distribution using the maximum likelihood method, conducting the Kolmogorov-Smirnov fit test, generating data with parameters as estimated and estimating the parameters 1000 times, then calculating the 95% confidence interval with the BPCI method. Based on the calculation results, it is obtained that the PM 2.5 air quality data in Semarang City in 2023 is suitable to be modeled with IG distribution and BPCI confidence interval with 95% confidence level for KV of the data is between the range [0.31 ;0.42].

Keywords : *Invers Gaussian Distribution*; *Bootstrap Percentile Confidence Interval*; *Air Quality.*

Corresponding Author : *abdulkudus@unisba.ac.id
Indexed : Garuda, Crossref, Google Scholar
DOI : https://doi.org/10.29313/jrs.v4i1.3854

A. Pendahuluan

Penaksiran adalah suatu proses yang digunakan untuk memperkirakan nilai dari suatu parameter. Penaksiran parameter yaitu suatu tahapan terpenting dalam menentukan model peluang yang tepat dari sekumpulan data. Data yang digunakan untuk melakukan penaksiran parameter merupakan suatu sampel yang akan digunakan untuk menentukan penaksiran terbaik berdasarkan nilai dari evaluasi metode penaksir terbaik . Penaksiran parameter dibagi menjadi dua yaitu penaksiran titik dan penaksiran interval. Penaksiran titik adalah nilai hasil dari perkiraan parameter suatu populasi yang diperoleh berdasarkan data sampel. Penaksiran interval adalah penaksiran suatu parameter berdasarkan nilai dalam suatu interval yang akan memberikan tingkat kepercayaan tertentu. Dalam jurnal ini, penulis akan melakukan penaksiran parameter dari distribusi invers gaussian (IG) dengan menggunakan data riil di bidang lingkungan.

Distribusi IG merupakan distribusi yang digunakan untuk memodelkan peluang dari variabel acak [1] [2]. Distribusi IG juga digunakan untuk memodelkan waktu kegagalan cairan isolasi, model kejadian pasar, partikel distribusi waktu siklus dalam darah, biaya kesehatan, dan polusi udara oleh [3]. Pada analisis distribusi polusi udara di pusat kota Athena, diperoleh hasil bahwa distribusi IG memberikan hasil yang lebih baik daripada distribusi beta, gamma dan distribusi Weibull, oleh karena itu penulis tertarik untuk mempelajari polusi PM 2.5 dengan menggunakan distribusi IG [4]. Distribusi IG pertama kali diperkenalkan oleh Schrodinger pada tahun 1915 digunakan untuk mengatasi data yang memiliki kemiringan positif atau condong kekanan [5]. Distribusi IG adalah pilihan yang sangat baik untuk memodelkan data positif dan data condong ke kanan, dan beberapa fitur statistiknya oleh [6].

Dalam kajian statistika ada yang dikenal dengan koefisien variasi (KV) yaitu perbandingan antara standar deviasi dengan nilai rata-rata. Koefisien variasi ini telah digunakan dalam berbagai bidang diantaranya pertanian, biologi, kedokteran dan keuangan. Banyak peneliti yang telah mengembangkan parameter koefisien variasi dan selang kepercayaan. Chankham W, Niwitpong S.a Niwitpong S (2022) menerapkan berbagai macam pendugaan confidence interval pada data polusi PM 2.5. Salah satu metode yang ada di penelitian tersebut yaitu selang kepercayaan The Bootstrap Percentile Confidence Interval (BPCI) [5]. Metode BPCI sendiri merupakan salah satu metode untuk membuat selang kepercayaan. Akan dibuat selang kepercayaan dari koefisien variasi yang berasal dari fungsi densitas distribusi IG yang akan diterapkan pada data kualitas udara PM 2.5 di Kota Semarang [7]. Kota Semarang dipilih karena dari beberapa kota di Indonesia, data yang memenuhi dan memilki kemiringan positif atau condong kekanan yaitu Kota Semarang.

Udara adalah campuran gas pada lapisan yang mengelilingi bumi (atmosfer), yang komposisi udaranya tidak selalu tetap [8]. Udara merupakan faktor lingkungan yang penting dalam kehidupan, maka perlu dijaga dan ditingkatkan (Wardoyo, 2016). Kualitas udara atau air quality adalah tingkat kandungan udara, berdasarkan konsentrasi polutan di lokasi tertentu [9]. Dalam Peraturan Menteri Lingkungan Hidup dan Kehutanan Nomor P.14/Menlhk/Setjen/Kum.1/7/2020 mengenai Indeks Standar Pencemar Udara (ISPU) dijelaskan bahwa ISPU adalah angka tanpa satuan yang menggambarkan keadaan kualitas udara ambien di lokasi tertentu yang berdampak terhadap kesehatan manusia, nilai estetik, dan makhluk hidup lainnya [10]. Parameter ISPU diantaranya yaitu Karbon monoksida (CO), Nitrogen dioksida ($\boldsymbol{NO_2}$), Sulfur dioksida ($\boldsymbol{SO_2}$), Hidrokarbon (HC), Ozon ($\boldsymbol{O_3}$), Sulfur dioksida ($\boldsymbol{SO_2}$) dan Partikulat PM 10 dan PM 2.5.

B. Metode Penelitian

Distribusi Invers Gaussian (IG)

Misalkan $X = (x_1, x_2, \ldots, x_n)$ adalah peubah acak dari distribusi IG yang dinotasikan sebagai $IG \sim (\mu, \lambda)$ di mana memiliki dua parameter yaitu μ adalah parameter rata-rata dan λ adalah parameter skala [11][12]. Fungsi densitas probabilitas dari X yang mengikuti distribusi IG dengan parameter μ dan λ adalah sebagai berikut:

$$f(x;\mu,\lambda) = \left\{\frac{\lambda}{2\pi X^3}\right\}^{\frac{1}{2}} exp\left\{\frac{-\lambda(X_i-\mu)^2}{2\mu^2 X}\right\}, x > 0, \mu > 0, \lambda > 0 \quad (1)$$

Edgemen et al (1988) dan Mudholkar et al (2001) menyatakan bahwa fungsi distribusi kumulatif dari peubah acak X yang mengikuti distribusi IG seperti pada persamaan (2.1) adalah terkait dengan fungsi dari distribusi kumulatif normal baku, yaitu:

$$F(x;\mu,\lambda)=\Phi\left[\left(\frac{x}{\mu}-1\right)\left(\frac{\lambda}{x}\right)^{\frac{1}{2}}\right]+\exp\left(\frac{2\lambda}{\mu}\right)\Phi\left[\left(\frac{x}{\mu}+1\right)\left(\frac{\lambda}{x}\right)^{\frac{1}{2}}\right];untuk>0\ dan\ \mu,\lambda>0 \quad (2)$$

Di mana $\Phi(t)$ adalah fungsi distribusi kumulatif normal baku yang dievaluasi pada t. Rata-rata dan varians dari distribusi IG sebagai berikut:

$$E(X)=\mu \quad (3)$$

$$Var(X)=\frac{\mu^3}{\lambda} \quad (4)$$

Taksiran Taksiran Parameter Distribusi Invers Gaussian

Parameter dari distribusi IG yaitu μ dan λ, yang dinotasikan $X_i \sim IG(\mu,\lambda)$ dengan fungsi densitas seperti pada persamaan (2.1), akan dilakukan taksiran parameter menggunakan metode kemungkinan maksimum *likelihood* (MLE). Fungsi *likelihood* untuk distribusi IG yaitu sebagai berikut:

$$L(\mu,\lambda)=\left(\frac{\lambda}{2\pi}\right)^{\frac{n}{2}}\left(\prod_{i=1}^{n}X_i^{-\frac{3}{2}}\right)exp\left\{-\sum_{i=1}^{n}\frac{\lambda(X_i-\mu)^2}{2\mu^2X_i}\right\} \quad (5)$$

Fungsi *log-likelihood* untuk distribusi IG yaitu:

$$\ln L(\mu,\lambda)=\frac{n}{2}ln\left(\frac{\lambda}{2\pi}\right)-\frac{3}{2}\sum_{i=1}^{n}lnX_i-\sum_{i=1}^{n}\frac{\lambda(X_i-\mu)^2}{2\mu^2X_i} \quad (6)$$

Turunan fungsi *log-likelihood* terhadap parameter μ yaitu

$$\frac{\partial lnL(\mu,\lambda)}{\partial\mu}=\sum_{i=1}^{n}\lambda\left\{\frac{2(X_i-\mu)(-1)2\mu^2X_i-(X_i-\mu)^2 4\mu X_i}{(2\mu^2X_i)^2}\right\}=0$$

$$\sum_{i=1}^{n}\lambda\left\{\frac{-4\mu^2X_i(X_i-\mu)-4\mu X_i(X_i-\mu)^2}{4\mu^4X_i^2}\right\}=0 \quad (7)$$

$$n\mu-\sum_{i=1}^{n}X_i=0$$

$$\hat{\mu}=\bar{X}=\frac{\sum_{i=1}^{n}X_i}{n}$$

Turunan fungsi *log-likelihood* terhadap parameter λ yaitu

$$\frac{\partial lnL(\mu,\lambda)}{\partial\lambda}=\frac{n}{2}\ln\left(\frac{\lambda}{2\pi}\right)-\sum_{i=1}^{n}\left\{\frac{\lambda(X_i-\mu)^2}{2\mu^2X_i}\right\}=0$$

$$\frac{\partial lnL(\mu,\lambda)}{\partial\lambda}=\frac{n}{2\lambda}-\sum_{i=1}^{n}\left\{\frac{(X_i-\mu)^2}{2\mu^2X_i}\right\}=0 \quad (8)$$

$$\frac{1}{\lambda}=\frac{1}{n}\left(\sum_{i=1}^{n}\left(\frac{X_i}{2\mu^2}-\frac{2}{\mu}+\frac{1}{X_i}\right)\right)$$

Di mana nilai $\hat{\mu}=\bar{X}=\frac{\sum_{i=1}^{n}X_i}{n}$

$$\frac{1}{\lambda} = \frac{1}{n}\left[\sum_{i=1}^{n}\left(\frac{1}{X_i}\right) - \frac{n}{\bar{X}}\right]$$
Atau (9)
$$\hat{\lambda}^{-1} = \frac{1}{n}\left[\sum_{i=1}^{n}\left(\frac{1}{X_i} - \frac{1}{\bar{X}}\right)\right]$$

Maka, taksiran parameter untuk μ dan λ adalah $\hat{\mu} = \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ dan $\hat{\lambda} = \frac{n}{\sum_{i=1}^{n}\left(\frac{1}{X_i} - \frac{1}{\bar{X}}\right)}$.

Uji Kecocokan Kolmogorov Smirnov

Misalkan sampel acak yang berukuran n yaitu $X_1, X_2, \ldots, X_n$ di mana realisasi dari sampel acak tersebut adalah $x_1, x_2, \ldots, x_n$ yang berkaitan dengan suatu fungsi distribusi yang tidak diketahui $F(x)$ dan misalkan $F^*(x)$ adalah suatu fungsi sebaran yang dihipotesiskan secara lengkap. Adapun rumusan hipotesis yang digunakan pada uji *Kolmogorov-Smirnov* sebagai berikut:

$H_0: F(x) = F^*(x)$;Data berasal dari suatu populasi berdistribusi tertentu.

$H_1: F(x) \neq F^*(x)$;Data tidak berasal dari suatu populasi berdistribusi tertentu.

Statistik uji untuk *Kolmogorov Smirnov* dinyatakan dalam bentuk:

$$D = \max_{1 \le i \le n} |F_n(x_i) - F^*(x_i)| \tag{10}$$

Di mana:

$$F_n(x_i) = \frac{Banyaknya\ pengamatan \le\ x_i}{n}$$

Keterangan:

$F_n(x_i) =$ fungsi distribusi empiris untuk data pengamatan ke-i.

$F^*(x_i) =$ fungsi distribusi kumulatif yang ditentukan untuk data pengamatan ke-i berdasarkan fungsi distribusi IG di mana parameter μ dan λ disubtitusi oleh taksiran parameter $\hat{\mu}$ dan $\hat{\lambda}$.

Kriteria uji yang digunakan yaitu jika nilai D lebih kecil dari nilai kritisnya pada taraf nyata α yang sudah ditentukan maka disimpulkan bahwa hipotesis nol diterima artinya data berasal dari suatu populasi berdistribusi tertentu.

Koefisien Variasi

Salah satu ukuran keragaman atau variasi dari suatu kelompok data dikenal dengan koefisien variasi (KV). Koefisien variasi merupakan perbandingan antara standar deviasi (σ) dengan nilai rata-rata (μ). Koefisien variasi dapat dihitung dengan rumus yaitu:

$$KV = \frac{\sigma}{\mu} \tag{11}$$

Keterangan:

$KV =$ koefisien variasi.

$\sigma =$ standar deviasi.

$\mu =$ rata-rata.

Dapat dihitung KV untuk X mengikuti distribusi IG:

$$KV(X) = \theta = \frac{\sqrt{Var(X)}}{E(X)}$$
$$KV(X) = \sqrt{\frac{\mu}{\lambda}} \tag{12}$$

Metode Bootsrap

Metode *bootstrap* mulai diperkenalkan oleh Bradley Efron pada tahun 1979, sebagai suatu metode pengambilan contoh ulang secara acak dengan pemulihan. Metode *bootstrap* merupakan metode yang digunakan untuk mengestimasi suatu distribusi populasi yang tidak diketahui dengan distribusi empiris yang diperoleh dari proses penyampelan ulang [13].

Secara garis besar *bootstrap* dibagi dua, yaitu *bootstrap* parametrik dan *bootstrap* nonparametrik. *Bootstrap* parametrik mengasumsikan bahwa penaksir, yaitu $\hat{\theta}$ berasal dari distribusi tertentu [14]. Dalam kasus *bootstrap* parametrik asumsi tersebut mungkin bukan terkait dengan parameter, tetapi terkait dengan model.

***Bootstrap* Parametrik**

Misalkan $x_1, x_2, \ldots, x_n$ adalah data sampel dari suatu populasi. Diasumsikan data berasal dari suatu distribusi tertentu. Misalkan parameter yang menjadi penelitian adalah θ dengan penaksirnya adalah $\hat{\theta} = s(x_1, x_2, \ldots, x_n)$ [15]. Tahapan dalam *bootstrap* parametrik adalah sebagai berikut:

Kesatu, Lakukan penaksiran parameter distribusi dengan metode yang sudah tersedia seperti maksimum *likelihood*, MKT atau yang lainnya berdasarkan data sampel yang ada.

Kedua, Tentukan banyaknya $bootstrap$ resample sebesar B, misalkan $B = 1000$ atau $B = 2000$. Set $b = 1$.

Ketiga, Misalkan data $x_1, x_2, \ldots, x_n$ adalah data sampel dari suatu populasi dengan distribusi sebagaimana pada langkah 1, diasumsikan data berasal dari suatu distribusi tertentu.

Keempat, Lakukan penaksiran nilai θ, yaitu $\hat{\theta}_b^* = s(x_1^*, x_2^*, \ldots, x_n^*)$. Simpan nilai $\hat{\theta}_b^*$.

Keempat, Set $b = b + 1$.

Kelima, Ulang langkah 3-5 sebanyak B kali.

Statistik $\hat{\theta}^*$ yang diperoleh dari *bootstrap*, yaitu $\hat{\theta}_1^*, \hat{\theta}_2^*, \ldots, \hat{\theta}_B^*$ akan membentuk distribusi sampling bagi $\hat{\theta}^*$. Karena diperoleh dari proses *bootstrap*, maka distribusi dari $\hat{\theta}_1^*, \hat{\theta}_2^*, \ldots, \hat{\theta}_B^*$ disebut sebagai distribusi *bootstrap* [16].

Selang Kepercayaan untuk KV dari Distribusi IG Menggunakan Metode BPCI

Misal $X = (x_1, x_2, \ldots, x_n)$ adalah sampel acak berukuran n dari distribusi IG. Data hasil sampling dinotasikan sebagai $X^* = (x_1^*, x_2^*, \ldots, x_n^*)$ yang berasal dari *bootstrap resample* sebesar B kali [17]. Diketahui bahwa minimum sekitar B = 1000 sampel *bootstrap* biasanya cukup untuk mendapatkan selang kepercayaan yang akurat untuk KV distribusi IG [18].

Selang kepercayaan $100(1-\alpha)\%$ untuk KV dari distribusi IG berdasarkan BPCI didefinisikan sebagai berikut:

$$\mathrm{CI_{BPCI}} = [\mathrm{L_{BPCI}}, \mathrm{U_{BPCI}}] = \left[KV_{\left(\frac{\alpha}{2}\right)}, KV_{\left(1-\frac{\alpha}{2}\right)}\right] \tag{13}$$

Keterangan bahwa $KV_{\left(\frac{\alpha}{2}\right)}$ dan $KV_{\left(1-\frac{\alpha}{2}\right)}$ masing-masing adalah persentil ke-$100\left(\frac{\alpha}{2}\right)$ dan persentil ke-$100\left(1-\frac{\alpha}{2}\right)$ dari parameter KV-nya.

Tahapan dalam BPCI sebagai berikut: Misalkan data $x_1, x_2, \ldots, x_n$ dari distribusi IG dengan parameter (μ, λ). Hitung $\hat{\mu}$ menggunakan persamaan (2.7). Hitung $\hat{\lambda}$ menggunakan persamaan (2.8). Hitung $\widehat{KV} = \sqrt{\frac{\hat{\mu}}{\hat{\lambda}}}$ menggunakan persamaan (2.12). Bangkitkan data berukuran n dari distribusi IG dengan parameter sebagaimana pada langkah 2 dan 3. Hitung taksiran parameter μ^* berdasarkan data yang dibangkitkan. Hitung taksiran parameter λ^* berdasarkan data yang dibangkitkan. Hitung taksiran KV^* dan simpan. Ulangi langkah 5,6,7 dan 8 sebanyak $B = 1000$ kali.

Kualitas Udara

Definisi kualitas udara bisa dipahami dari pengertian kata ‘kualitas’ dan ‘udara’. Dikutip dari Kamus Besar Bahasa Indonesia (KBBI), kualitas merupakan tingkat baik atau buruk dari suatu hal. Sedangkan udara diartikan sebagai campuran berbagai gas yang tidak berwarna dan tidak berbau, memenuhi seluruh ruang di atas bumi. Dilansir dari situs UCAR *Center for Science Education*, kualitas udara atau *air quality* merupakan kadar kandungan udara, yang didasarkan pada konsentrasi polutan di lokasi tertentu. Kualitas udara ini disesuaikan dengan Indeks Kualitas Udara atau *Air Quality Index* (AQI). Pemerintah telah menentukan Indeks Standar Pencemar Udara (ISPU) sesuai Keputusan Menteri Negara Lingkungan Hidup Nomor: KEP 45/MENLH/1997 tentang Indeks Standar Pencemar Udara.Indeks Standar Pencemar Udara (ISPU) adalah angka yang tidak mempunyai satuan yang menggambarkan kondisi kualitas udara ambien di lokasi dan waktu tertentu yang didasarkan kepada dampak terhadap kesehatan manusia, nilai estetika dan makhluk hidup lainnya. ISPU ditetapkan dengan cara mengubah kadar pencemar udara yang terukur menjadi suatu angka yang tidak berdimensi. Rentang ISPU dapat dilihat pada tabel sebagai berikut:

Tabel 1. Indeks Standar Pencemar Udara

Rentang	Kategori
1-50	Baik
51-100	Sedang
101-200	Tidak Sehat
201-300	Sangat Tidak Sehat
301+	Berbahaya

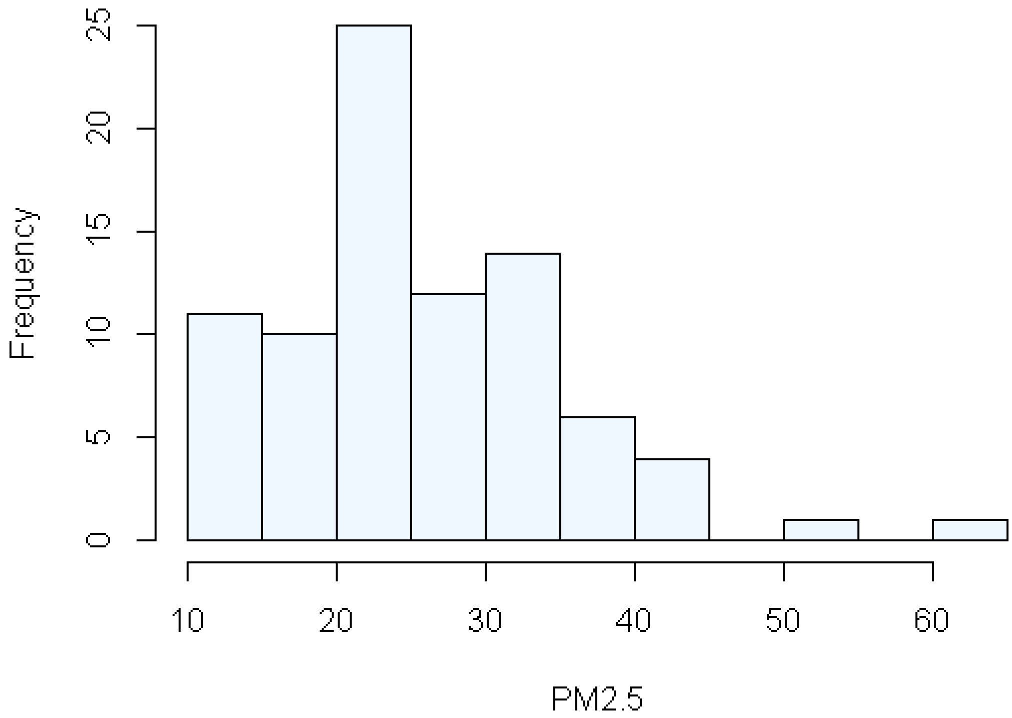
Sumber: ditppu.menlhk.go.id

PM atau *Particulate Matter* 2.5 merupakan partikel udara berukuran 2,5 mikron (mikrometer) atau lebih kecil. Tidak seperti PM 10, PM 2.5 tidak bisa dilihat dengan mata telanjang. Hal tersebut menyebabkan partikel ini dianggap sebagai polutan udara paling berbahaya. PM 2.5 mampu bertahan di udara untuk waktu yang tidak terbatas dan ketika dihirup dapat menembus jaringan paru-paru, memasuki aliran darah, menyebabkan kerusakan jantung, pernapasan, hingga otak.

C. Hasil dan Pembahasan

Sebaran Data

Berikut menampilkan histogram data kualitas udara PM 2.5 pada bulan Mei-Juli di Kota Semarang tahun 2023.



Gambar 1. Histogram Data Kualitas Udara PM 2.5

Berdasarkan histogram dapat dilihat bahwa sebaran dari distribusi datanya yaitu miring kekanan atau memiliki kemiringan positif dan dari ekor sebarannya berada di sebelah kanan. Oleh karena itu, data kualitas udara PM 2.5 pada bulan Mei-Juli di Kota Semarang tahun 2023 cocok digunakan pada distribusi IG.

Nilai Taksiran Parameter Distribusi IG

$$KV(X) = \sqrt{\frac{\hat{\mu}}{\hat{\lambda}}} = \sqrt{\frac{25,8714}{190,6387}} = 0,3684$$

Berdasarkan hasil perhitungan diatas diperoleh nilai koefisien variasi (KV) untuk data kualitas udara PM 2.5 pada bulan Mei-Juli di Kota Semarang tahun 2023 adalah 0,3684.

Uji Kecocokan Distribusi IG

Hipotesis pengujiannya sebagai berikut:

$H_0: F(x) = F^*(x)$;Data PM 2.5 Kota Semarang tahun 2023 berasal dari suatu populasi berdistribusi IG.

$H_1: F(x) \neq F^*(x)$;Data PM 2.5 Kota Semarang tahun 2023 tidak berasal dari suatu populasi berdistribusi IG.

Sebagai contoh untuk data urutan pertama yaitu $x_1 = 10,7$ dengan bantuan software Microsoft Excel diperoleh nilai fungsi distribusi empiris yaitu:

$$F_n(x_1) = F_n(10,7) = \frac{Banyaknya\ pengamatan \leq 10,7}{84}$$

$$F_n(x_1) = F_n(10,7) = \frac{1}{84} = 0,0119$$

Perhitungan untuk nilai fungsi distribusi kumulatif dari IG sebagai berikut:

$$F^*(x_1) = \Phi\left[\left(\frac{x}{\hat{\mu}} - 1\right)\left(\frac{\hat{\lambda}}{x}\right)^{\frac{1}{2}}\right] + \exp\left(\frac{2\hat{\lambda}}{\hat{\mu}}\right)\Phi\left[\left(\frac{x}{\hat{\mu}} + 1\right)\left(\frac{\hat{\lambda}}{x}\right)^{\frac{1}{2}}\right]$$

$$F^*(x_1) = \Phi\left[\left(\frac{10,7}{25,8714} - 1\right)\left(\frac{190,6387}{10,7}\right)^{\frac{1}{2}}\right] + \exp\left(\frac{2(190,6387)}{25,8714}\right)\Phi\left[\left(\frac{10,7}{25,8714} + 1\right)\left(\frac{190,6387}{10,7}\right)^{\frac{1}{2}}\right]$$

$$F^*(x_1) = 0,0097$$

Tabel 2. Nilai Fungsi Kumulatif dari IG

No	x_i	f_i	$Fkum$	$F^*(xi)$	$F_n(xi)$	$F_n(xi) - F^*(xi)$	$\lvert F_n(xi) - F^*(xi)\rvert$
1	10.7	1	1	0.0097	0.0119	0.0022	0.0022
2	11.8	1	2	0.0205	0.0238	0.0033	0.0033
3	13.4	1	3	0.0475	0.0357	-0.0118	0.0118
4	13.6	1	4	0.0519	0.0476	-0.0043	0.0043
5	14.3	2	6	0.0692	0.0714	0.0022	0.0022
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
72	63.4	1	84	0.9968	1	0.0032	0.0032

Keterangan:

x_i = Data PM 2.5 yang diurutkan dari nilai terkecil hingga terbesar dan diambil nilai yang uniknya saja.

f_i = Frekuensi atau banyaknya data ke$-i$.

$Fkum$ = Frekuensi kumulatif dari distribusi IG.
$F_n(x_i)$ = Fungsi distribusi secara empiris dari data PM 2.5 di kota Semarang.
$F^*(x_1)$ = Fungsi distribusi kumulatif dari distribusi IG.
$|F_n(x_i) - F^*(x_1)|$ = Nilai absolut dari nilai fungsi empiris dikurangi nilai fungsi distribusi kumulatif dari distribusi IG.

Diperoleh nilai statistik uji *Kolmogorov-Smirnov* menggunakan persamaan (2.9) yaitu:

$$D = \max_{1 \le i \le n} |F_n(x_i) - F^*(x_i)| = 0{,}0602$$

Dengan taraf nyata yang digunakan adalah $\alpha = 5\% = 0{,}05$ dan $n = 84$ diperoleh nilai kritis berdasarkan tabel nilai kritis uji *Kolmogorov-Smirnov* adalah 0,1484. Kriteria uji adalah H_0 diterima jika nilai D lebih kecil dari nilai kritisnya pada taraf nyata $\alpha = 5\% = 0{,}05$. Sehingga, didapat nilai D sebesar $0{,}0602 <$ $0{,}484$. Maka, disimpulkan bahwa H_0 diterima artinya data PM 2.5 di kota Semarang tahun 2023 cocok dimodelkan dengan distribusi IG.

Nilai Selang Kepercayaan untuk KV dari Distribusi IG Menggunakan Metode BPCI

Untuk menghitung nilai selang kepercayaan dilakukan dengan bantuan *software* RStudio sebagai berikut: (1) Melakukan penaksiran parameter distribusi IG yaitu μ dan λ**.** (2) Membuat fungsi dari koefisien variasi (KV) untuk X yang mengikuti distribusi IG**.** (3) Membuat fungsi dari uji kecocokan *Kolmogorov-Smirnov* untuk distribusi IG. (4) Membangkitkan data berdasarkan dari distribusi IG dengan parameter $\hat{\mu}$ dan $\hat{\lambda}$ sebagaimana yang sudah diperoleh pada langkah 2. Selanjutnya dilakukan penaksiran parameter $\hat{\mu}$, $\hat{\lambda}$ dan KV. Pembangkitan dan pendugaan parameter diulang sebanyak B = 1000 sampel *bootstrap*.

Diperoleh output untuk himpunan nilai KV sebanyak 1000 nilai sebagai berikut:

Tabel 3. Nilai $\hat{\theta}^*$ Sebelum Diurutkan

No	Nilai $\hat{\theta}^*$ Sebelum Diurutkan
1	0.3322292
2	0.4152481
3	0.3222934
.	.
.	.
.	.
1000	0.3497469

(5) Kemudian dilakukan pembuatan selang kepercayaan untuk KV dari distribusi IG menggunakan metode BPCI dengan tingkat kepercayaan 95%. Terlebih dahulu dilakukan pengurutan nilai KV dari yang terkecil hingga terbesar. (6) Menginputkan nilai $\alpha = 5\% = 0{,}05$. (7) Menghitung batas bawah dan batas atas

Diperoleh *output* nilai batas bawah dan batas atas adalah sebagai berikut:

Tabel 4. Batas Atas dan Batas Bawah

Batas Bawah	0.3108984
Batas Atas	0.4211863

D. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan dapat disimpulkan bahwa data kualitas udara PM 2.5 bulan Mei-Juli di Kota Semarang tahun 2023 cocok dimodelkan dengan distribusi Invers Gaussian(IG). Dengan tingkat kepercayaan 95% pembuatan selang kepercayaan dengan metode BPCI untuk koefisien variasi dari distribusi IG pada data kualitas udara PM 2.5 di Kota Semarang tahun 2023 adalah berada di dalam rentang [0,31 ;0,42] dengan batas bawah terletak pada persentil 25 dan batas bawah terletak pada persentil 975.

Daftar Pustaka

[1] C. J. Schwarz and M. Samanta, “An Inductive Proof of the Sampling Distributions for the MLE’s of the Parameters in an Inverse Gaussian Distribution,” Am Stat, vol. 45, no. 3, pp. 223–225, Aug. 1991, doi: 10.1080/00031305.1991.10475807.

[2] G. S. Datta and S. K. Sarkar, “A General Proof of Some Known Results of Independence Between Two Statistics,” Am Stat, vol. 62, no. 2, pp. 141–143, May 2008, doi: 10.1198/000313008X303955.

[3] J. L. Folks and R. S. Chhikara, The Inverse Gaussian Distribution. New York: Marcel Dekker, 1989.

[4] Gavriil I, Grivas G, Kassomenos P, Chaloulakou A, and Spyrellis N, “An application of theoretical probability distributions, to the study of PM10 and PM2.5 time series in Athens, Greece,” Global NEST Journal, vol. 8, no. 3, pp. 241–251, Apr. 2013, doi: 10.30955/gnj.000401.

[5] W. Chankham, S.-A. Niwitpong, and S. Niwitpong, “Measurement of Dispersion of PM 2.5 in Thailand using Confidence Intervals for The Coefficient of Variation of an Inverse Gaussian Distribution,” PeerJ, vol. 10, p. e12988, Feb. 2022, doi: 10.7717/peerj.12988.

[6] M. C. K. Tweedie, “Statistical Properties of Inverse Gaussian Distributions,” The Annals of Mathematical Statistics, vol. 28, no. 2, pp. 362–377, Jun. 1957, doi: 10.1214/aoms/1177706964.

[7] M. Bernardino, “Penerapan CUSUM-Tukey’s Control Chart untuk Mendeteksi Perubahan Rata-Rata Proses pada Data Non-Normal,” Jurnal Riset Statistika, vol. 3, no. 2, pp. 119–124, Dec. 2023, doi: 10.29313/jrs.v3i2.2955.

[8] H. J. Putri and Y. Ramdani, “Prediksi Hujan Rencana Tahunan Di Stasiun Palolo, Sulawesi Tengah menggunakan Metode Gumbel, Log Normal, Dan Log Pearson III,” DataMath: Journal of Statistics and Mathematics, vol. 2, no. 1, pp. 17–24, 2024.

[9] V. K. M. Putri and S. Gischa, “Kualitas Udara: Pengertian, Parameter, dan Cara Menjaganya,” Kompas.com. Accessed: Jul. 18, 2024. [Online]. Available: https://www.kompas.com/skola/read/2021/07/09/110000769/kualitas-udara--pengertian-parameter-dan-cara-menjaganya

[10] D. Chaniago, A. Zahara, and I. S. Ramadhani, “Indeks Standar Pencemar Udara (ISPU) sebagai Informasi Mutu Udara Ambien di Indonesia,” Kementrian Lingkungan Hidup dan Kehutanan. Accessed: Jul. 18, 2024. [Online]. Available: https://ditppu.menlhk.go.id/portal/read/indeks-standar-pencemar-udara-ispu-sebagai-informasi-mutu-udara-ambien-di-indonesia.

[11] A. K. Mutaqin, “Distribusi Invers Gaussian Sebagai Salah Satu Distribusi Kegagalan,” Statistika, vol. 1, no. 1, 2001.

[12] A. K. Mutaqin, “Penurunan Fungsi Distribusi Kumulatif dari Distribusi Invers Gaussian,” Statistika, vol. 4, no. 1, 2004.

[13] D. L. Fajri, “Mengenal PM 2.5 dan PM 10, Partikel Berbahaya bagi Tubuh,” Katadata.co.id.

[14] S. Nugroho, Statistika Nonparametrika, 1st ed. Bengkulu: UNIB Press, 2008.

[15] E. Damanik and E. Simamora, “Estimasi Interval Kepercayaan Parameter Selisih Rata-rata IPK Kelas Pendidikan Reguler dan Ekstensi FMIPA Angkatan 2010 Unimed dengan Bootstrap Pensentil,” Karismatika, vol. 5, no. 3, pp. 1–9, 2019.

[16] C. Putri, K. A. Notodiputro, and L. O. A. Rahman, “Bootstrap Parametrik dan Nonparametrik untuk Pendugaan Kuadrat Tengah Galat dalam Statistik Area Kecil dengan Respon Bersebaran Lognormal,” Forum Statistika dan Komputasi : Indonesian Journal of Statistic, vol. 18, no. 1, pp. 28–35, 2013.

[17] W. Chankham, S.-A. Niwitpong, and S. Niwitpong, "Confidence Intervals for Coefficient of Variation of Inverse Gaussian Distribution," in Proceedings of the 3rd International Conference on Vision, Image and Signal Processing, New York, NY, USA: ACM, Aug. 2019, pp. 1–6. doi: 10.1145/3387168.3387254.

[18] B. Efron and R. Tibshirani, "Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy," Statistical Science, vol. 1, no. 1, pp. 54–77, Feb. 1986, doi: 10.1214/ss/1177013815.