

LAPORAN TUGAS BESAR KOMPUTASI STATISTIK
PENERAPAN METODE PEMBANGKIT BILANGAN ACAK UNTUK
PEMODELAN UMUR DETEKSI DAN ASAL PENYANDANG
THALASSEMIA DI PROVINSI LAMPUNG



Disusun oleh:

1. Anadia Carana (123450019)
2. Fadil Prasetyo Alfarizzi (123450048)
3. Muhammad Hanif Dzaky (123450064)
4. Wan Nashwa Alhasni Yuska (123450077)

PROGRAM STUDI SAINS DATA
FAKULTAS SAINS
INSTITUT TEKNOLOGI SUMATERA
2025

ABSTRAK

Penelitian ini bertujuan memodelkan umur saat terdeteksi dan asal daerah penyandang Thalassemia di Provinsi Lampung dengan menggunakan metode pembangkit bilangan acak untuk menghasilkan data sintetis yang merepresentasikan pola data riil. Data primer dikumpulkan melalui kuesioner Google Form dari penyandang Thalassemia atau keluarganya, kemudian dibersihkan dan dianalisis untuk memperoleh parameter deskriptif umur serta probabilitas empiris asal kabupaten/kota. Algoritma Linear Congruential Generator (LCG) diimplementasikan secara manual dalam bahasa R untuk membangkitkan bilangan acak berdistribusi seragam $U(0,1)$, yang selanjutnya ditransformasikan menjadi variabel umur (kontinu) menggunakan Metode Transformasi Invers Empiris berbasis CDF umur deteksi, dan variabel asal daerah (diskrit) menggunakan Inverse Transform Diskrit berdasarkan probabilitas kumulatif masing-masing wilayah. Hasil simulasi $N = 5.000$ data sintetis menunjukkan bahwa distribusi umur terdeteksi yang dihasilkan memiliki bentuk sebaran yang sangat berimpit dengan data asli yang bersifat miring ke kanan dengan mayoritas kasus terdeteksi pada usia dini, sedangkan distribusi asal daerah berhasil mempertahankan dominasi kategori Bandar Lampung dan proporsi relatif wilayah lainnya. Secara keseluruhan, dataset sintetis yang dibangun mampu merepresentasikan pola distribusi data riil sehingga berpotensi dimanfaatkan sebagai alternatif data dalam perencanaan layanan kesehatan dan penelitian lanjutan terkait epidemiologi Thalassemia di Provinsi Lampung.

Keyword: Thalassemia, data sintetis, Linear Congruential Generator, Metode Transformasi Invers, distribusi empiris, Provinsi Lampung

BAB I

PENDAHULUAN

1.1. Latar Belakang

Thalassemia adalah kelainan genetik yang mempengaruhi sintesis hemoglobin, menyebabkan anemia kronis dan berbagai komplikasi kesehatan serius. Menurut studi epidemiologi, Indonesia terletak di "Thalassemia Belt" dengan prevalensi pembawa sifat gen β -thalassemia sebesar 3–10% populasi, serta diperkirakan tiap tahun lahir sekitar 2.500 bayi baru dengan thalassemia mayor. Kondisi ini berdampak besar pada kesehatan masyarakat, mendatangkan beban ekonomi dan sosial yang tinggi bagi keluarga maupun sistem layanan kesehatan[1].

Peningkatan jumlah kasus thalassemia Indonesia juga tercermin secara nasional setiap tahunnya. Studi kualitas hidup anak Indonesia dengan thalassemia menunjukkan banyak pasien membutuhkan transfusi rutin dan perawatan lanjutan, sementara akses ke terapi kelasi besi yang optimal masih terbatas. Dengan tren kasus yang terus meningkat, upaya skrining dini dan penyebaran pemahaman epidemiologi thalassemia di seluruh wilayah sangat penting agar pencegahan dan penanganan dapat berjalan efektif [2].

Salah satu tantangan terbesar dalam penelitian thalassemia adalah keterbatasan data primer yang dapat diakses, baik karena regulasi privasi maupun fragmentasi layanan kesehatan. Pembangkitan data sintetis (synthetic data generation) berbasis algoritma acak di bidang medis menjadi solusi inovatif, karena dapat menggantikan data riil untuk berbagai analisis dan prediksi tanpa berisiko mengungkap data sensitif pasien. Data sintetis didesain sehingga properti statistik dan distribusinya tetap mencerminkan populasi asli, sehingga sangat berguna untuk validasi model klinis dan riset kesehatan berbasis skala besar[3].

Dalam proses pembangkitan data acak, Linear Congruential Generator (LCG) adalah salah satu algoritma yang populer digunakan di simulasi data medis, karena mampu menghasilkan urutan bilangan acak yang cukup stabil dan representatif untuk pemodelan statistik berbasis simulasi. Studi membuktikan bahwa kualitas LCG dapat mempengaruhi hasil analisis statistik dan simulasi biologis, sehingga pemilihan dan pengujian PRNG menjadi aspek krusial dalam penelitian berkualitas tinggi[4].

Dengan demikian, riset di bidang thalassemia yang menggunakan algoritma pembangkitan data acak seperti LCG dan model distribusi empiris dapat memberikan pemahaman baru terhadap pola umur deteksi dan persebaran wilayah asal penderita di Lampung. Seluruh pendekatan ini berkontribusi dalam pengembangan strategi penanggulangan thalassemia berbasis data, memperkaya literatur epidemiologi, dan mendukung kebijakan layanan kesehatan Indonesia secara menyeluruh.[5]

1.2. Rumusan Masalah

Berdasarkan uraian pada latar belakang, maka permasalahan yang akan dikaji dalam penelitian ini dapat dirumuskan sebagai berikut:

1. Bagaimana distribusi probabilitas variabel umur deteksi dan asal daerah penyandang Thalassemia di Provinsi Lampung berdasarkan data?
2. Bagaimana mengimplementasikan algoritma Linear Congruential Generator (LCG) untuk membangkitkan bilangan acak berdistribusi Uniform $U(0,1)$?
3. Bagaimana penerapan Metode Transformasi Invers untuk membangkitkan data simulasi umur (kontinu) dan asal daerah (diskrit) sesuai karakteristik data asli?
4. Apakah dataset simulasi yang dihasilkan mampu merepresentasikan pola distribusi yang mendekati data riil?

1.3. Tujuan Penelitian

Berdasarkan rumusan masalah yang telah disusun, tujuan penelitian ini adalah sebagai berikut:

1. Mengidentifikasi parameter data meliputi karakteristik statistik deskriptif dari data riil penyandang Thalassemia, termasuk distribusi frekuensi kumulatif (CDF) untuk variabel umur serta proporsi probabilitas untuk variabel asal daerah.
2. Mengimplementasikan algoritma pembangkit bilangan acak, yaitu Metode Kongruensi Campuran, secara manual menggunakan bahasa pemrograman R untuk menghasilkan bilangan acak berdistribusi seragam $U(0,1)$.

3. Menerapkan Metode Transformasi Invers untuk mensimulasikan variabel umur dan asal daerah melalui konversi bilangan acak ke nilai umur berdasarkan CDF empiris serta pemetaan ke kategori wilayah menggunakan interval probabilitas kumulatif.
4. Melakukan validasi model dengan menilai kesesuaian pola distribusi dataset sintetis yang dihasilkan terhadap karakteristik data asli.

1.4. Manfaat Penelitian

Penelitian ini diharapkan memberikan kontribusi dalam aspek teoritis maupun praktis sebagai berikut:

a. Manfaat Teoritis

1. Memberikan penguatan pemahaman mengenai konsep algoritmik yang mendasari pembangkit bilangan acak, termasuk mekanisme matematis pada Metode Kongruensi Campuran yang menjadi dasar fungsi *runif* dalam R.
2. Menunjukkan aplikasi komputasi statistika melalui penerapan simulasi Pembangkit Bilangan Acak dan teknik *Inverse Transform Sampling* untuk pemodelan data yang melibatkan variabel kontinu dan diskrit.

b. Manfaat Praktis

1. Menyediakan pendekatan alternatif untuk mengatasi keterbatasan akses terhadap data sensitif dengan menghasilkan dataset sintetis yang tetap merepresentasikan pola riil namun tidak mengungkap informasi pribadi pasien.
2. Memberikan dukungan dalam pemetaan kondisi kesehatan, khususnya berupa gambaran simulatif distribusi umur dan asal daerah penyandang Thalassemia yang dapat dimanfaatkan oleh instansi terkait dalam perencanaan distribusi layanan, obat, serta fasilitas perawatan secara lebih efektif.

BAB II

TINJAUAN PUSTAKA

2.1. Thalassemia

Thalassemia adalah kelainan genetik yang diturunkan, yang memengaruhi produksi hemoglobin, protein pembawa oksigen dalam sel darah merah. Kelainan ini disebabkan oleh mutasi pada gen globin, yang menghasilkan sintesis rantai globin (alfa atau beta) yang kurang atau tidak berfungsi secara normal [5]. Kondisi ini menyebabkan sel darah merah rapuh dan hancur terlalu cepat (*hemolisis*), yang berujung pada kondisi anemia kronis. Secara klinis, Thalassemia diklasifikasikan berdasarkan rantai globin yang terpengaruh (Alfa atau Beta) dan tingkat keparahannya (minor, intermedia, atau mayor) [5].

Dalam konteks komputasi statistika dan pemodelan, data terkait umur saat terdeteksi dan asal penyandang Thalassemia merupakan peubah acak penting. Distribusi empiris dari peubah ini digunakan sebagai dasar untuk simulasi data sintetis, yang bertujuan untuk membantu estimasi beban kasus dan perencanaan layanan kesehatan, sebagaimana diterapkan dalam studi ini. Akurasi pemodelan bergantung pada kemampuan untuk mereplikasi karakteristik distribusi empiris ini melalui teknik pembangkitan bilangan acak.

2.2. Pembangkitan Bilangan Acak Semu (Pseudo-Random Number Generator, PRNG)

Pembangkit Bilangan Acak Semu (PRNG) adalah algoritma deterministik yang dirancang untuk menghasilkan urutan bilangan yang terlihat acak, tetapi sebenarnya dapat direproduksi jika kondisi awalnya (*seed*) diketahui [6]. Meskipun urutan yang dihasilkan tidak sepenuhnya acak (*random*), bilangan tersebut harus memenuhi properti statistik tertentu, seperti distribusi seragam (uniform) pada interval (0, 1), dinotasikan $U \sim U(0, 1)$, dan memiliki korelasi yang rendah di antara bilangan yang berurutan [7].

Kualitas suatu PRNG ditentukan oleh panjang periode (jumlah bilangan yang dihasilkan sebelum urutan tersebut berulang), kecepatan komputasi, dan kemampuan untuk lolos dalam uji

keacakan statistik. Algoritma yang baik harus memiliki periode yang sangat panjang untuk menjamin bahwa urutan tidak berulang selama simulasi berlangsung [6].

2.3. Linear Congruential Generator (LCG)

Linear Congruential Generator (LCG) adalah salah satu algoritma PRNG tertua dan paling dasar yang digunakan secara luas karena kesederhanaannya [7]. LCG menghasilkan urutan bilangan bulat X_n yang kemudian dinormalisasi untuk mendapatkan bilangan seragam U_n .

Relasi rekursif LCG didefinisikan sebagai berikut:

$$X_{n+1} = (aX_n + c) \bmod m, n \geq 0 \quad (1)$$

Di mana:

- X_{n+1} : Bilangan acak pada iterasi berikutnya.
- X_n : Bilangan acak saat ini.
- a : Pengganda (Multiplier).
- c : Inkremen (Increment).
- m : Modulus.
- X_0 : Nilai awal (Seed).

Urutan bilangan seragam $U_n \sim U(0, 1)$ diperoleh melalui:

$$U_n = \frac{X_n}{m} \quad (2)$$

Periode LCG dapat mencapai nilai maksimum m (disebut periode penuh) jika parameter a , c , dan m memenuhi kriteria tertentu. Dalam studi ini, LCG digunakan sebagai generator manual untuk membandingkan kinerja dengan generator bawaan R, seperti runif.

2.4. Metode Inverse Transform

Metode Inverse Transform (Inverse Transform Method) adalah teknik fundamental untuk membangkitkan peubah acak X dengan Fungsi Distribusi Kumulatif (CDF), $F_x(x)$, yang telah ditentukan [8]. Metode ini didasarkan pada prinsip bahwa jika U adalah peubah acak seragam pada $(0, 1)$, maka $F_x^{-1}(U)$ akan memiliki CDF $F_x(x)$.

Langkah-langkah Metode Inverse Transform:

1. Tentukan CDF target, $F_x(x)$, dan turunkan fungsi inversnya, $F_x^{-1}(U)$
2. Bangkitkan bilangan acak seragam $U \sim U(0, 1)$ menggunakan PRNG (misalnya, LCG atau runif).
3. Hitung peubah acak X yang diinginkan dengan:

$$X = F_x^{-1}(U) \quad (3)$$

Dalam konteks simulasi data Thalassemia, metode ini diterapkan secara non-parametrik, yaitu sebagai Metode Inverse Transform Empiris. Dengan menggunakan Fungsi Distribusi Kumulatif Empiris $F_n(x)$ dari data observasi (umur deteksi), metode ini menghasilkan sampel sintetis yang secara efektif mereplikasi bentuk distribusi probabilitas yang diamati, tidak hanya rentang datanya saja [8].

2.5. Sampling Diskrit Berdasarkan Probabilitas Empiris

Sampling Diskrit Berdasarkan Probabilitas Empiris adalah metode simulasi yang digunakan untuk membangkitkan peubah acak diskrit atau kategorikal (seperti wilayah asal, jenis kelamin, atau status) sedemikian rupa sehingga distribusi probabilitas kategori dalam data simulasi secara akurat mencerminkan frekuensi relatif yang diamati dalam data asli (empiris) [8]. Teknik ini pada dasarnya adalah aplikasi dari Metode Inverse Transform untuk peubah acak diskrit.

Mekanisme Pembangkitan:

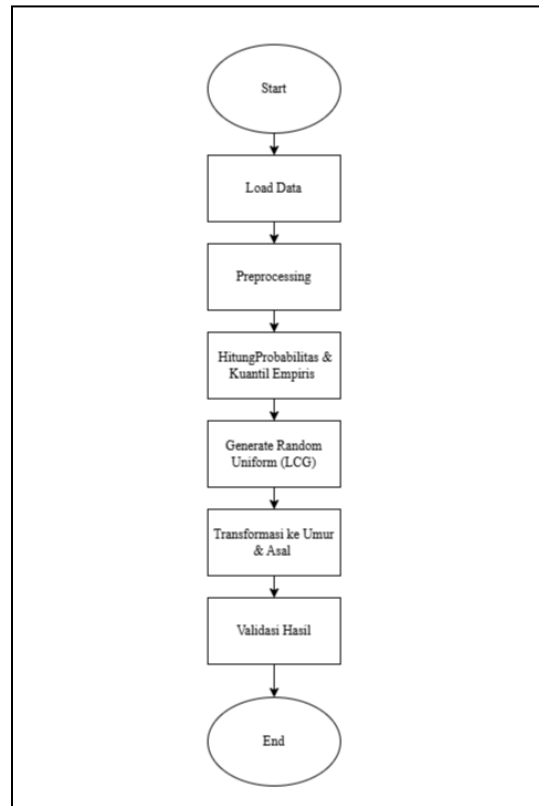
1. Estimasi Probabilitas Empiris: Hitung frekuensi relatif (probabilitas) p_i untuk setiap kategori diskrit k_i dari data observasi. Dalam studi kasus ini, ini adalah probabilitas $P(\text{Asal}=\text{Wilayah})$ dari data Thalassemia.
2. Konstruksi Fungsi Distribusi Kumulatif (CDF) Diskrit: Probabilitas p_i diakumulasikan secara berurutan, $F(k_i) = \sum_{j=1}^i p_j$, untuk membagi interval $(0, 1)$ menjadi segmen-segmen kumulatif.
3. Pemetaan Bilangan Acak: Sebuah bilangan acak seragam $U \sim U(0, 1)$ dibangkitkan. Kategori k_i dipilih (dibangkitkan) jika U jatuh ke dalam segmen kumulatif kategori tersebut, yang dapat dirumuskan sebagai:
Pilih k_i jika $F(k_{i-1}) < U \leq F(k_i)$

Metode ini memastikan bahwa kategori dengan proporsi terbesar (misalnya, 'Bandar Lampung' dalam variabel asal penyandang) akan memiliki segmen terpanjang dalam interval $(0, 1)$ dan dengan demikian memiliki probabilitas paling tinggi untuk dipilih dalam setiap iterasi simulasi. Teknik ini krusial untuk menghasilkan data sintetis yang mempertahankan struktur dan proporsi kategorikal data asli.

BAB III

METODE PENELITIAN

3.1. Diagram Alir Penelitian



Gambar 1. Diagram Alir

3.2. Jenis dan Teknik Pengumpulan Data

Penelitian ini menggunakan data primer, yaitu data yang dikumpulkan secara langsung melalui kuesioner Google Form. Responden merupakan penyandang Thalasemia atau pihak keluarga yang mengetahui informasi terkait, sesuai dengan kebutuhan analisis variabel umur deteksi dan asal daerah.

Pengumpulan data dilakukan secara daring dengan menyebarkan tautan Google Form melalui komunitas pasien, fasilitas kesehatan, serta jaringan komunikasi masyarakat terkait. Setiap responden memberikan persetujuan pengisian (informed consent) sebelum mengisi kuesioner.

3.3. Variabel Penelitian

Variabel yang digunakan dalam penelitian ini terdiri atas:

- Umur Saat Terdeteksi (Tahun)
 - Jenis: Kuantitatif kontinu
 - Diperoleh melalui pertanyaan yang meminta usia responden/pasien pada saat pertama kali didiagnosis Thalassemia.
- Asal Kabupaten/Kota
 - Jenis: Kualitatif diskrit (kategori)
 - Diperoleh melalui pilihan wilayah asal responden sesuai kabupaten/kota di Provinsi Lampung.

3.4. Prosedur Penelitian

1. Pengumpulan dan Pembersihan Data Primer
 - Data diunduh dari Google Form dalam format CSV.
 - Kolom umur dikonversi menjadi numerik, dan entri yang tidak logis (umur ≤ 0 atau > 100 tahun) dihapus.
 - Kolom asal daerah diseragamkan (penulisan huruf, karakter khusus, spasi) untuk memastikan konsistensi kategori wilayah.
2. Pembentukan Distribusi Empiris
 - Statistik deskriptif (min, kuartil, median, mean, dan max) dihitung untuk variabel umur.
 - Frekuensi relatif dihitung untuk mendapatkan probabilitas empiris setiap kategori asal kabupaten/kota.
3. Pembangkitan Bilangan Acak
 - Diimplementasikan Linear Congruential Generator (LCG) secara manual dengan parameter (a,c,m) tertentu untuk menghasilkan bilangan uniform $U(0,1)$
 - Digunakan pula pembangkit bawaan R yaitu `runif()` untuk keperluan simulasi yang lebih stabil.
4. Simulasi Variabel Umur (Kontinu)

Bilangan uniform yang dihasilkan selanjutnya ditransformasikan menjadi nilai umur simulatif menggunakan:

- Transformasi linear (min–max) untuk ilustrasi, dan
- Metode Transformasi Invers, yaitu pemetaan nilai uniform ke kuantil umur berdasarkan CDF empiris hasil pengisian Google Form.

5. Simulasi Variabel Asal Daerah (Diskrit)

Pembangkitan asal wilayah dilakukan melalui:

- *Sampling* dengan probabilitas empiris tiap wilayah, dan
- Inverse Transform Diskrit, yaitu mencocokkan bilangan uniform dengan interval probabilitas kumulatif masing-masing kategori.

6. Validasi Dataset Simulatif

- Dataset sintetis yang dihasilkan dibandingkan dengan data primer berdasarkan pola distribusi umur dan proporsi wilayah.
- Validasi dilakukan menggunakan perbandingan histogram, kuantil, serta persentase kategori untuk menilai kesesuaian pola distribusi.

BAB IV

HASIL DAN PEMBAHASAN

4.1. Deskripsi Data Awal

Ringkasan statistik deskriptif untuk variabel kontinu Umur saat Terdeteksi disajikan pada tabel berikut, yang berfungsi sebagai dasar dalam memahami pola umum distribusi usia diagnosis pada penyandang Thalassemia:

Tabel 1. Statistika Deskriptif Data

Statistik	Nilai
Min	0,1667 Tahun
Kuartil 1 (Q1)	1 Tahun
Median	3 Tahun
Rata-Rata (Mean)	5,0948 Tahun
Kuartil 2 (Q3)	6 Tahun
Max	55 Tahun

Berdasarkan tabel tersebut, dapat disimpulkan bahwa sebagian besar kasus Thalassemia terdeteksi pada usia sangat muda, yang ditunjukkan oleh nilai median sebesar 3,00 tahun. Artinya, setengah dari keseluruhan responden telah mendapatkan diagnosis sebelum mencapai usia tiga tahun.

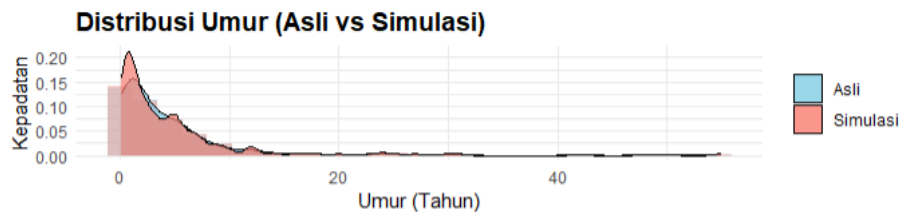
Selain itu, nilai rata-rata sebesar 5,0948 tahun yang lebih tinggi daripada median mengindikasikan bahwa distribusi umur deteksi bersifat miring ke kanan (*positively skewed*). Rentang nilai yang luas, dengan usia maksimum mencapai 55 tahun, menunjukkan adanya sejumlah kasus yang baru teridentifikasi pada usia dewasa hingga lanjut. Meskipun demikian, frekuensinya jauh lebih kecil dibandingkan kelompok usia anak-anak.

4.2. Hasil Simulasi dan Validasi

Pembangkitan $N=5.000$ data pasien sintetis menggunakan bilangan acak seragam $U \sim (0, 1)$ yang dihasilkan oleh *Linear Congruential Generator* (LCG). Validasi dilakukan melalui perbandingan distribusi variabel target (Umur Saat Terdeteksi dan Asal Daerah) antara data hasil simulasi dan data observasi empiris asli.

4.2.1. Pemodelan Variabel Umur (Kontinu)

Variabel Umur Saat Terdeteksi yang berdistribusi kontinu dievaluasi menggunakan dua pendekatan transformasi dari U : Transformasi Linear (Min-Max Scaling) dan Transformasi Invers Empiris.



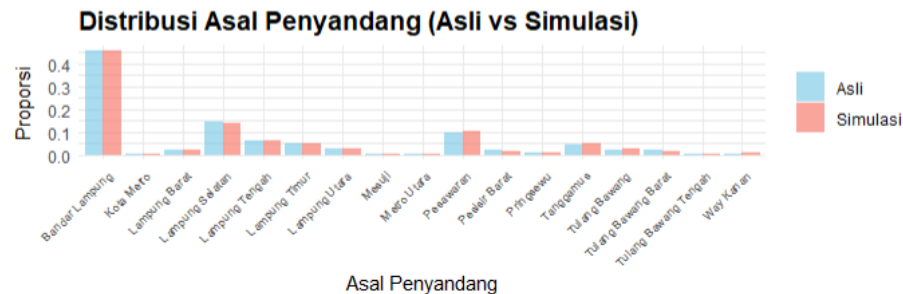
Gambar 2. Hasil Visualisasi Distribusi Umur (Asli vs Simulasi)

Simulasi membuktikan bahwa hanya Metode Inverse Transform Empiris yang berhasil mereplikasi bentuk distribusi umur asli. Metode ini memanfaatkan fungsi kuantil empiris (CDF invers) data umur asli, $X = F^{-1}(U)$, sehingga distribusi sampel simulasi akan mempertahankan karakteristik statistik dan bentuk kepadatan data observasi. Sebaliknya, transformasi linear hanya meniru rentang umur tanpa mereplikasi kepadatan (bentuk).

Kurva kepadatan data umur asli menunjukkan pola highly skewed ke kanan, dengan puncak kepadatan yang sangat tinggi pada usia dini (median 3 tahun, kuartil atas 6 tahun). Perbandingan kurva kepadatan menunjukkan bahwa kurva simulasi (merah) yang dibangkitkan dari LCG dan inverse transform sangat berimpit dengan kurva data asli (biru). Kecocokan ini memvalidasi bahwa Metode Inverse Transform Empiris efektif untuk mensimulasikan umur terdeteksi Thalassemia.

4.2.2. Pemodelan Variabel Asal Daerah

Pemodelan variabel Asal Kabupaten/Kota (kategori diskrit) dilakukan menggunakan teknik *Inverse Transform Diskrit* berdasarkan probabilitas empiris dari frekuensi data asli.



Gambar 3. Hasil Visualisasi Distribusi Asal Penyandang (Asli vs Simulasi)

Distribusi wilayah asli menunjukkan dominasi Bandar Lampung (46%). Perbandingan diagram batang proporsi wilayah menunjukkan bahwa teknik sampling kategori berbasis probabilitas empiris mengonservasi proporsi relatif antar-wilayah secara akurat.

- **Dominasi Wilayah:** Wilayah Bandar Lampung tetap menjadi kategori yang dominan dalam data simulasi, mempertahankan proporsi aslinya.
- **Kesesuaian Proporsi:** Proporsi wilayah-wilayah dengan frekuensi menengah (seperti Lampung Selatan dan Pesawaran) juga terjaga kesamaannya dengan data observasi.

BAB V

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan hasil implementasi algoritma pembangkit bilangan acak dan validasi model, diperoleh kesimpulan sebagai berikut:

1. Distribusi Data Asli: Distribusi probabilitas variabel umur deteksi Thalassemia memiliki karakteristik miring ke kanan (*positively skewed*) dengan mayoritas kasus terdeteksi pada usia sangat muda (Median = 3 tahun). Distribusi variabel asal daerah bersifat diskrit dan sangat didominasi oleh satu kategori utama (Bandar Lampung).
2. Implementasi LCG: Algoritma *Linear Congruential Generator* (LCG) berhasil diimplementasikan secara manual untuk menghasilkan bilangan acak seragam $U \sim (0, 1)$ yang menjadi variabel dasar simulasi.
3. Penerapan Metode Transformasi Invers:
 - Variabel Umur (Kontinu): Penerapan Metode Transformasi Invers Empiris((($X = F^{-1}(U)$)) terbukti efektif dan superior dibandingkan transformasi linear, menghasilkan distribusi simulasi yang sangat berimpit dengan pola *skewness* distribusi umur asli.
 - Variabel Asal Daerah (Diskrit): Penerapan Sampling Invers Diskrit berdasarkan probabilitas empiris berhasil mengonservasi proporsi relatif kategori wilayah, menjaga dominasi wilayah frekuensi tinggi.
4. Validasi Model: Dataset sintetis yang dihasilkan (berjumlah $N=5.000$) mampu merepresentasikan pola distribusi yang mendekati data riil, yang divalidasi melalui kecocokan kurva kepadatan (umur) dan proporsi kategori (asal daerah).

5.2. Saran

Berdasarkan keterbatasan dan potensi pengembangan hasil penelitian ini, diajukan beberapa saran sebagai berikut:

1. Penggunaan Generator Modern: Untuk aplikasi praktis di masa depan, direkomendasikan untuk menggunakan generator bilangan acak bawaan R (*runif*) yang lebih stabil dan memiliki periode yang jauh lebih panjang, meskipun LCG manual telah memenuhi tujuan ilustrasi dalam studi ini.
2. Penghalusan Probabilitas Kategori: Terkait dengan sensitivitas simulasi pada kategori wilayah dengan frekuensi sangat kecil (misalnya, hanya 1 kasus), disarankan untuk melakukan agregasi wilayah atau menerapkan penghalusan probabilitas (misalnya, menggunakan pendekatan Bayesian) sebelum proses *sampling* diskrit untuk meningkatkan stabilitas estimasi pada data langka.
3. Pengembangan Model: Penelitian selanjutnya dapat mengembangkan pemodelan ini dengan mempertimbangkan hubungan multivariat atau variabel acak lain (seperti distribusi jenis kelamin atau status sosial) serta menguji metode pembangkitan lain seperti *Acceptance-Rejection* atau metode transformasi spesifik distribusi.
4. Pemanfaatan Data Sintetis: Hasil simulasi ini dapat digunakan lebih lanjut oleh instansi terkait sebagai alat pendukung keputusan dan perencanaan distribusi layanan kesehatan (seperti obat dan fasilitas) secara lebih efektif di Provinsi Lampung.

Daftar Pustaka

- [1] P. A. Wahidiyat *et al.*, "Thalassemia in Indonesia," *Hemoglobin*, vol. 46, no. 1, pp. 39–44, Jan. 2022, doi: 10.1080/03630269.2021.2023565.
- [2] R. D. Wulandari and Bagian Biomedik Penelitian Biomolekuler Fakultas Kedokteran Universitas Wijaya Kusuma Surabaya, "Kelainan pada Sintesis Hemoglobin: Thalassemia dan Epidemiologi Thalassemia," *Jurnal Ilmiah Kedokteran Wijaya Kusuma*, vol. 5, no. 2, pp. 33–43, 2019.
- [3] Qian, Z., Callender, T., Cebere, B. *et al.* Synthetic data for privacy-preserving clinical risk prediction. *Sci Rep* 14, 25676 (2024).
- [4] R. Rahmah and S. N. N. Makiyah, "Quality Of Life of Children with Thalassemia in Indonesia: Review," *IJNP (Indonesian Journal of Nursing Practices)*, vol. 6, no. 1, pp. 1–7, Jun. 2022, doi: 10.18196/ijnp.v6i1.10477.
- [5] D. J. Weatherall dan J. B. Clegg, "The Thalassemia Syndromes: Genetic Basis, Clinical Manifestations, and Therapy," *Nature Medicine*, vol. 2, no. 8, pp. 847–852, Agt. 1996.
- [6] D. C. Montgomery and G. C. Runger, *Applied Statistics and Probability for Engineers*, 6th ed. New York: Wiley, 2014.
- [7] P. L'Ecuyer, "Software for Uniform Random Number Generation," *Proc. Winter Simulation Conf.*, pp. 197–205, 2018.
- [8] B. L. Nelson, *Stochastic Modeling: Analysis & Simulation*. Dover Publications, 2019.