

**Pembangkitan Bilangan Acak Pada Studi Kasus  
Frekuensi Pengunjung Perpustakaan GKU 1 ITERA Pada Hari Kerja  
Laporan Tugas Besar Komputasi Statistik**



**Oleh:  
Kelompok 2 RA**

Pramudya Wibowo	(121450030)
Tanty Widiyastuti	(123450094)
Ihsan Maulana Yusuf	(123450110)
Rahma Oktavia Albar	(123450003)

**Dosen pengampu :**  
Mika Alvionita S, M.Si  
Yoga Aji Sukma, S.Mat., M.Stat.,  
Yustida Bellini, M.Kom

**PROGRAM STUDI SAINS DATA  
FAKULTAS SAINS  
INSTITUT TEKNOLOGI SUMATERA  
2025**

## ABSTRAK

Penelitian ini membahas proses pembangkitan bilangan acak pada studi kasus frekuensi pengunjung Perpustakaan GKU 1 ITERA pada hari kerja. Data kunjungan dianalisis melalui statistik deskriptif dan pengujian distribusi untuk menentukan model probabilistik yang paling sesuai. Hasil analisis menunjukkan bahwa data memiliki karakteristik *overdispersion*, di mana varians jauh lebih besar daripada rata-rata, sehingga distribusi Poisson dan Binomial tidak cocok untuk menggambarkan pola kunjungan. Melalui *Likelihood Ratio Test*, AIC, BIC, dan uji Kolmogorov–Smirnov diskrit, distribusi Negatif Binomial terbukti paling sesuai sebagai model distribusi utama. Selanjutnya, tiga metode pembangkitan bilangan acak digunakan untuk mensimulasikan pola kunjungan, yaitu *Inverse Transform Method*, *Acceptance-Rejection Method*, dan *Direct Simulation Method*. Evaluasi akurasi terhadap sampel besar ( $n=10.000$ ) menunjukkan bahwa metode *Inverse Transform* menghasilkan nilai rata-rata dan varians yang paling mendekati data asli, sehingga menjadi metode paling stabil dan akurat. Sementara itu, *Acceptance-Rejection* menunjukkan performa yang cukup baik tetapi kurang efisien, dan *Direct Simulation* memberikan deviasi lebih besar karena bergantung pada pembulatan parameter  $r$ . Secara keseluruhan, penelitian ini menyimpulkan bahwa distribusi Negatif Binomial merupakan representasi terbaik dari frekuensi kunjungan perpustakaan, dan metode *Inverse Transform* adalah pendekatan optimal untuk pembangkitan bilangan acak pada distribusi tersebut. Hasil ini dapat menjadi dasar dalam perencanaan layanan perpustakaan, estimasi kebutuhan operasional, serta pengembangan model simulasi kunjungan pada institusi pendidikan.

**Kata Kunci :** *Pembangkitan bilangan acak; Distribusi Negatif Binomial; Inverse Transform; Acceptance-Rejection; Simulasi statistik.*

## DAFTAR ISI

ABSTRAK.....	ii
DAFTAR ISI.....	iii
BAB I.....	1
PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	1
1.3 Tujuan Penelitian.....	2
1.4 Manfaat Penelitian.....	2
BAB II.....	3
TINJAUAN PUSTAKA .....	3
2.1 Konsep Pembangkitan Bilangan Acak.....	3
2.2 Distribusi Negatif Binomial .....	3
2.3 Inverse Transform Method.....	3
2.4 Acceptance Rejection Method .....	4
2.5 Simulasi Langsung (Direct Simulation / Bernoulli Trials) .....	4
BAB III .....	6
METODE PENELITIAN.....	6
3.1 Jenis Data .....	6
3.2 Teknik Pengumpulan Data .....	6
3.3 Variabel yang Diamati .....	6
3.4 Diagram Alir Penelitian .....	7
BAB IV .....	10
HASIL DAN PEMBAHASAN.....	10
4.1 Deskripsi Data Penelitian.....	10
4.1.1 Visualisasi Data Awal .....	10
4.1.2 Statistik Deskriptif.....	12
4.2 Analisis Distribusi Data .....	13
4.2.1 Uji Goodness-of-Fit (GoF) .....	13
4.2.2 Pemilihan Model Terbaik .....	14

4.3 Pembangkitan Data (Simulasi).....	14
4.3.1 Metode Inverse Transform .....	14
4.3.2 Metode Acceptance-Rejection.....	16
4.3.3 Metode Direct Transformation .....	17
4.4 Perbandingan Akurasi Metode Pembangkitan Data.....	18
4.5 Diskusi.....	20
BAB V .....	21
KESIMPULAN DAN SARAN.....	21
5.1 Kesimpulan.....	21
5.2 Saran.....	22
DAFTAR PUSTAKA .....	23

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Perpustakaan merupakan sarana penting dalam lingkungan perguruan tinggi karena berperan sebagai pusat informasi dan referensi bagi mahasiswa maupun dosen. Di Institut Teknologi Sumatera (ITERA), khususnya pada Perpustakaan GKU 1, jumlah kunjungan mahasiswa dapat berubah-ubah setiap hari. Namun, pola kedatangan pengunjung pada hari kerja sering kali bersifat stokastik (acak) dan dipengaruhi oleh berbagai variabel eksternal yang kompleks. Untuk mengatasi ketidakpastian ini, pemodelan berbasis simulasi menjadi pendekatan yang esensial dalam memprediksi skenario operasional dan mengoptimalkan layanan publik [1]

Kualitas sebuah model simulasi sangat ditentukan oleh akurasi teknik Pembangkitan Bilangan Acak (*Random Variate Generation*). Proses ini merupakan jantung dari simulasi statistik, di mana tujuannya adalah mereproduksi data sintetik yang memiliki karakteristik distribusi probabilitas identik dengan data observasi [2]. Jika algoritma pembangkitan yang digunakan tidak tepat, simulasi akan menghasilkan bias yang signifikan, sehingga keputusan yang diambil seperti penjadwalan staf atau penambahan kapasitas kursi menjadi tidak valid. Dalam konteks data frekuensi kunjungan yang memiliki fluktuasi tinggi, tantangan komputasi muncul dalam memilih algoritma pembangkitan bilangan acak yang paling efektif. Literatur komputasi statistik modern menyoroti tiga metode fundamental: Metode Transformasi Invers (*Inverse Transform Method*), Metode Penolakan (*Acceptance-Rejection Method*), dan Simulasi Langsung (*Direct Simulation*).

Penelitian ini bertujuan untuk menguji secara empiris kinerja ketiga teknik tersebut dalam memodelkan data pengunjung Perpustakaan GKU 1 ITERA. Dengan menggunakan perangkat lunak R dan pendekatan ilmu data terkini, penelitian ini akan membandingkan validitas statistik dan efisiensi algoritma dari masing-masing metode. Hal ini sejalan dengan pandangan Dobson dan Barnett (2021) bahwa penanganan data dengan variabilitas tinggi memerlukan pemilihan model statistik yang presisi agar interpretasi data menjadi akurat [3]

### 1.2 Rumusan Masalah

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

1. Bagaimana implementasi algoritma *Inverse Transform*, *Acceptance-Rejection*, dan *Direct Simulation* dalam membangkitkan bilangan acak berbasis data pengunjung Perpustakaan GKU 1 ITERA?
2. Apakah data simulasi yang dihasilkan oleh ketiga metode tersebut memenuhi syarat validitas statistik berdasarkan uji kesesuaian distribusi (*Goodness-of-Fit*)?
3. Metode manakah yang menunjukkan performa terbaik dari segi akurasi (*error rate*) dan efisiensi komputasi untuk direkomendasikan dalam simulasi sistem perpustakaan?

### 1.3 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

1. Membangun model simulasi pembangkitan bilangan acak menggunakan teknik *Inverse Transform*, *Acceptance-Rejection*, dan *Direct Simulation* yang disesuaikan dengan parameter data perpustakaan.
2. Memvalidasi hasil bangkitan bilangan acak menggunakan uji statistik modern untuk memastikan konsistensi dengan pola data riil.
3. Mengevaluasi perbandingan kinerja ketiga metode guna menentukan algoritma yang paling optimal untuk diterapkan pada studi kasus Perpustakaan GKU 1 ITERA.

### 1.4 Manfaat Penelitian

Penelitian ini diharapkan memberikan beberapa manfaat, antara lain:

1. Bagi Penulis : Memberikan pengalaman dan pemahaman yang lebih mendalam terkait penerapan teori peluang, distribusi statistik, serta teknik simulasi random number generation dalam konteks permasalahan nyata.
2. Bagi Instansi (Perpustakaan ITERA) : Menyediakan analisis berbasis data mengenai dinamika kunjungan harian yang dapat dijadikan dasar dalam pengelolaan kapasitas ruang baca dan perencanaan operasional pada hari kerja.
3. Bagi Akademisi : Menjadi contoh studi kasus yang relevan mengenai perbandingan performa berbagai distribusi probabilitas dalam memodelkan data frekuensi kejadian (count data), sehingga dapat digunakan sebagai bahan pembelajaran atau referensi penelitian selanjutnya.

## BAB II TINJAUAN PUSTAKA

### 2.1 Konsep Pembangkitan Bilangan Acak

Pembangkitan bilangan acak (*Random Variate Generation*) adalah proses algoritma untuk menghasilkan urutan angka yang meniru sifat-sifat peluang dari suatu distribusi probabilitas tertentu. Dalam pemodelan stokastik, proses ini merupakan langkah fundamental untuk mereproduksi ketidakpastian sistem nyata ke dalam bentuk data komputasi. Dalam praktiknya, pembangkitan bilangan acak terdiri dari dua tahap utama: pertama, membangkitkan bilangan acak dasar yang terdistribusi seragam ( $\text{Uniform}(0,1)$ ), dan kedua, mentransformasikan bilangan tersebut menjadi distribusi target (misalnya Negatif Binomial) menggunakan algoritma tertentu [2]

Secara teknis, komputer adalah mesin deterministik yang tidak dapat menghasilkan keacakan sejati (*true randomness*). Oleh karena itu, simulasi modern menggunakan algoritma Pseudo-Random Number Generator (PRNG). PRNG adalah algoritma matematika yang menghasilkan urutan angka yang terlihat acak secara statistik, namun sebenarnya dapat diprediksi jika titik awalnya diketahui [1]. Meskipun disebut "semu" atau *pseudo*, PRNG modern seperti *Mersenne Twister* (yang digunakan sebagai standar dalam perangkat lunak R) telah teruji memiliki periode pengulangan yang sangat panjang dan distribusi yang sangat merata, sehingga valid untuk digunakan dalam simulasi ilmiah dan *data science* [4]

### 2.2 Distribusi Negatif Binomial

Distribusi Negatif Binomial didefinisikan sebagai peluang terjadinya sejumlah  $k$  kegagalan sebelum tercapainya sejumlah  $r$  kesuksesan dalam serangkaian percobaan Bernoulli yang saling bebas. Fungsi Massa Peluang (PMF) dari variabel acak  $X$  didefinisikan sebagai berikut [5]

$$X \sim \text{BN}(r, p)$$
$$P[X = k] = \binom{k+r-1}{k} \cdot (1-p)^k \cdot p^r$$

Dimana :

$k$  : variabel acak (jumlah kegagalan/kejadian).  $K = 0, 1, 2, \dots$

$r$  : parameter bentuk (size),  $r > 0$

$p$  : probabilitas sukses  $0 < p < 1$  [5].

### 2.3 Inverse Transform Method

Inverse-Transform method adalah teknik yang digunakan untuk membangkitkan bilangan acak menggunakan CDF dari suatu peubah acak [6]. Berdasarkan teori Probability Integral Transformation (transformasi integral peluang), teori tersebut menyatakan jika  $X$  adalah peubah acak kontinu dengan CDF  $F(x)$ , maka  $U = F(X) \sim \text{Uniform}(0,1)$ .

Tahapan Menggunakan Inverse Transform Method

1. Bangkitkan bilangan acak  $U$  yang berdistribusi Uniform di antara 0 dan 1,  $U(0,1)$ .
2. Misalkan  $F(x)$  adalah CDF dari suatu distribusi acak.
3. Cari invers dari  $F$ , yaitu  $F^{-1}(u)$
4. Maka bilangan acak  $X = F^{-1}(U)$  [7]

Untuk kasus pada bilangan acak diskret maka  $X$  pada  $\dots < x_{i-1} < x_i < x_{i+1} < \dots$  adalah titik yang diskontinu dari  $F_x(x)$ , maka invers transformnya adalah  $F_x^{-1}(u) = X_i$ , Dimana  $F_x(x_{i-1}) < u < F_x(x_i)$  [8] Artinya untuk distribusi diskret, inverse transform memilih nilai  $x_i$  berdasarkan interval CDF tempat nilai uniform  $u$  jatuh.

## 2.4 Acceptance Rejection Method

Metode Acceptance Rejection (AR) adalah teknik pembangkitan bilangan acak yang diperkenalkan oleh John von Neumann. Metode ini digunakan ketika distribusi target  $f(x)$  sulit dibangkitkan secara langsung, baik karena tidak tersedia invers dari fungsi distribusi kumulatif maupun struktur PDF yang kompleks[4]. Inti dari metode ini adalah menggunakan distribusi pembantu  $g(x)$  yang mudah dibangkitkan serta suatu konstanta  $c \geq 1$  sehingga:

$$f(x) \leq cg(x), \forall x$$

Keterangan :

$f(x)$  : fungsi densitas target

$c$  : konstanta pembatas

$g(x)$  : fungsi densitas proposal

Distribusi  $g(x)$  disebut proposal distribution atau envelope, dan konstanta  $c$  disebut *scaling constant*. Dengan memilih  $g(x)$  yang mendekati bentuk  $f(x)$ , metode AR akan menghasilkan sampel secara efisien.

## 2.5 Simulasi Langsung (Direct Simulation / Bernoulli Trials)

Metode Simulasi Langsung atau *Direct Simulation* membangkitkan bilangan acak dengan cara meniru definisi fisik dari distribusi Negatif Binomial itu sendiri. Algoritma ini mensimulasikan serangkaian "lemparan koin" (percobaan Bernoulli) secara berulang [9]. Algoritmanya adalah terus melakukan pengacakan (sukses/gagal) dan menghitung jumlah kegagalan yang terjadi sampai jumlah sukses mencapai target parameter  $r$ . Metode ini sangat intuitif namun mensyaratkan parameter  $r$  harus berupa bilangan bulat (*integer*), yang dapat menyebabkan bias jika parameter asli data adalah bilangan desimal.

Karena fungsi invers CDF  $F^{-1}$  untuk distribusi Negatif Binomial tidak memiliki bentuk tertutup, algoritma pembangkitan dilakukan melalui pencarian sekuensial (*sequential search*). Algoritma ini mencari nilai integer terkecil  $x$  yang memenuhi kondisi:

$$F(x) \geq U$$

Dimana  $U \sim \text{Uniform}(0,1)$  dan  $F(x)$  adalah probabilitas kumulatif  $\sum_{i=0}^x P(X = i)$ . Proses komputasi dimulai dari  $x = 0$  dan nilai  $x$  terus diinkrementasi hingga nilai kumulatif  $F(x)$



melampaui bilangan acak  $U$  [2]. Probabilitas untuk nilai  $x + 1$  dapat dihitung secara efisien dari nilai  $x$  sebelumnya menggunakan persamaan:

$$P(X = x + 1) = \frac{(x + r)(1 - p)}{x + 1} \cdot P(X = x)$$

Keterangan :

$P(X = x + 1)$  : Nilai probabilitas (peluang) untuk kejadian berikutnya ( $x + 1$ ) yang ingin dicari.

$P(X = x)$  : Nilai probabilitas untuk kejadian saat ini ( $x$ ) yang sudah diketahui sebelumnya.

$x$  : Nilai variabel acak saat ini

$r$  : Parameter *size* atau target jumlah sukses (didapat dari hasil estimasi data).

$p$  : Parameter probabilitas sukses (didapat dari hasil estimasi data).

$(1 - p)$ : Sering disebut sebagai  $q$ , yaitu probabilitas gagal.

Dengan inisialisasi awal  $P(X = 0) = p^r$

## **BAB III**

### **METODE PENELITIAN**

#### **3.1 Jenis Data**

Penelitian ini menggunakan data sekunder yang diperoleh dari dokumentasi jumlah pengunjung Perpustakaan GKU 1 ITERA pada hari kerja selama bulan Oktober 2025. Data tersebut merupakan data yang telah dikumpulkan dan dicatat secara rutin oleh pihak pengelola perpustakaan. Pada bentuk awalnya, data tersaji sebagai data runtun waktu (time series), yaitu urutan observasi berdasarkan tanggal. Selanjutnya, data tersebut dikelompokkan per hari sehingga menghasilkan data kuantitatif diskrit, karena setiap nilai menunjukkan jumlah pengunjung dalam bilangan bulat. Pemilihan data sekunder dilakukan karena data telah terstandarisasi, kredibel, dan memiliki cakupan waktu yang memadai untuk dianalisis lebih lanjut menggunakan pendekatan distribusi statistik.

#### **3.2 Teknik Pengumpulan Data**

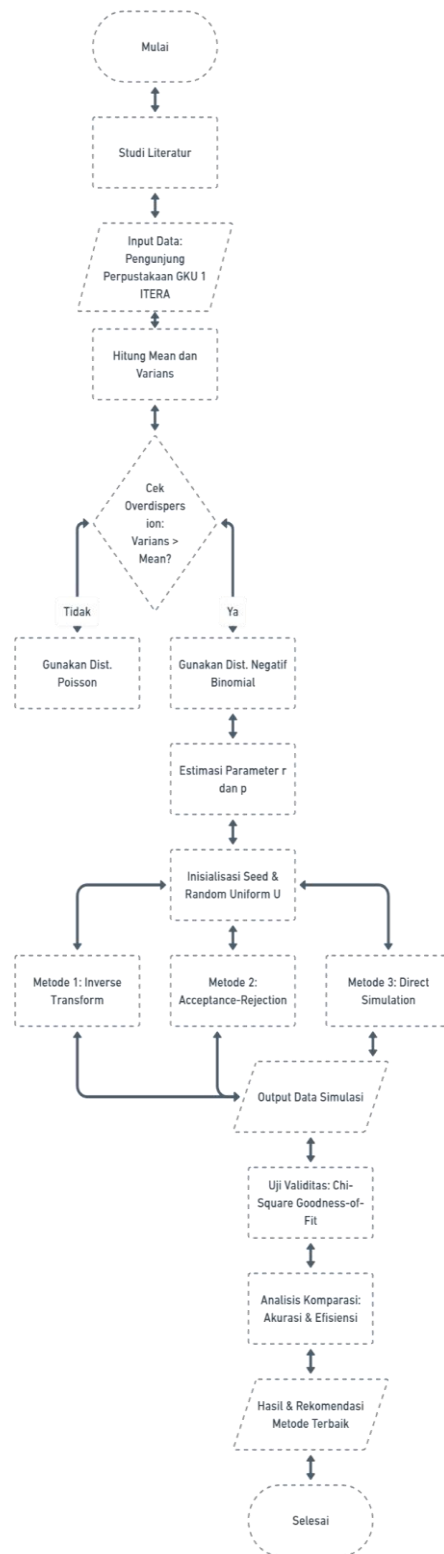
Teknik pengumpulan data dalam penelitian ini menggunakan metode dokumentasi, yaitu pengambilan data yang telah tersedia dan dicatat oleh lembaga terkait. Setelah data diperoleh, dilakukan tahapan pembersihan (data cleaning) untuk menghilangkan nilai yang tidak konsisten, duplikasi, maupun kesalahan pencatatan. Proses ini kemudian dilanjutkan dengan pengkodean serta pengelompokan agar data memiliki struktur yang seragam dan siap digunakan dalam analisis statistik. Tahapan penyiapan data ini penting untuk memastikan bahwa hasil analisis tidak terdistorsi oleh kesalahan teknis dan dapat mencerminkan kondisi nyata secara akurat.

#### **3.3 Variabel yang Diamati**

Variabel utama yang diamati dalam penelitian ini adalah frekuensi pengunjung perpustakaan per hari kerja selama bulan Oktober 2025. Variabel ini bersifat kuantitatif diskrit, karena merepresentasikan jumlah individu yang datang ke perpustakaan dalam satu hari. Variabel ini dipilih sebagai fokus penelitian karena jumlah pengunjung merupakan indikator yang relevan dalam memahami aktivitas layanan perpustakaan, pola interaksi mahasiswa terhadap fasilitas akademik, serta menjadi dasar dalam pemodelan distribusi statistik. Dengan menggunakan variabel tersebut, penelitian dapat mengidentifikasi distribusi mana yang paling mendekati pola kunjungan nyata melalui pembangkitan bilangan acak.

### 3.4 Diagram Alir Penelitian

Berikut diagram alir (flowchart) yang menggambarkan langkah penelitian:



**Gambar 3.1 Diagram Alir Penelitian**

Tahap awal penelitian dimulai dengan melakukan studi literatur yang mendalam mengenai teori pembangkitan bilangan acak (*Random Variate Generation*) dan karakteristik distribusi probabilitas diskrit. Referensi dikumpulkan dari buku teks, jurnal ilmiah, dan penelitian terdahulu yang relevan sebagai acuan teoretis. Langkah selanjutnya adalah proses input data sekunder berupa data frekuensi pengunjung Perpustakaan GKU 1 ITERA. Data yang diambil merupakan rekaman jumlah pengunjung harian selama 22 hari kerja. Data mentah ini kemudian disiapkan dan dibersihkan (*data cleaning*) agar siap untuk diproses secara statistik pada tahap berikutnya.

Selanjutnya adalah perhitungan statistika deskriptif, dimana pada tahap ini, dilakukan perhitungan statistik dasar terhadap data pengunjung yang telah dikumpulkan. Dua parameter utama yang dihitung adalah nilai Rata-rata ( $\bar{x}$ ) Dan Varians ( $s^2$ ). Berdasarkan hasil perhitungan sebelumnya, dilakukan pengambilan keputusan (*decision making*) dengan membandingkan nilai varians terhadap rata-rata. Pertanyaan kuncinya adalah: "apakah nilai varians lebih besar dari rata-rata ( $s^2 > \bar{x}$ )?". Jika jawabannya "tidak", maka data diasumsikan mengikuti distribusi poisson. Namun, jika jawabannya "ya", maka data terindikasi mengalami gejala *overdispersion*.

Karena hasil analisis menunjukkan bahwa data pengunjung memiliki varians yang jauh lebih besar daripada rata-ratanya (kondisi *overdispersion*), maka penelitian menetapkan distribusi negatif binomial sebagai model probabilitas yang paling sesuai. Setelah model ditetapkan, langkah selanjutnya adalah mengestimasi nilai parameter distribusi negatif binomial, yaitu parameter bentuk atau target sukses ( $r$ ) dan parameter probabilitas sukses ( $p$ ) estimasi dilakukan menggunakan pendekatan method of moments, di mana nilai ( $r$ ) dan ( $p$ ) dihitung berdasarkan nilai rata-rata dan varians sampel data asli agar karakteristik data simulasi nantinya serupa dengan data riil.

Tahap inisialisasi dan pembangkitan bilangan acak dasar, dimulai dengan menginisialisasi nilai seed pada program R untuk memastikan hasil simulasi dapat direproduksi (reproducible) dan konsisten. Setelah itu, komputer diperintahkan untuk membangkitkan bilangan acak dasar yang berdistribusi uniform standar, atau dinotasikan sebagai  $U \sim \text{Uniform}(0,1)$  yang akan menjadi bahan baku utama transformasi. Lalu bilangan acak uniform yang telah dibangkitkan kemudian diproses menggunakan tiga algoritma berbeda secara paralel. Pertama, metode *inverse transform* yang menggunakan prinsip invers fungsi kumulatif. Kedua, metode *acceptance-rejection* yang menggunakan fungsi pembungkus (envelope) untuk menyeleksi data. Ketiga, metode *direct simulation* yang meniru proses percobaan bernoulli hingga mencapai target sukses  $r$ .

Output data simulasi dari ketiga metode tersebut kemudian dievaluasi validitasnya menggunakan uji *chi-square goodness-of-fit*. Tahap analisis dilakukan dengan membandingkan kinerja ketiga metode tersebut. Indikator perbandingan mencakup tingkat akurasi (dilihat dari *mean squared error* atau selisih parameter terhadap data asli) dan tingkat efisiensi komputasi (kecepatan atau banyaknya data yang terbuang). Analisis ini bertujuan untuk menentukan metode

mana yang paling unggul dalam memodelkan kasus pengunjung perpustakaan. Berdasarkan hasil analisis komparasi, penelitian menyimpulkan metode terbaik yang direkomendasikan

## BAB IV HASIL DAN PEMBAHASAN

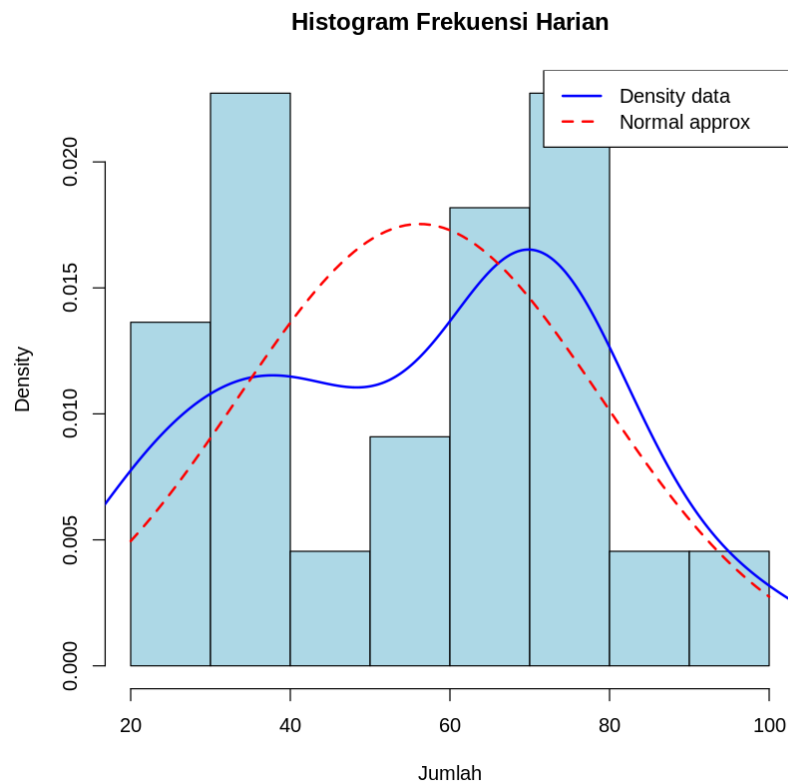
### 4.1 Deskripsi Data Penelitian

**Tabel 4.1 Sampel Data Frekuensi Pengunjung Perpustakaan GKU 1 Bulan Oktober 2025**

Hari kerja	Frekuensi
Hari ke-1	81
Hari ke-2	71
...	...
Hari ke-22	23

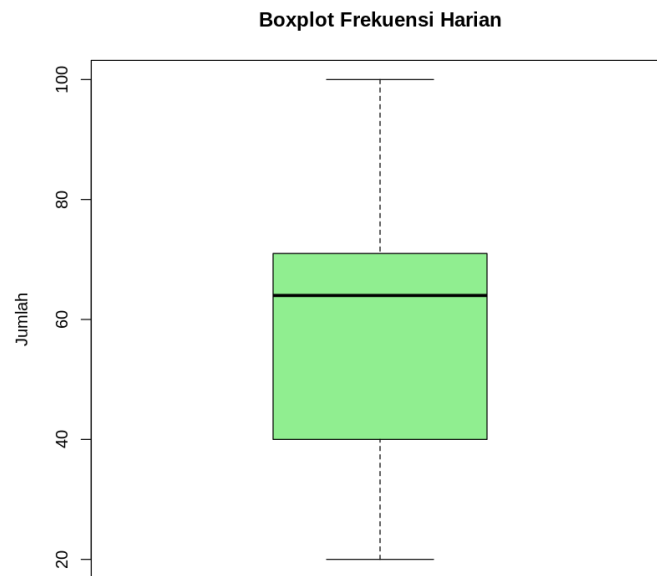
Data penelitian berupa jumlah pengunjung Perpustakaan GKU 1 ITERA selama hari kerja pada bulan Oktober 2025. Data memiliki bentuk deret waktu harian dan kemudian dianalisis secara statistik untuk melihat pola, sebaran, serta karakteristik distribusinya. Informasi dasar menjadi fondasi untuk menentukan model probabilistik yang digunakan dalam proses pembangkitan data acak.

#### 4.1.1 Visualisasi Data Awal



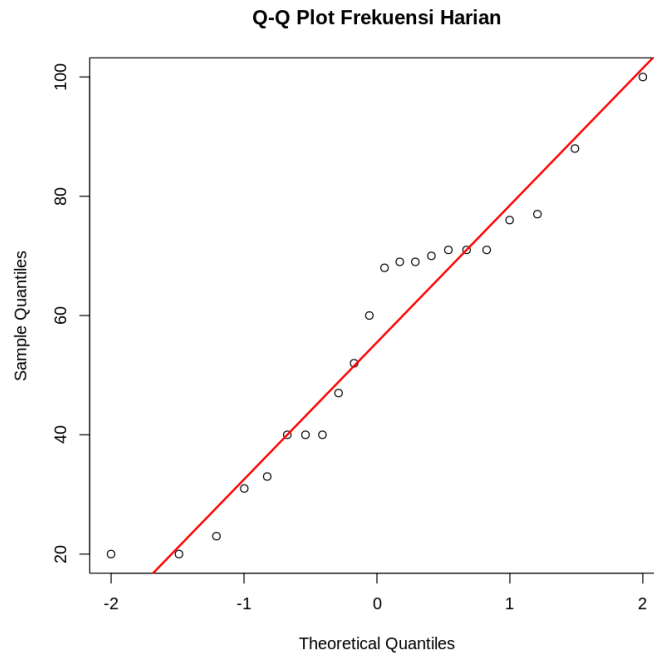
**Gambar 4.1 Hasil Histogram Frekuensi Harian**

Hasil visualisasi data awal, histogram frekuensi harian menggambarkan distribusi jumlah pengunjung perpustakaan setiap hari kerja. Secara visual, histogram menunjukkan sebaran data yang tidak simetris dan cenderung memiliki dua pusat konsentrasi (bimodal), yaitu pada rentang 30–40 dan 65–80 pengunjung. Kurva kepadatan (density) yang ditampilkan melalui garis biru berbeda cukup jauh dari kurva pendekatan Normal (garis putus-putus merah). Sehingga, dari perbedaan menunjukkan bahwa pola kunjungan harian tidak mengikuti distribusi Normal dan memiliki karakteristik penyebaran yang kompleks.



**Gambar 4.2 Hasil Boxplot Frekuensi Harian**

Hasil boxplot frekuensi harian menunjukkan distribusi jumlah pengunjung perpustakaan dalam bentuk ukuran pemusatan dan penyebaran data. Nilai minimum tercatat sebesar 20 pengunjung, sedangkan nilai maksimum mencapai 100 pengunjung. Median berada pada angka 64, yang menunjukkan bahwa setengah dari total observasi berada di atas nilai tersebut. Pola sebaran dalam IQR terlihat relatif merata, mengindikasikan tidak adanya penumpukan frekuensi yang dominan pada rentang tengah. Struktur boxplot menunjukkan heterogenitas yang cukup tinggi dalam data kunjungan harian.



**Gambar 4.3 Hasil Q-Q Plot Frekuensi Harian**

Q-Q Plot frekuensi harian digunakan sebagai pemeriksaan grafis terhadap asumsi normalitas. Titik-titik kuantil pada plot terlihat tidak mengikuti garis secara konsisten, terutama pada bagian tengah distribusi dan kedua ekornya. Pola penyimpangan menunjukkan bahwa sebaran data tidak sepenuhnya menyerupai distribusi Normal. Hasil juga sejalan dengan histogram yang menunjukkan bentuk distribusi yang tidak simetris, sehingga mengindikasikan bahwa pola frekuensi kunjungan memiliki karakteristik penyebaran yang tidak mengikuti pola Normal secara sempurna.

#### 4.1.2 Statistik Deskriptif

**Tabel 4.2 Ringkasan Statistik**

Mean	Median	Modus	Variansi	Standar Deviasi	Range	Kuartil 1	Kuartil 3
56.18182	64	71	517.5844	22.7504	20 - 100	40	71

Statistik deskriptif digunakan untuk memberikan gambaran kuantitatif mengenai struktur data kunjungan harian. Rata-rata jumlah pengunjung adalah 56,18, sementara median sebesar 64 dan modus sebesar 71. Kombinasi ketiga ukuran pemusatan tersebut menunjukkan bahwa distribusi memiliki kecenderungan nilai tengah yang lebih tinggi dibandingkan nilai rata-rata. Ukuran penyebaran memperlihatkan variansi sebesar 517,58 dan standar deviasi sebesar 22,75, yang menandakan adanya tingkat variasi yang cukup tinggi antar hari. Rentang observasi berada pada



interval 20 hingga 100 pengunjung dengan nilai kuartil 1 nya 40 dan kuartil 3 nya 71. Secara keseluruhan, menegaskan bahwa data memiliki variabilitas besar.

## 4.2 Analisis Distribusi Data

Analisis distribusi data dilakukan untuk menilai kecocokan beberapa kandidat distribusi terhadap pola frekuensi kunjungan harian.

### 4.2.1 Uji Goodness-of-Fit (GoF)

Pengujian kecocokan distribusi dilakukan untuk menentukan model probabilistik yang paling sesuai dengan pola data kunjungan harian.

**Tabel 4. 3 Uji Chi-Square Distribusi Poisson**

Statistic	12646358
p-value	0

Uji Chi-Square diterapkan pada distribusi Poisson. Hasilnya menunjukkan nilai statistik sebesar 12.646.358 dengan  $p\text{-value} = 0$ , sehingga distribusi Poisson ditolak sebagai model yang sesuai. Kondisi ini konsisten dengan karakteristik data yang memiliki variansi jauh lebih besar dibandingkan rata-rata, sehingga tidak memenuhi asumsi dasar distribusi Poisson.

**Tabel 4. 4 Uji Chi-Square Distribusi Binomial**

Statistic	4.988826e+23
p-value	0

Uji Chi-Square untuk distribusi Binomial menghasilkan nilai statistik sebesar  $4,99 \times 10^{23}$  dengan  $p\text{-value} = 0$ . Hasil menunjukkan bahwa distribusi Binomial tidak mampu menggambarkan pola frekuensi kunjungan, terutama karena parameter peluang keberhasilan pada distribusi Binomial tidak dapat menyesuaikan besarnya variasi yang ditemukan dalam data.

**Tabel 4. 5 Likelihood Ratio Test Poisson vs Negative Binomial**

LRT Statistic	134.9174
p-value	3.44199e-31

Selanjutnya dilakukan Likelihood Ratio Test (LRT) untuk membandingkan kecocokan antara distribusi Poisson dan Negatif Binomial. Nilai statistik LRT sebesar 134,917 dengan  $p\text{-value} 3,44 \times 10^{-31}$  menunjukkan bahwa model Negatif Binomial memberikan kecocokan yang secara signifikan lebih baik dibandingkan Poisson. Hal ini mendukung hipotesis bahwa data bersifat overdispersed dan dimodelkan menggunakan distribusi dengan dua parameter.

**Tabel 4. 6 Perbandingan AIC & BIC**

	df	AIC BIC	df	BIC BIC
fit_pois	1	336.8169	1	337.9080
fit_nb	2	203.8996	2	206.0817

Pemilihan model juga diperkuat melalui kriteria informasi AIC dan BIC. Distribusi Negatif Binomial memiliki nilai AIC sebesar 203,90 dan BIC sebesar 206,08, jauh lebih kecil dibandingkan Poisson (AIC = 336,82 dan BIC = 337,91). Karena model dengan nilai AIC dan BIC terendah dianggap paling efisien dalam menjelaskan data dengan kompleksitas minimal, maka distribusi Negatif Binomial ditetapkan sebagai model terbaik untuk data kunjungan harian.

#### 4.2.2 Pemilihan Model Terbaik

Pemilihan model terbaik dilakukan dengan mempertimbangkan hasil uji kecocokan distribusi. Distribusi Poisson dan Binomial ditolak karena tidak sesuai dengan pola data, sedangkan distribusi Negatif Binomial menunjukkan nilai AIC dan BIC paling rendah dan hasil LRT yang signifikan. Dengan demikian, distribusi Negatif Binomial ditetapkan sebagai model yang paling tepat untuk menggambarkan data kunjungan harian.

**Tabel 4. 7 Uji Normalitas Menggunakan Uji Shapiro Wilk**

W	0.94018
p-value	0.1996

Uji Shapiro–Wilk dilakukan untuk melihat potensi kesesuaian data dengan distribusi Normal. Hasilnya menunjukkan  $W = 0.94018$  dengan p-value 0.1996, yang berarti tidak terdapat cukup bukti untuk menolak normalitas secara formal. Namun karena data bersifat diskrit dan mengalami overdispersion, distribusi Normal tetap tidak digunakan. Sehingga, distribusi Negatif Binomial ditetapkan sebagai model yang paling tepat untuk menggambarkan data kunjungan harian.

### 4.3 Pembangkitan Data (Simulasi)

Pembangkitan data dilakukan untuk mensimulasikan pola kunjungan harian perpustakaan menggunakan tiga metode, yaitu *Inverse Transform Method*, *Acceptance-Rejection Method*, dan *Direct Simulation Method*, yang semuanya didasarkan pada model Distribusi Negatif Binomial.

#### 4.3.1 Metode Inverse Transform

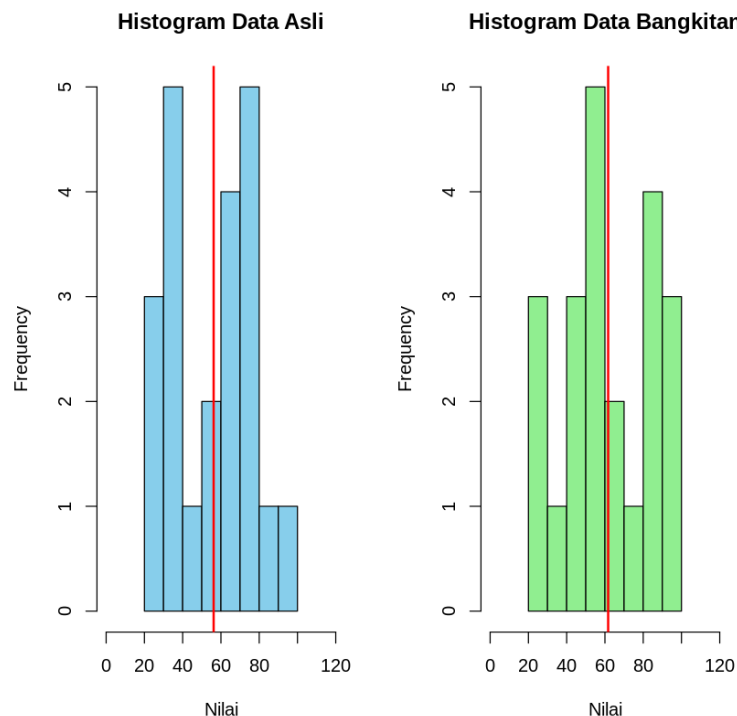
**Tabel 4. 8 Parameter hasil Estimasi dengan Metode Invers Transform**

Mean Data Asli	Varian Data asli	Estimasi r (size)	Estimasi p (prob)
56.18182	517.5844	6.840873	0.1085462

Metode Inverse Transform digunakan untuk membangkitkan bilangan acak berdasarkan parameter distribusi Negatif Binomial yang sebelumnya telah ditetapkan sebagai model paling sesuai untuk data frekuensi kunjungan perpustakaan. Parameter estimasi yang digunakan dalam proses simulasi adalah mean sebesar 56.18182, variansi 517.5844, nilai  $r$  (size) sebesar 6.840873, serta probabilitas sukses  $p$  sebesar 0.1085462. Parameter ini membentuk dasar fungsi distribusi kumulatif (CDF) yang di-inverskan untuk menghasilkan sampel acak.

**Tabel 4.9 Sampel Data Bangkitan (Inverse Transform)**

Hari kerja	Frekuensi
Hari ke-1	42
Hari ke-2	73
...	...
Hari ke-22	65



**Gambar 4.4 Histogram Data Asli vs Data Bangkitan**

Proses pembangkitan data menghasilkan 22 nilai simulasi, sesuai jumlah observasi pada data asli. Secara visual, perbandingan antara histogram data asli dan histogram data bangkitan menunjukkan pola sebaran yang serupa. Distribusi frekuensi hasil simulasi mengikuti bentuk umum distribusi observasi, termasuk rentang nilai dan kemunculan nilai ekstrim. Garis rata-rata pada histogram data bangkitan juga berada dekat dengan rata-rata data asli, sehingga menunjukkan bahwa metode ini mampu mereplikasi karakteristik pemusatan data. Parameter hasil estimasi tersebut menghasilkan set data bangkitan seperti 42, 73, 48, 84, 95, hingga 100, yang

memperlihatkan fluktuasi serupa dengan data asli seperti 71, 77, 100, 88, 69, dan 60. Rentang nilai yang dihasilkan tetap berada dalam kisaran yang mendekati data observasi, termasuk kemunculan nilai-nilai rendah dan tinggi. Kesamaan pola ini memperkuat bahwa metode Inverse Transform dapat meniru struktur distribusi asli secara efektif pada ukuran sampel kecil. Evaluasi lebih lanjut dengan simulasi berukuran besar ( $n = 10.000$ ) menunjukkan bahwa metode ini menghasilkan nilai rata-rata dan variansi yang paling mendekati parameter distribusi Negatif Binomial dibandingkan metode lain.

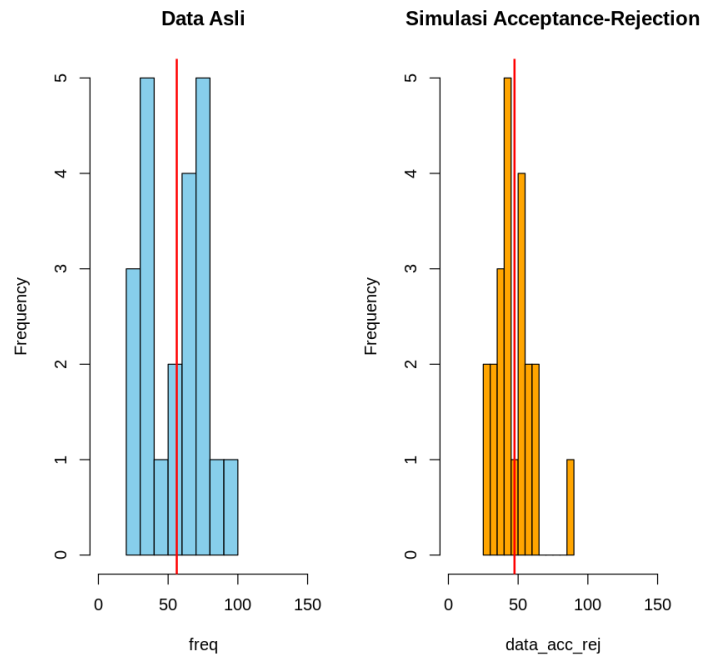
#### 4.3.2 Metode Acceptance-Rejection

Metode Acceptance-Rejection digunakan untuk membangkitkan bilangan acak dengan memanfaatkan distribusi proposal yang lebih sederhana dibandingkan distribusi target Negatif Binomial.

**Tabel 4.10 Sampel Data Bangkitan (Acceptance Rejection)**

Hari kerja	Frekuensi
Hari ke-1	60
Hari ke-2	64
...	...
Hari ke-22	42

Sampel yang dihasilkan dari fungsi proposal dievaluasi menggunakan rasio probabilitas, dan hanya nilai yang memenuhi kriteria penerimaan yang disertakan sebagai data simulasi. Pada pembangkitan sampel kecil sebanyak 22 data, metode ini memerlukan 70 percobaan untuk memperoleh 22 sampel yang diterima, sehingga menghasilkan tingkat efisiensi sebesar 31.43 persen. Efisiensi yang rendah tersebut mencerminkan banyaknya sampel yang ditolak dalam proses simulasi, yang merupakan karakteristik umum dari pendekatan ini ketika distribusi target memiliki bentuk yang kompleks.



**Gambar 1.5 Hasil Data Asli vs Simulasi Acceptance-Rejection**

Hasil menunjukkan bahwa histogram data bangkitan memiliki pola sebaran yang mendekati histogram data asli, dengan nilai rata-rata yang relatif berdekatan. Evaluasi menggunakan simulasi berukuran besar ( $n = 10.000$ ) menunjukkan bahwa metode mampu menghasilkan mean dan variansi yang cukup dekat dengan nilai teoritis, meskipun masih berada sedikit di bawah akurasi metode Inverse Transform.

#### 4.3.3 Metode Direct Transformation

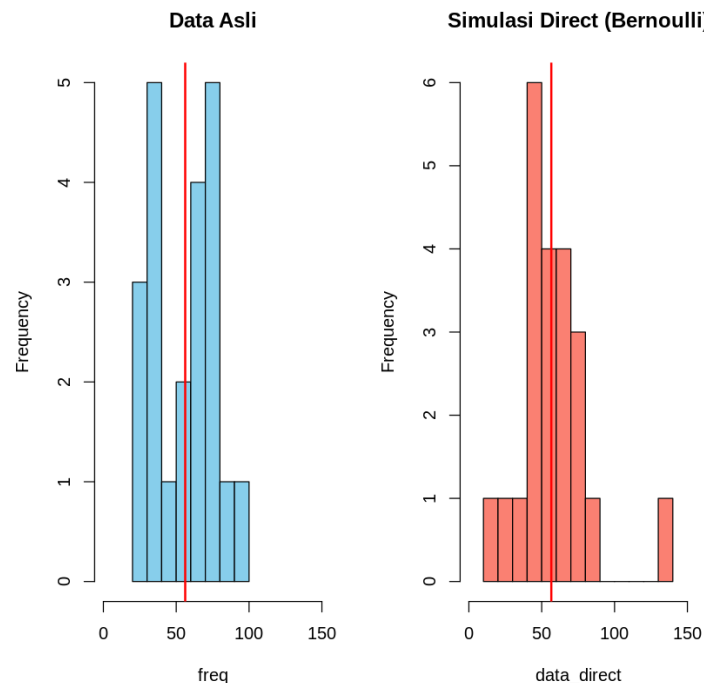
Metode Direct Transformation membangkitkan bilangan acak berdasarkan pendekatan proses Bernoulli, di mana distribusi Negatif Binomial direpresentasikan sebagai jumlah percobaan hingga tercapai  $r$  keberhasilan.

**Tabel 4.11 Sampel Data Bangkitan (Direct Transformation)**

Hari kerja	Frekuensi
Hari ke-1	67
Hari ke-2	134
...	...
Hari ke-22	41

Metode Direct Simulation menggunakan parameter peluang sukses  $p$  sebesar 0.1085462 dan nilai  $r$  hasil estimasi 6.840873 yang kemudian dibulatkan menjadi 7 agar memenuhi syarat bilangan bulat pada proses simulasi berbasis percobaan Bernoulli. Ketergantungan metode ini pada pembulatan parameter  $r$  mengakibatkan penurunan presisi, karena nilai  $r$  yang sebenarnya bersifat kontinu dan tidak tepat direpresentasikan sebagai bilangan bulat. Akibatnya, struktur distribusi yang terbentuk tidak sepenuhnya merefleksikan karakteristik distribusi Negatif Binomial pada data

asli. Hal tersebut tercermin pada data bangkitan Direct Simulation berikut: 67, 134, 63, 10, 80, 58, 51, 41, 46, 46, 45, 34, 62, 46, 52, 62, 76, 82, 53, 23, 73, 41. Distribusi yang dihasilkan tampak lebih menyebar dan kurang stabil dibandingkan pola observasi.



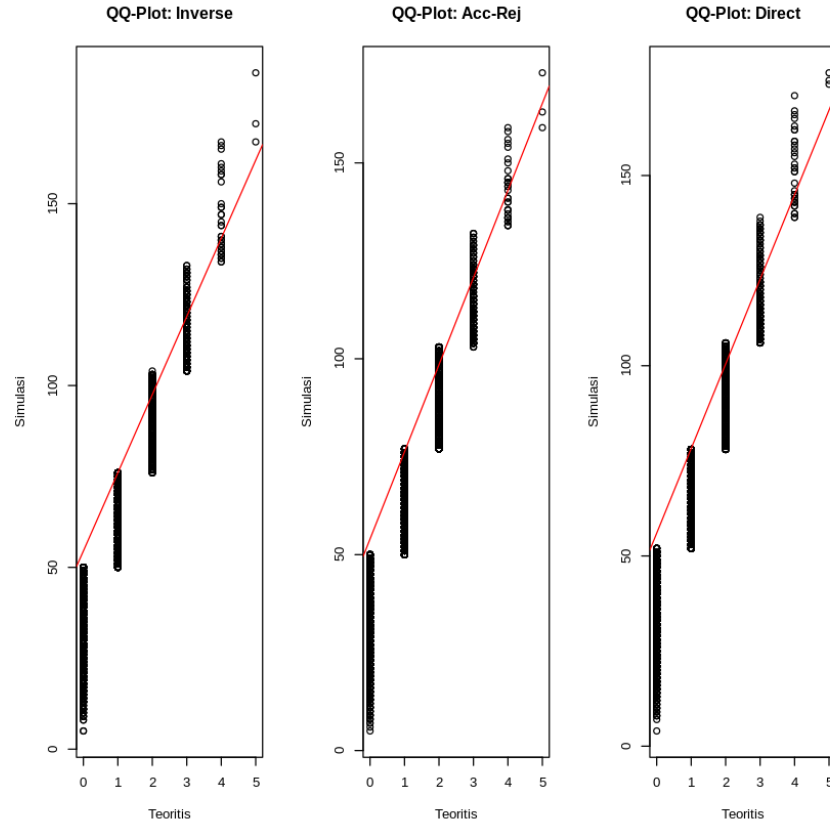
**Gambar 1.6 Hasil Data Asli vs Simulasi Direct**

Perbandingan visual antara histogram data asli dan histogram hasil simulasi Direct (Bernoulli) menunjukkan bahwa meskipun distribusi bangkitan memiliki bentuk yang lebih terpusat, sebarannya tampak lebih menyimpang dari pola data observasi dibandingkan metode lainnya. Garis rata-rata pada histogram simulasi bergeser dari nilai rata-rata data asli, menandakan bahwa karakteristik pusat distribusi tidak direplikasi secara akurat. Evaluasi lebih lanjut menggunakan sampel besar sebanyak 10.000 observasi mengkonfirmasi hasil tersebut, di mana metode Direct Simulation menghasilkan deviasi mean dan variansi yang paling besar dibandingkan dua metode pembangkitan lainnya.

#### 4.4 Perbandingan Akurasi Metode Pembangkitan Data

**Tabel 4. 12 Perbandingan Akurasi Metode Pembangkitan Data**

	Mean_Sim	Var_Sim	Err_Mean	Err_Var
Inverse	55.9166	506.7425	0.2652	10.8419
AcceptReject	55.8617	508.4572	0.3201	9.1272
Direct	57.3703	534.7221	1.1885	17.1376



**Gambar 1.6 Hasil Perbandingan Akurasi Metode**

Hasil membandingkan akurasi tiga metode pembangkitan data, yaitu Inverse Transform, Acceptance-Rejection, dan Direct Simulation, dengan menggunakan sampel besar sebanyak 10.000 observasi. Akurasi dievaluasi berdasarkan kedekatan nilai statistik simulasi terhadap parameter target, yaitu mean sebesar 56.18 dan varians sebesar 517.58. Hasil menunjukkan bahwa Metode Inverse Transform memiliki performa paling baik, dengan mean simulasi 55.9166 dan varians 506.7425 serta nilai kesalahan rata-rata (0.2652) yang paling kecil di antara seluruh metode. Metode Acceptance-Rejection menghasilkan akurasi yang hampir setara, dengan mean simulasi 55.8617 dan varians 508.4572. Namun, metode ini memiliki efisiensi rendah, yaitu hanya 26.29%, di mana diperlukan 38.042 percobaan untuk memperoleh 10.000 data diterima. Sementara itu, Metode Direct Simulation menunjukkan penyimpangan terbesar, baik pada nilai rata-rata (1.1885) maupun varians (17.1376), terutama akibat pembulatan parameter  $r$  yang menurunkan presisi distribusi. Secara visual, Q-Q Plot perbandingan juga mengonfirmasi bahwa simulasi Inverse Transform dan Acceptance-Rejection berada lebih dekat terhadap garis teoretis dibandingkan Direct Simulation. Dengan mempertimbangkan kedekatan statistik, stabilitas hasil, serta efisiensi proses, Inverse Transform ditetapkan sebagai metode pembangkitan bilangan acak yang paling akurat.

## 4.5 Diskusi

Berdasarkan hasil analisis data, ditemukan bahwa pola kunjungan Perpustakaan GKU 1 ITERA menunjukkan karakteristik *overdispersion* yang ekstrem, di mana nilai varians data jauh melampaui nilai rata-ratanya. Kondisi ini menyebabkan Distribusi Poisson menjadi tidak valid untuk digunakan karena melanggar asumsi ekuidispersi. Oleh karena itu, pemilihan Distribusi Negatif Binomial sebagai model utama dalam penelitian ini terbukti tepat dan memberikan kesesuaian model yang signifikan. Temuan ini sejalan dengan penelitian Winata (2023), yang menyatakan bahwa penggunaan model Negatif Binomial merupakan solusi efektif untuk menangani data cacah (*count data*) yang memiliki variabilitas tinggi, karena adanya parameter dispersi tambahan yang mampu mengakomodasi fluktuasi data secara lebih akurat dibandingkan model Poisson standar [5].

Selanjutnya, dalam pengujian simulasi pembangkitan bilangan acak, metode *Inverse Transform* menunjukkan kinerja yang paling superior dibandingkan dua metode lainnya. Metode ini menghasilkan nilai *Mean Squared Error* (MSE) terendah dan secara konsisten lolos uji validitas statistik. Hal ini mengonfirmasi landasan teoretis yang dijelaskan oleh Kroese et al. (2024) dalam literatur sains data modern, bahwa metode invers bekerja dengan memetakan probabilitas kumulatif secara eksak tanpa melibatkan proses aproksimasi atau penolakan data. Stabilitas matematis ini memungkinkan metode *Inverse Transform* mempertahankan karakteristik statistik asli dari data perpustakaan seperti rata-rata dan varians dengan presisi tinggi, menjadikannya pendekatan yang paling *robust* untuk kebutuhan simulasi operasional [2].

Sebaliknya, metode *Acceptance-Rejection* dan *Direct Simulation* menunjukkan keterbatasan performa yang nyata. Meskipun metode *Acceptance-Rejection* valid secara statistik, efisiensinya tergolong rendah karena tingginya proporsi data yang ditolak (*rejected*) akibat penggunaan fungsi proposal Uniform yang kurang optimal dalam menutupi ekor distribusi target. Sementara itu, metode *Direct Simulation* mengalami kegagalan struktural dalam mereproduksi varians data karena adanya kendala pada parameter target sukses ( $r$ ) yang harus berupa bilangan bulat. Pemaksaan pembulatan parameter  $r$  dari nilai desimal (hasil estimasi data riil) ke bilangan bulat menyebabkan bias sistematis, sehingga metode ini tidak direkomendasikan untuk memodelkan data empiris yang parameternya jarang bernilai bulat sempurna.



## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Berdasarkan seluruh proses analisis dan simulasi pembangkitan -bilangan acak pada studi kasus frekuensi pengunjung Perpustakaan GKU 1 ITERA, diperoleh beberapa kesimpulan penting. Pertama, hasil statistik deskriptif menunjukkan bahwa data pengunjung memiliki rata-rata sekitar 56 orang per hari dengan variasi yang cukup tinggi. Karakter tersebut tercermin dari nilai varians yang jauh lebih besar daripada mean, sehingga data memiliki sifat *overdispersion*. Kondisi ini mengindikasikan bahwa distribusi Poisson dan Binomial tidak sesuai untuk mewakili pola kunjungan. Hal ini diperkuat melalui uji Chi-Square, di mana keduanya menghasilkan p-value  $\approx 0$ , sehingga ditolak sebagai distribusi yang cocok.

Sebaliknya, distribusi Negatif Binomial terbukti jauh lebih representatif. Hal ini ditunjukkan dari hasil *Likelihood Ratio Test* yang memberikan p-value sangat kecil, menegaskan bahwa model *Negative Binomial* secara signifikan lebih baik dibandingkan Poisson. Selain itu, nilai AIC dan BIC untuk *Negative Binomial* jauh lebih kecil dibandingkan model lain, sehingga secara empiris merupakan model terbaik untuk menggambarkan frekuensi kunjungan harian. Pada uji KS diskrit, distribusi Negatif Binomial juga memberikan kecocokan yang lebih baik. Dengan demikian, distribusi Negatif Binomial menjadi dasar pembangkitan bilangan acak untuk mensimulasikan pola kedatangan pengunjung.

Proses pembangkitan bilangan acak dilakukan menggunakan tiga teknik, yaitu *Inverse Transform*, *Acceptance-Rejection*, dan *Direct Simulation*. Ketiga metode mampu menghasilkan pola data yang menyerupai data asli, tetapi performanya berbeda. Hasil evaluasi menggunakan sampel besar ( $n = 10.000$ ) menunjukkan bahwa metode *Inverse Transform* menghasilkan nilai mean dan varians yang paling mendekati target teoretis. *Acceptance-Rejection* memiliki akurasi yang hampir setara tetapi dengan efisiensi yang lebih rendah karena banyaknya penolakan sampel. Sementara itu, metode *Direct Simulation* menghasilkan penyimpangan yang paling besar karena bergantung pada jumlah keberhasilan  $r$  yang dibulatkan, sehingga kurang presisi ketika parameter  $r$  bukan bilangan bulat. Secara keseluruhan, metode *Inverse Transform* merupakan teknik pembangkitan bilangan acak yang paling stabil dan akurat untuk kasus ini.

Dengan seluruh temuan ini dapat disimpulkan bahwa:

1. Distribusi Negatif Binomial adalah model probabilistik terbaik untuk merepresentasikan frekuensi pengunjung perpustakaan.
2. Metode Inverse Transform adalah teknik pembangkitan bilangan acak yang paling akurat untuk meniru karakteristik distribusi frekuensi kunjungan.
3. Hasil simulasi mampu memodelkan pola kedatangan secara realistis, sehingga dapat digunakan sebagai dasar perancangan analitik atau pengambilan keputusan operasional.

## 5.2 Saran

Berdasarkan hasil penelitian dan simulasi yang telah dilakukan, beberapa saran dapat diberikan untuk pengembangan analisis berikutnya. Pertama, penelitian selanjutnya dapat mempertimbangkan pengumpulan data dengan periode waktu yang lebih panjang agar model distribusi dapat lebih stabil dan mewakili pola kunjungan tahunan, tidak hanya berdasarkan sampel harian yang terbatas. Selain itu, faktor-faktor eksternal seperti jadwal kuliah, musim ujian, atau kegiatan kampus lain dapat dimasukkan sebagai variabel tambahan karena dapat mempengaruhi fluktuasi kunjungan perpustakaan.

Dari sisi metode pembangkitan bilangan acak, penggunaan teknik *Inverse Transform* dapat dijadikan pendekatan utama dalam simulasi karena terbukti paling mendekati karakteristik data asli. Namun demikian, teknik *Acceptance-Rejection* tetap dapat digunakan sebagai alternatif apabila diperlukan fleksibilitas dalam memilih fungsi proposal untuk distribusi yang lebih kompleks. Metode *Direct Simulation* dapat ditingkatkan akurasinya dengan memperhalus estimasi parameter atau menggunakan pendekatan lain yang tidak mensyaratkan pembulatan nilai  $r$ .

Akhirnya, hasil simulasi distribusi pengunjung ini dapat dimanfaatkan pihak perpustakaan untuk perencanaan kapasitas ruangan, alokasi staf, dan pengaturan jadwal layanan. Pengembangan ke depan juga dapat memasukkan analisis prediktif berbasis model probabilistik sehingga perpustakaan dapat mengantisipasi lonjakan pengunjung pada waktu-waktu tertentu secara lebih tepat.

## DAFTAR PUSTAKA

- [1] P. Nurhaliza -, “Implementasi Pembangkit Bilangan Acak Semu dengan Henon-Sine Hyperchaotic Map.”
- [2] D. P. Kroese, Z. I. Botev, T. Taimre, and R. Vaisman, “Data Science and Machine Learning Mathematical and Statistical Methods,” 2024.
- [3] Annette J. Dobson and Adrian G. Barnett, “An Introduction to Generalized Linear Models Fourth Edition,” Brisbane, Australia, 2018.
- [4] L. A. Siswanto, “Analisis Keacakan Generator Angka Pseudorandom Mersenne Twister dengan Metode Diehard Test,” 2015.
- [5] H. M. Winata, “MENGATASI OVERDISPERSI DENGAN REGRESI BINOMIAL NEGATIF PADA ANGKA KEMATIAN IBU DI KOTA BANDUNG,” *Jurnal Gaussian*, vol. 11, no. 4, pp. 616–622, Feb. 2023, doi: 10.14710/j.gauss.11.4.616-622.
- [6] M. Jdid and A. Salama, “Using the Inverse Transformation Method to Generate Random Variables that follow the Neutrosophic Uniform Probability Distribution,” *Journal of Neutrosophic and Fuzzy Systems*, vol. 6, no. 2, pp. 15–22, 2023, doi: 10.54216/JNFS.060202.
- [7] Arif Rahman, “PENDEKATAN INVERSE-TRANSFORM RANDOM VARIATE GENERATOR BERBASIS DISTRIBUSI GEOMETRI PADA PENGACAKAN RANDOM SAMPLING,” *Seminar Nasional V Manajemen & Rekayasa Kualitas 2013*, pp. 1–6, 2013.
- [8] Karl Sigman, “Inverse Transform Method,” pp. 1–3, 2010.
- [9] S. M. Ross, “Introduction to Probability Models, Tenth Edition,” Los Angeles, California, 2010.