

A Comparative Analysis of Bayesian and Frequentist Methods for Independent K-Sample Testing

Pardomoan Robinson Sihombing^{1*}

¹Badan Pusat Statistik Indonesia, Jakarta Pusat, Indonesia

*Corresponding author: robinson@bps.go.id

Abstract

Regional socioeconomic inequality remains a critical development challenge in Indonesia, with significant variations observed in Sumatra, Java-Bali, and other regions. This study conducted a comprehensive methodological comparison between Bayesian and frequentist approaches to test the average differences among three independent regional groups using five economic and social indicators from 2024-2025: Human Development Index (HDI), Gini ratio, poverty rate, unemployment rate, and economic growth. The authors evaluated Bayesian ANOVA against parametric ANOVA *F* and four nonparametric alternatives, including the Kruskal-Wallis test, permutation test, and median test, by assessing its performance under different assumption violations. Statistical analysis showed that HDI and Gini ratios showed significant regional differences across all methods, while poverty rates showed method-dependent outcomes, and unemployment and economic growth showed no significant regional differences. These findings demonstrate the importance of method selection based on assumption testing and highlight the advantages of the Bayesian approach for regional analysis with small samples. Policymakers should prioritize targeted interventions for disadvantaged regions while addressing income inequality in more developed regions.

Keywords: Bayesian, ANOVA, Inequality, Kruskal-Wallis, and K-Sample

Article History:

Received : 07 July 2026

Revised : 11 August 2026

Accepted : 31 August 2026

© 2026 Sihombing. Published by Program
Studi Statistika Fakultas MIPA Universitas
Lambung Mangkurat.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



How to Cite:

Sihombing, P. R. A Comparative Analysis of Bayesian and Frequentist Methods for Independent K-Sample Testing, RAGAM: Journal of Statistics and Its Application, vol. 5, no. 2, pp. 160-174, 2026, [doi: 10.20527/ragam.v5i2.19876](https://doi.org/10.20527/ragam.v5i2.19876)

1. PENDAHULUAN

Geografi kepulauan Indonesia yang luas mencakup 38 provinsi telah menyebabkan ketimpangan regional yang persisten dalam pembangunan sosial-ekonomi, yang merupakan salah satu tantangan kebijakan paling mendesak bagi negara ini [1]. Indeks Pembangunan Manusia (IPM), ketimpangan pendapatan yang diukur dengan rasio Gini, tingkat kemiskinan, tingkat pengangguran, dan pertumbuhan ekonomi bervariasi secara signifikan antar provinsi, mencerminkan akses yang berbeda terhadap infrastruktur, pendidikan, layanan kesehatan, dan peluang ekonomi. Pada tahun 2025, Indeks Pembangunan Manusia nasional Indonesia mencapai 75.90, namun skor provinsi bervariasi secara dramatis dari 54.91 di Papua Pegunungan hingga 85,05 di DKI Jakarta, menggambarkan kesenjangan pembangunan setara dengan hampir tiga dekade [2]. Demikian pula, tingkat kemiskinan bervariasi dari 3.72 persen di Bali hingga 30.03 persen di Papua Pegunungan. Sementara rasio Gini berkisar dari 0.261 di Kalimantan Utara hingga 0.441 di DKI Jakarta, menunjukkan konsentrasi pendapatan yang parah di kawasan metropolitan [3]. Ketimpangan ini sering dianalisis dengan menggabungkan

provinsi ke dalam kategori regional yang lebih luas yaitu Sumatera, Jawa-Bali, dan wilayah lain yang mencakup Kalimantan, Sulawesi, Nusa Tenggara, Maluku, dan Papua untuk mengidentifikasi pola perkembangan sistematis dan mendukung kebijakan regional berbasis bukti.

Inferensi statistik untuk membandingkan rata-rata antar kelompok independen secara tradisional mengandalkan analisis variansi satu arah (ANOVA), yang mengasumsikan data berdistribusi normal dengan varians homogen antar kelompok [4]. Namun, indikator sosioekonomi seringkali melanggar asumsi parametrik ini akibat kemiringan data, outlier, ukuran sampel kecil dalam kategori regional, atau heteroskedastisitas yang timbul dari karakteristik provinsi yang beragam [5]. Ketika asumsi parametrik gagal, peneliti harus menggunakan alternatif yang tangguh seperti uji Kruskal-Wallis, yang menggunakan inferensi berbasis peringkat tanpa persyaratan distribusi [6], atau uji permutasi [7] yang membangun distribusi nol empiris melalui resampling. Selain itu, uji median menyediakan opsi nonparametrik lain dengan membandingkan median kelompok daripada rata-rata, menawarkan ketahanan terhadap nilai ekstrem [8]. Kemajuan terbaru dalam statistik Bayesian telah memperkenalkan ANOVA Bayesian sebagai kerangka kerja fleksibel yang tidak secara ketat memerlukan normalitas, menyediakan kuantifikasi bukti kontinu melalui faktor Bayes, dan berkinerja baik dengan sampel kecil yang khas dalam analisis regional [9], [10]. Berbeda dengan nilai p frekuentis yang hanya menunjukkan apakah hipotesis nol harus ditolak, faktor Bayes secara langsung mengkuantifikasi bukti relatif untuk hipotesis yang bersaing sambil memasukkan informasi prior dan menghasilkan distribusi probabilitas posterior yang dapat diinterpretasikan [11].

Meskipun ada pengakuan yang semakin luas tentang pluralisme metodologis dalam statistik terapan, perbandingan sistematis antara pendekatan Bayesian dan multiple frequentist untuk analisis sosioekonomi regional masih jarang, terutama menggunakan data Indonesia. Penelitian sebelumnya oleh [12] menganalisis hubungan timbal balik antara investasi langsung asing dan Indeks Pembangunan Manusia (IPM) di berbagai wilayah Indonesia menggunakan model panel ekonometrik, dan menemukan heterogenitas regional yang signifikan dalam pola perkembangan. Laporan ketidaksetaraan Badan Pusat Statistik (BPS) mencatat kesenjangan regional yang semakin melebar, terutama antara kawasan metropolitan berbasis Jawa dan provinsi-provinsi timur, menunjukkan ketidaksetaraan struktural [13]. Namun, studi-studi ini sebagian besar menggunakan pendekatan regresi atau analisis deskriptif tanpa membandingkan secara sistematis metode statistik untuk perbedaan rata-rata kelompok di bawah kondisi asumsi yang berbeda.

Secara metodologis, tutorial terbaru dalam penelitian psikologi dan medis telah menunjukkan keunggulan ANOVA Bayesian untuk studi dengan sampel kecil, dengan [14] menunjukkan bahwa faktor Bayes memberikan interpretasi bukti yang lebih nuansatif daripada ambang batas nilai p biner. Studi yang membandingkan metode parametrik dan nonparametrik menekankan bahwa pemilihan uji harus didasarkan pada karakteristik data, dengan studi simulasi mengungkapkan perbedaan daya dan tingkat kesalahan tipe I

di bawah pelanggaran asumsi yang berbeda (Barzizza et al., 2023) [16]. Terdapat celah penelitian kritis dalam menerapkan wawasan metodologis ini pada data ekonomi regional dunia nyata, di mana ukuran sampel secara inheren dibatasi oleh batas geografis-administratif dan karakteristik data mungkin tidak memenuhi asumsi parametrik.

Studi ini bertujuan mengatasi kesenjangan yang diidentifikasi melalui perbandingan komprehensif lima metode statistik: Bayesian ANOVA, parametric F-ANOVA, uji Kruskal-Wallis, uji permutasi, dan uji median. Untuk mendeteksi perbedaan sosioekonomi di antara tiga kelompok regional di Indonesia menggunakan data provinsi tahun 2024-2025. Terdapat tiga tujuan penelitian ini: pertama, mengevaluasi validitas asumsi melalui uji normalitas dan homogenitas formal, untuk menentukan pemilihan metode yang tepat; kedua, membandingkan kesimpulan statistik antar metode, mengidentifikasi konsistensi dan ketidaksesuaian dalam hasil pengujian hipotesis; dan ketiga, mengukur signifikansi praktis melalui ukuran efek dan kekuatan bukti Bayesian, memungkinkan interpretasi yang relevan bagi kebijakan di luar signifikansi statistik. Keunikan penelitian ini terletak pada perbandingan metodologis komprehensif menggunakan data provinsi Indonesia yang autentik dengan batasan ukuran sampel yang inheren dan potensi pelanggaran asumsi, memberikan panduan berbasis bukti bagi peneliti dan pembuat kebijakan yang terlibat dalam analisis pengembangan regional. Dengan membandingkan secara sistematis pendekatan Bayesian dan frequentist pada lima indikator sosioekonomi dengan sifat distribusi yang bervariasi, kami berkontribusi pada literatur metodologis tentang perbandingan kelompok dan pemahaman substantif tentang tantangan pembangunan regional Indonesia, pada akhirnya mendukung inferensi statistik yang lebih ketat dan sesuai konteks dalam ekonomi regional terapan.

2. METODE PENELITIAN

2.1 Sumber Data dan Variabel

Studi ini menggunakan data sekunder dari Badan Pusat Statistik (BPS) Indonesia untuk periode 2024-2025, mencakup seluruh 38 provinsi. Lima indikator sosioekonomi digunakan sebagai variabel dependen: Indeks Pembangunan Manusia (IPM) untuk tahun 2025, diukur sebagai indeks komposit berkisar antara 0 hingga 100 yang menggabungkan dimensi kesehatan, pendidikan, dan pendapatan; Rasio Gini untuk tahun 2025, menunjukkan ketimpangan pendapatan pada skala dari 0 (kesetaraan sempurna) hingga 1 (ketimpangan sempurna); tingkat kemiskinan untuk tahun 2025, diungkapkan sebagai persentase penduduk yang hidup di bawah garis kemiskinan nasional; Tingkat Pengangguran Terbuka (TPT) untuk tahun 2025, mewakili persentase angkatan kerja yang aktif mencari pekerjaan tetapi tidak dapat menemukannya; dan pertumbuhan ekonomi untuk tahun 2024, diukur sebagai perubahan persentase tahunan Produk Domestik Regional Bruto (PDRB) provinsi pada harga konstan.

Provinsi dikategorikan ke dalam tiga kelompok regional independen berdasarkan klasifikasi geografis-administratif konvensional Indonesia: Sumatera ($n = 10$ provinsi: Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Bangka Belitung, dan Kepulauan Riau), Jawa-Bali ($n = 7$ provinsi:

DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, dan Bali), dan Lainnya atau wilayah lain ($n = 21$ provinsi: lima provinsi masing-masing di wilayah Kalimantan, Sulawesi, dan Papua, ditambah dua di Nusa Tenggara dan dua di Maluku). Semua data lengkap di 38 provinsi untuk setiap indikator, menghilangkan kekhawatiran data yang hilang dan memastikan validitas analitis yang seimbang.

2.2 Uji Asumsi

Sebelum pengujian parametrik, kami mengevaluasi dua asumsi dasar ANOVA: normalitas distribusi dalam kelompok dan homogenitas varians antar kelompok. Normalitas dievaluasi secara global menggunakan uji Shapiro-Wilk [17], yang menganalisis korelasi antara data observasi dan statistik urutan normal yang diharapkan dengan statistik uji [18]:

H_0 : data berdistribusi normal

H_1 : data tidak berdistribusi normal

$$W = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (1)$$

dimana $x_{(i)}$ adalah observasi yang diurutkan dan a_i adalah bobot dari statistik urutan normal yang diharapkan. Untuk setiap uji, tingkat signifikansi di bawah $\alpha = 0,05$ menunjukkan penyimpangan dari normalitas.

Homogenitas varians dievaluasi menggunakan uji Levene [19], yang melakukan ANOVA pada penyimpangan absolut dari median kelompok [20], dihitung sebagai:

H_0 : ragam data sama (homogen/ *equal varian*)

H_1 : ragam data tidak sama (heterogen/ *unequal varian*)

$$F = \frac{(N-k) \cdot \sum_{i=1}^k n_i (Z_i - Z_{..})^2}{(k-1) \cdot \sum_{i=1}^k \sum_{j=1}^{n_i} (Z_{ij} - Z_i)^2} \quad (2)$$

dimana $Z_{ij} = |Y_{ij} - \tilde{Y}_i|$ mewakili penyimpangan absolut dari median kelompok, dengan $k = 3$ kelompok dan $N = 38$ pengamatan total.

2.3 Metode Statistik

ANOVA Parametrik F (ANOVA Fisher)

Penelitian ini menggunakan analisis variansi satu arah untuk menguji hipotesis nol:

$H_0: \mu_1 = \mu_2 = \mu_3$ bahwa rata-rata populasi ketiga kelompok regional sama

H_1 : setidaknya 1 pasang $\mu_i \neq \mu_j$

Statistik F [21] dihitung sebagai rasio antara kuadrat rata-rata antar-kelompok dan kuadrat rata-rata dalam-kelompok:

$$F = \frac{MS_{between}}{MS_{within}} = \frac{SS_{between}/(k-1)}{SS_{within}/(N-k)} = \frac{\sum_{i=1}^k n_i (\bar{X}_i - \bar{X})^2 / (k-1)}{\sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2 / (N-k)} \quad (3)$$

dimana $SS_{between}$ mengukur variansi yang dijelaskan oleh pengelompokan regional dan SS_{within} mengukur variansi sisa dalam kelompok. Ketika asumsi homogenitas tidak terpenuhi, maka dapat digunakan ANOVA Welch, yang menyesuaikan derajat kebebasan menggunakan perkiraan Welch [22] untuk memberikan inferensi yang andal di bawah asumsi heteroskedastisitas.

Uji Kruskal-Wallis

Uji nonparametrik bebas distribusi ini digunakan untuk menguji apakah sampel berasal dari populasi yang identik dengan membandingkan jumlah peringkat antar kelompok, hanya memerlukan pengukuran ordinal. Semua N pengamatan terlebih dahulu diurutkan dari 1 hingga N tanpa memandang keanggotaan kelompok, dengan peringkat rata-rata diberikan untuk pengamatan yang sama. Statistik uji [6] kemudian dihitung sebagai:

$H_0: \tilde{\mu}_1 = \tilde{\mu}_2 = \tilde{\mu}_3$ bahwa median populasi ketiga kelompok regional sama

H_1 : setidaknya 1 pasang $\tilde{\mu}_i \neq \tilde{\mu}_j$

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1) = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{(\sum_{j=1}^{n_i} \text{rank}(X_{ij}))^2}{n_i} - 3(N+1) \quad (4)$$

di mana R_i menandakan jumlah peringkat untuk kelompok i dan n_i adalah ukuran sampel kelompok. Di bawah hipotesis nol distribusi identik, H mendekati distribusi chi-kuadrat dengan $k-1 = 2$ derajat kebebasan untuk sampel sedang hingga besar.

Uji Permutasi

Inferensi berbasis permutasi [23] menyediakan pendekatan bebas distribusi yang tepat dengan membangun distribusi nol empiris melalui resampling ekshaustif atau acak dari penugasan kelompok. Untuk setiap dari 10.000 permutasi acak, kami mengacak label kelompok regional sambil mempertahankan nilai data yang diamati, lalu menghitung ulang statistik F berdasarkan penugasan kelompok yang diubah:

$$F_{perm} = \frac{MS_{between,perm}}{MS_{within,perm}} \quad (5)$$

Prosedur ini menghasilkan distribusi empiris statistik F yang mewakili variabilitas sampling di bawah hipotesis nol tidak adanya perbedaan regional, tanpa bergantung pada asumsi distribusi parametrik [24]. Nilai p permutasi dihitung sebagai proporsi statistik F yang dipertukarkan yang sama dengan atau melebihi statistik F yang diamati:

$$p_{perm} = \frac{1 + \sum_{b=1}^B I(F_{perm,b} \geq F_{obs})}{B+1} \quad (6)$$

dimana $B = 10.000$ permutasi dan $I(\cdot)$ adalah fungsi indikator.

Uji Median

Uji median mengevaluasi apakah k kelompok memiliki median populasi yang sama melalui uji chi-kuadrat kemandirian yang diterapkan pada tabel kontingensi $2 \times k$. Pertama, median keseluruhan dihitung untuk semua $N = 38$ pengamatan. Setiap pengamatan kemudian diklasifikasikan sebagai di atas atau sama dengan/di bawah median keseluruhan, menciptakan variabel kategorikal biner [8]. Tabel kontingensi dibangun dengan menyeberangkan kelompok regional (3 kategori) dengan posisi median (2 kategori: di atas vs. sama dengan/di bawah median keseluruhan), menghasilkan enam sel dengan frekuensi teramati O_{ij} . Statistik chi-kuadrat dihitung sebagai [25]:

$H_0: \tilde{\mu}_1 = \tilde{\mu}_2 = \tilde{\mu}_3$ bahwa median populasi ketiga kelompok regional sama

H_1 : setidaknya 1 pasang $\tilde{\mu}_i \neq \tilde{\mu}_j$

$$\chi^2 = \sum_{i=1}^k \sum_{j=1}^2 \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \sum_{i=1}^k \sum_{j=1}^2 \frac{(O_{ij} - \frac{n_{i.} \cdot n_{.j}}{N})^2}{\frac{n_{i.} \cdot n_{.j}}{N}} \quad (7)$$

di mana E_{ij} mewakili frekuensi yang diharapkan di bawah asumsi kemandirian, dengan $n_{i.}$ dan $n_{.j}$ menandakan total marjinal baris dan kolom.

Analisis Varian Bayesian:

Analisis varian Bayesian memperkirakan distribusi posterior untuk rata-rata dan varians kelompok dengan menggabungkan distribusi prior dengan data yang diamati melalui teorema Bayes [9]. Penelitian ini menggunakan distribusi prior default yang lemah informatif sebagaimana diimplementasikan dalam perangkat lunak JASP: distribusi Cauchy multivariat untuk ukuran efek yang distandarisasi dengan parameter skala $r = \frac{\sqrt{2}}{2}$, mencerminkan keyakinan prior moderat bahwa ukuran efek berkumpul di sekitar nol tetapi memungkinkan massa probabilitas yang signifikan untuk efek yang lebih besar. Faktor Bayes BF_{10} mengukur bukti relatif untuk hipotesis alternatif H_1 (perbedaan regional ada) versus hipotesis nol H_0 (tidak ada perbedaan) melalui rasio kemungkinan marjinal:

$$BF_{10} = \frac{P(Data|H_1)}{P(Data|H_0)} = \frac{\int P(Data|\theta_1, H_1)P(\theta_1|H_1)d\theta_1}{\int P(Data|\theta_0, H_0)P(\theta_0|H_0)d\theta_0} \quad (8)$$

di mana integrasi rata-rata atas ruang parameter θ_1 dan θ_0 yang diweighted oleh densitas prior. Mengikuti konvensi Jeffreys (1961) yang disempurnakan oleh praktik Bayesian modern, penelitian ini menafsirkan faktor Bayes sebagai: $BF_{10} < 1$ mendukung H_0 , 1-3 memberikan bukti anekdot untuk H_1 , 3-10 menunjukkan bukti moderat, 10-30 menyarankan bukti kuat, 30-100 mewakili bukti sangat kuat, dan $BF_{10} > 100$ merupakan bukti menentukan untuk perbedaan regional.

2.4 Uji Post Hoc dan Ukuran Efek

Ketika uji omnibus menunjukkan perbedaan regional yang signifikan ($p < 0,05$ untuk uji frekuentis atau $BF_{10} > 3$ untuk analisis Bayesian), penelitian ini menggunakan perbandingan berpasangan post hoc untuk mengidentifikasi pasangan regional mana yang berbeda secara signifikan. Untuk ANOVA parametrik, kami menerapkan uji Tukey's Honestly Significant Difference (HSD) [26], yang mengontrol tingkat kesalahan keluarga (family-wise error rate) pada $\alpha = 0,05$ untuk semua perbandingan berpasangan ($\binom{k}{2} = 3$) dengan menyesuaikan nilai kritis berdasarkan distribusi rentang yang disesuaikan (studentized range distribution). Perbedaan kritis HSD dihitung sebagai:

$H_0: \mu_i = \mu_j$ bahwa rata-rata populasi kedua kelompok regional sama

$H_1: \mu_i \neq \mu_j$ bahwa rata-rata populasi kedua kelompok regional tidak sama

$$HSD = q_{\alpha, k, N-k} \sqrt{\frac{MS_{within}}{n_h}} \quad (9)$$

di mana $q_{\alpha, k, N-k}$ adalah statistik rentang studentized dan n_h adalah rata-rata harmonik ukuran sampel kelompok untuk kelompok yang tidak sama. Untuk uji Kruskal-Wallis, riset ini menggunakan uji post hoc Dunn [27] dimana uji Dunn digunakan untuk membandingkan kelompok i dan j adalah:

$H_0: \tilde{\mu}_i = \tilde{\mu}_j$ bahwa median populasi kedua kelompok regional sama

$H_1: \tilde{\mu}_i \neq \tilde{\mu}_j$ bahwa median populasi kedua kelompok regional tidak sama

$$z_{ij} = \frac{|\bar{R}_i - \bar{R}_j|}{\sqrt{\frac{N(N+1)}{12} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}} \quad (10)$$

di mana $\bar{R}_i$ menandakan rata-rata peringkat untuk kelompok i . Untuk ANOVA Bayesian, penelitian ini menganalisis rasio peluang posterior yang disesuaikan untuk pengujian multipel mengikuti Westfall et al [28], dengan $BF_{10} > 3$ menunjukkan bukti yang signifikan untuk perbedaan antarpasangan.

Ukuran efek mengukur besarnya dan signifikansi praktis ketidakmerataan regional di luar pengujian signifikansi statistik. Penelitian ini menghitung Cohen's d untuk perbandingan berpasangan menggunakan rumus simpangan baku gabungan:

$$d = \frac{\bar{X}_i - \bar{X}_j}{s_{pooled}} = \frac{\bar{X}_i - \bar{X}_j}{\sqrt{\frac{(n_i-1)s_i^2 + (n_j-1)s_j^2}{n_i + n_j - 2}}} \quad (11)$$

Mengikuti konvensi Cohen [29], kami menginterpretasikan $|d| < 0,5$ sebagai efek kecil, $0,5 \leq |d| < 0,8$ sebagai efek sedang, dan $|d| \geq 0,8$ sebagai efek besar. Untuk uji Kruskal-Wallis, penelitian ini menghitung epsilon-squared (ϵ^2) [30] sebagai ukuran efek yang mewakili proporsi varians total yang dijelaskan oleh perbedaan antar kelompok:

$$\epsilon^2 = \frac{H - k + 1}{N - k} \quad (12)$$

di mana H adalah statistik Kruskal-Wallis. Semua analisis dilakukan menggunakan JASP versi 0.18.3 untuk ANOVA Bayesian dan Jamovi versi 2.5 untuk uji frekuentis, dengan skrip R kustom (versi 4.3.2) untuk uji permutasi dan median menggunakan paket termasuk stats, coin, dan DescTools, memastikan reproduibilitas penuh dan transparansi komputasi.

3. HASIL DAN PEMBAHASAN

Analisis deskriptif menunjukkan heterogenitas yang signifikan dalam indikator sosioekonomi di antara tiga kelompok regional, dengan perbedaan yang mencolok dalam kecenderungan pusat dan variabilitas, seperti yang ditunjukkan dalam Tabel 1. Indeks Pembangunan Manusia (IPM) untuk tahun 2025 menunjukkan hierarki regional yang jelas, dengan Jawa-Bali mencapai rata-rata tertinggi 78.71 (SD = 3.81, rentang: 74.77-85.05), mewakili wilayah paling berkembang di Indonesia yang didukung oleh IPM luar biasa DKI Jakarta sebesar 85.05, satu-satunya provinsi yang melebihi 85.00 secara nasional. Sumatera menempati posisi menengah dengan rata-rata IPM 76.16 (SD = 1.80, rentang: 73.98-80.53), menunjukkan variasi yang jauh lebih kecil dibandingkan Jawa-Bali, kemungkinan mencerminkan tingkat perkembangan yang lebih homogen di antara provinsi-provinsi Sumatera. Wilayah lain mencatat rata-rata IPM terendah sebesar 71.85 (SD = 5.41, rentang: 54.91-79.39), dengan simpangan baku yang signifikan menunjukkan ketimpangan ekstrem antara provinsi yang relatif berkembang seperti Kalimantan Timur (IPM = 79.39) dan Papua Pegunungan yang tertinggal parah (IPM = 54.91), provinsi dengan peringkat terendah di negara ini yang mewakili kesenjangan pembangunan hampir 30 poin IPM.

Ketidaksetaraan pendapatan yang diukur dengan rasio Gini pada tahun 2025 menunjukkan pola yang tidak terduga, dengan wilayah Jawa-Bali yang paling maju mencatat ketidaksetaraan rata-rata tertinggi sebesar 0.385 (SD = 0.042, rentang: 0.330-0.441), didominasi oleh rasio Gini DKI Jakarta sebesar 0.441, provinsi paling tidak merata di negara ini mencerminkan konsentrasi pendapatan yang ekstrem di ibu kota. Wilayah lain menunjukkan ketimpangan menengah sebesar 0.335 (SD = 0.043), sementara Sumatera mencatat rata-rata Gini terendah sebesar 0.302 (SD = 0.042, rentang: 0.222-0.382), menunjukkan distribusi pendapatan yang lebih merata meskipun tingkat perkembangan keseluruhan lebih rendah, dengan Kepulauan Bangka Belitung mencatat distribusi pendapatan paling merata di negara ini sebesar 0.222.

Tabel 1. Statistik Deskriptif Indikator Sosial Ekonomi Menurut Wilayah

Variabel	Wilayah	<i>n</i>	<i>mean</i>	<i>SD</i>	Min	Maks
IPM (2025)	Sumatera	10	76.16	1.80	73.98	80.53
	Jawa-Bali	7	78.71	3.81	74.77	85.05
	Lainnya	21	71.85	5.41	54.91	79.39
Gini (2025)	Sumatera	10	0.302	0.042	0.222	0.382
	Jawa-Bali	7	0.385	0.042	0.330	0.441
	Lainnya	21	0.335	0.043	0.261	0.412
Kemiskinan (2025)	Sumatera	10	8.01	2.93	4.44	12.33
	Jawa-Bali	7	7.12	2.67	3.72	10.23
	Lainnya	21	13.01	7.73	3.84	30.03
TPT (2025)	Sumatera	10	4.72	0.98	3.41	6.45
	Jawa-Bali	7	4.71	1.94	1.49	6.77
	Lainnya	21	4.27	1.37	1.68	6.96
Pertumbuhan Ekonomi (2024)	Sumatera	10	4.21	1.29	0.77	5.03
	Jawa-Bali	7	5.00	0.22	4.79	5.48
	Lainnya	21	6.19	4.05	3.60	20.80

Sumber: Badan Pusat Statistik (data diolah)

Tingkat kemiskinan untuk tahun 2025 menunjukkan ketimpangan regional yang paling dramatis, dengan wilayah lain mengalami rata-rata kemiskinan sebesar 13.01% (SD = 7.73, rentang: 3.84-30.03%), lebih dari dua kali lipat tingkat kemiskinan di Sumatra (8.01%, SD = 2.93) dan Jawa-Bali (7.12%, SD = 2.67). Angka maksimum ekstrem 30,03% di Papua Pegunungan, di mana hampir sepertiga penduduknya hidup di bawah garis kemiskinan, kontras secara mencolok dengan angka 3.72% di Bali, menggambarkan kemiskinan sebagai indikator yang paling terpolarisasi di lanskap regional Indonesia.

Tingkat pengangguran (TPT) untuk tahun 2025 menunjukkan variasi regional yang relatif moderat, dengan Sumatera (4.72%, SD = 0.98), Jawa-Bali (4.71%, SD = 1.94), dan wilayah lain (4.27%, SD = 1.37) menunjukkan rata-rata yang serupa berkisar di sekitar 4.5%, meskipun simpangan baku Jawa-Bali yang lebih besar mencerminkan heterogenitas provinsi yang lebih besar, dari 1.49% yang sangat rendah di Bali hingga 6.77% di Jawa Barat. Pertumbuhan ekonomi pada tahun 2024 menunjukkan koefisien variasi terbesar di wilayah lain (rata-rata = 6.19%, SD = 4.05), didorong oleh pertumbuhan luar biasa Papua Barat sebesar 20.80% yang mencerminkan booming ekstraksi sumber daya di Indonesia timur, berbanding terbalik dengan pertumbuhan stabil dan seragam Jawa-Bali sekitar 5.00% (SD = 0.22) dan pertumbuhan intermediet Sumatera sebesar 4.21% (SD = 1.29).

Pada Tabel 2 menunjukkan Uji normalitas formal menghasilkan hasil yang bervariasi di antara indikator dan kelompok regional, menunjukkan penyimpangan yang signifikan dari normalitas untuk beberapa variabel seperti tercantum dalam Tabel 2. Uji *Shapiro-Wilk* global menunjukkan bahwa Indeks Pembangunan Manusia (IPM), tingkat kemiskinan, dan pertumbuhan ekonomi melanggar asumsi normalitas pada $\alpha = 0.05$, sementara gini rasio dan tingkat pengangguran memenuhi asumsi normalitas. Uji *Shapiro-Wilk* spesifik kelompok menunjukkan pola heterogen.

Tabel 2. Ringkasan Hasil Uji Normalitas Menurut Variabel dan Wilayah

Variabel	Global	Hasil Spesifik Wilayah		
	Shapiro-Wilk (W, p)	Sumatera	Jawa-Bali	Lainnya
IPM	0.856. p=0.001**	W=0.862. p=0.080	W=0.901. p=0.338	W=0.808. p=0.001**
Gini	0.980. p=0.712	W=0.929. p=0.434	W=0.917. p=0.447	W=0.967. p=0.665
Kemiskinan	0.935. p=0.028*	W=0.906. p=0.257	W=0.894. p=0.294	W=0.901. p=0.036*
TPT	0.967. p=0.323	W=0.935. p=0.495	W=0.927. p=0.525	W=0.950. p=0.340
Pertumbuhan Ekonomi	0.623. p<0.001**	W=0.645. p=0.001**	W=0.755. p=0.014*	W=0.564. p<0.001**

Catatan: ** = $p < 0,01$; * = $p < 0,05$.

Uji homogenitas varians menunjukkan pola yang kontras, dengan sebagian besar indikator memenuhi asumsi seperti yang ditunjukkan dalam Tabel 3. Uji Levene menunjukkan varians homogen untuk IPM, rasio Gini, tingkat pengangguran, dan pertumbuhan ekonomi meskipun pertumbuhan mendekati ambang batas signifikansi. Namun, tingkat kemiskinan menunjukkan heteroskedastisitas yang signifikan. Uji asumsi gabungan menunjukkan bahwa ANOVA parametrik berpotensi valid untuk rasio Gini dan tingkat pengangguran, yang memenuhi kedua asumsi normalitas dan homogenitas, tetapi dipertanyakan untuk IPM (variansi normal tetapi distribusi non-normal), kemiskinan (distribusi normal tetapi variansi heterogen), dan terutama pertumbuhan ekonomi (melanggar kedua asumsi secara serius), sehingga memerlukan alternatif nonparametrik atau Bayesian untuk inferensi yang *robust*.

Tabel 3. Hasil Uji Homogenitas Varians (Uji Levene)

Variabel	Statistik F	df1	df2	nilai p	Homogen
IPM	1.968	2	35	0.155	Ya
Gini	0.494	2	35	0.615	Ya
Kemiskinan	6.838	2	35	0.003**	Tidak
TPT	1.651	2	35	0.206	Ya
Pertumbuhan Ekonomi	3.048	2	35	0.060	Ya

Tabel 4 menampilkan perbandingan komprehensif antar lima metode statistik untuk semua indikator sosioekonomi, menunjukkan konsistensi yang signifikan untuk IPM dan rasio Gini, namun kesimpulan yang bergantung pada metode untuk variabel lain. Untuk IPM, semua metode secara bulat mendeteksi perbedaan regional yang sangat signifikan. Konsistensi yang mencolok di antara metode-metode tersebut meskipun pelanggaran normalitas IPM di wilayah lain menunjukkan ketahanan semua pendekatan terhadap

ketidaknormalan moderat ketika ukuran sampel kelompok melebihi 7-10 pengamatan dan perbedaan cukup besar. Rasio Gini juga menunjukkan perbedaan signifikan yang konsisten di seluruh metode. Hal ini memberikan bukti konvergen yang kuat tentang ketidaksetaraan pendapatan regional meskipun BF Bayesien sedikit lebih rendah daripada untuk IPM, menunjukkan kekuatan bukti yang sedikit kurang meyakinkan.

Tabel 4. Perbandingan Komprehensif Metode Statistik pada Lima Indikator

Variabel	F-ANOVA		Kruskal-Wallis		Permutasi		Uji Median		Bayesian	
	F	p	H, ϵ^2	p	F	p	χ^2	p	BF ₁₀	Penafsiran
IPM	7.390	0.002	16.11; 0.435	<0.001	7.390	0.0024	12.933	0.0016	21.22	Kuat
Gini	7.764	0.002	10.57; 0.286	0.005	7.764	0.0013	7.219	0.0271	17.28	Kuat
Kemiskinan	3.654	0.036	5.60; 0.151	0.061	3.654	0.0332	2.876	0.2374	2.34	Anekdotal
TPT	0.490	0.617	1.56; 0.042	0.459	0.490	0.6188	1.633	0.4421	0.28	Moderat (H ₀)
Pertumbuhan Ekonomi	1.443	0.250	3.59; 0.097	0.166	1.443	0.2189	2.836	0.2422	0.55	Anekdotal (H ₀)

Tingkat kemiskinan menunjukkan hasil yang bergantung pada metode, yang menyoroti pentingnya pemilihan metode berdasarkan asumsi. Uji F-ANOVA parametrik mendeteksi signifikansi marjinal demikian pula uji permutasi, keduanya melampaui ambang batas $\alpha = 0.05$. Namun, uji *Kruskal-Wallis* menghasilkan hasil yang tidak signifikan, mendekati tetapi tidak mencapai signifikansi, sementara uji median menunjukkan bukti yang lebih lemah. Ketidaksesuaian ini kemungkinan mencerminkan heteroskedastisitas yang parah pada kemiskinan serta dikombinasikan dengan kemiringan positif ekstrem di wilayah lain di kelompok lain), yang secara berbeda memengaruhi uji berbasis peringkat versus uji berbasis rata-rata. ANOVA Bayesien memberikan $BF_{10} = 2.34$, diklasifikasikan sebagai bukti anekdotal untuk perbedaan regional, menunjukkan bahwa data hanya 2,3 kali lebih mungkin di bawah H₁ daripada H₀, tidak cukup untuk kesimpulan yang pasti.

Kegagalan uji median dalam mendeteksi perbedaan menunjukkan bahwa median regional lebih serupa daripada rata-rata, dengan nilai kemiskinan ekstrem di provinsi Papua yang meningkatkan rata-rata di wilayah lain tetapi tidak mengubah median secara signifikan. Pola ini menggambarkan bagaimana pilihan metode secara kritis memengaruhi kesimpulan ketika asumsi dilanggar, dengan uji parametrik berpotensi terinfeksi oleh heteroskedastisitas dan uji median terlalu konservatif dengan mengabaikan informasi *magnitude*.

Tingkat pengangguran dan pertumbuhan ekonomi menunjukkan hasil yang tidak signifikan secara konsisten di semua metode, meskipun dengan alasan yang berbeda. Untuk TPT, kelima metode menunjukkan hasil yang tidak signifikan. Pada faktor Bayes menunjukkan bukti moderat yang mendukung hipotesis nol, menyarankan bahwa tingkat pengangguran benar-benar serupa di seluruh wilayah. Konsistensi antar metode, dikombinasikan dengan TPT yang memenuhi asumsi normalitas dan homogenitas, memberikan keyakinan tinggi bahwa pengelompokan regional menjelaskan varians minimal dalam pola pengangguran, yang kemungkinan lebih bergantung pada struktur

ekonomi lokal, ketidakcocokan pendidikan-pekerjaan, dan kebijakan tenaga kerja provinsi daripada geografi regional yang luas.

Pertumbuhan ekonomi juga tidak menunjukkan perbedaan yang signifikan, meskipun kesimpulan ini rumit karena pelanggaran asumsi yang parah, termasuk ketidaknormalan ekstrem dan adanya *outlier* Papua Barat sebesar 20.80% yang berada 3.7 standar deviasi di atas rata-rata wilayah lain. Faktor Bayes sebesar 0.55 menunjukkan bukti lemah hingga anekdotal yang mendukung tidak adanya perbedaan regional, meskipun mungkin dipertanyakan apakah ini mencerminkan kesamaan sejati atau keterbatasan daya statistik akibat heterogenitas varians yang ekstrem.

Selanjutnya pada Tabel 5 Analisis *post hoc* untuk dua variabel yang menunjukkan uji omnibus signifikan yang konsisten yaitu IPM dan rasio Gini. Hal ini mengidentifikasi pasangan regional spesifik yang mendasari perbedaan yang diamati, seperti terperinci dalam Tabel 5. Untuk IPM, uji *Tukey's HSD* menunjukkan bahwa daerah lainnya secara signifikan berbeda dengan wilayah Sumatera dan Jawa Bali. Uji *post hoc Dunn* setelah *Kruskal-Wallis* mengonfirmasi temuan ini dengan pola serupa. Perbandingan *post hoc Bayesian* menggunakan rasio peluang posterior mengonfirmasi bukti yang meyakinkan untuk perbedaan wilayah lainnya terhadap Sumatera dan Jawa-Bali.

Tabel 5. Perbandingan Pasangan Pasca-Uji untuk IPM dan Rasio Gini

Variabel	Perbandingan	Selisih Rata-rata	<i>Tukey p</i>	<i>Dunn p</i>	BF ₁₀ U	<i>Cohen's d</i>
IPM	Jawa-Bali vs Lainnya	6.86	0.004	0.001	8.69	1.52 (Besar)
	Jawa-Bali vs Sumatera	2.55	0.489	1.000	1.26	0.79 (Sedang)
	Sumatera vs Lainnya	4.31	0.044	0.012	2.92	1.03 (Besar)
Gini	Jawa-Bali vs Lainnya	0.050	0.029	0.036	4.20	1.16 (Besar)
	Jawa-Bali vs Sumatera	0.083	0.001	0.003	24.51	1.96 (Besar)
	Sumatera vs Lainnya	0.033	0.125	0.220	1.53	0.78 (Sedang)

Analisis ukuran efek mengukur signifikansi praktis ketidaksetaraan IPM regional, menunjukkan efek besar untuk perbandingan kunci. *Cohen's d* untuk wilayah lainnya terhadap Sumatera dan Jawa-Bali nilainya di atas 1. Untuk rasio Gini, perbandingan *post hoc* menunjukkan pola yang berbeda yang menonjolkan ketidaksetaraan ekstrem Jawa-Bali. Uji *Tukey's HSD* dan Uji *Dunn* mendeteksi perbedaan signifikan antara Jawa-Bali dan Sumatera. Pola ini menunjukkan bahwa Jawa-Bali menonjol sebagai wilayah yang unik dalam ketidaksetaraan, sementara Sumatera dan wilayah lain menunjukkan tingkat ketidaksetaraan yang serupa meskipun memiliki profil IPM yang berbeda. Perbandingan *post hoc Bayesian* secara kuat mendukung interpretasi ini, dengan perbandingan Jawa-Bali *versus* Sumatera menghasilkan faktor Bayes yang tidak disesuaikan terbesar dalam studi ini. Hal ini menunjukkan bukti yang sangat kuat untuk perbedaan ini.

Untuk tingkat kemiskinan, di mana hasil uji omnibus bergantung pada metode dan marginal, perbandingan *post hoc* eksploratif (tidak ditampilkan dalam tabel karena uji *Kruskal-Wallis omnibus* tidak signifikan) menyarankan bahwa signifikansi ANOVA parametrik didorong terutama oleh perbandingan antara wilayah lain dan dua kelompok lain. Dengan uji *Dunn* menunjukkan bukti signifikan marginal untuk Jawa-Bali *versus* wilayah lain

Perbandingan metodologis komprehensif mengungkapkan beberapa wawasan kritis bagi penelitian ekonomi regional terapan dan analisis kebijakan. Pertama, tidak ada metode statistik tunggal yang mendominasi di semua skenario; sebaliknya, kesesuaian metode bergantung pada karakteristik data dan tujuan penelitian. Ketika asumsi parametrik terpenuhi (seperti pada gini rasio dan TPT), F-ANOVA parametrik memberikan daya statistik maksimum dan interpretasi yang familiar. Namun, ketika asumsi dilanggar (seperti pada non-normalitas IPM dan heteroskedastisitas kemiskinan), alternatif nonparametrik seperti *Kruskal-Wallis* menawarkan inferensi valid yang tahan terhadap penyimpangan distribusi, sementara uji permutasi memberikan nilai p yang tepat tanpa bergantung pada aproksimasi asimtotik. Hasil tes median yang secara konsisten konservatif (hanya mendeteksi perbedaan IPM dan secara marginal perbedaan gini) menggambarkan biaya daya statistik dari mengabaikan informasi magnitudo, menyarankan agar tes ini disimpan untuk situasi dengan outlier parah atau distribusi ekor berat di mana median lebih baik mewakili kecenderungan pusat daripada rata-rata.

Kedua, ANOVA Bayesian memberikan keunggulan inferensial unik melalui kuantifikasi bukti kontinu via faktor Bayes dan pernyataan probabilitas langsung via distribusi posterior. Berbeda dengan nilai p frekuentis yang hanya menunjukkan apakah hipotesis nol ditolak pada ambang batas arbitrer, faktor Bayes mengkuantifikasi kekuatan bukti pada skala yang dapat diinterpretasikan, dengan $BF_{10} = 21.22$ untuk IPM menunjukkan bukti yang jauh lebih kuat daripada $BF_{10} = 2.34$ untuk kemiskinan meskipun keduanya melampaui ambang batas signifikansi konvensional. Selain itu, metode Bayesian secara alami mendukung hipotesis nol ketika $BF_{10} < 1$ (seperti pada TPT dengan $BF_{10} = 0.28$), memberikan bukti afirmatif tentang tidak adanya perbedaan regional daripada sekadar kegagalan untuk menolak, yang lebih sesuai dengan kebutuhan kebijakan untuk membedakan "tidak ada bukti efek" dari "bukti tidak adanya efek".

Dari perspektif kebijakan substantif, deteksi konsisten perbedaan IPM yang besar di semua metode ($BF_{10} = 21.22$, Cohen's d hingga 1.52) memberikan bukti yang kuat yang memerlukan intervensi pengembangan regional yang ditargetkan. Selisih IPM sebesar 6.86 poin antara Jawa-Bali (78.71) dan wilayah lain (71.85) setara dengan sekitar 10-12 tahun kemajuan IPM berdasarkan laju pertumbuhan historis Indonesia sebesar 0,55 poin per tahun, menunjukkan bahwa wilayah lain, terutama provinsi-provinsi timur, tertinggal lebih dari satu dekade dalam pembangunan manusia dibandingkan Jawa-Bali. Kasus ekstrem Papua Pegunungan (IPM = 54.91) yang tertinggal 30 poin dari DKI Jakarta (85.05) mewakili jurang pembangunan yang mengkhawatirkan, sebanding dengan perbedaan antara negara berpendapatan menengah bawah dan negara berpendapatan tinggi secara global, yang menuntut intervensi multisektor intensif dalam akses pendidikan, infrastruktur kesehatan, dan diversifikasi ekonomi di Papua dan provinsi timur terpencil. Sebaliknya, temuan yang bertentangan dengan intuisi tentang rasio Gini yang lebih tinggi di Jawa-Bali yang lebih maju (0.385 versus 0.302 di Sumatera) sejalan dengan hipotesis Kuznets (1955) bahwa perkembangan ekonomi awalnya meningkatkan ketidaksetaraan sebelum akhirnya menurun, menunjukkan bahwa pertumbuhan cepat di kawasan metropolitan berbasis Jawa telah mengkonsentrasikan manfaat di kalangan

pekerja terampil dan pemilik modal sementara meninggalkan pekerja informal. Polanya ini menuntut kebijakan redistribusi progresif, termasuk pajak kekayaan, perlindungan sosial yang ditingkatkan, dan penegakan upah minimum, khususnya di kawasan metropolitan dengan pertumbuhan tinggi seperti DKI Jakarta (Gini = 0.441), di mana konsentrasi pendapatan paling ekstrem.

Ketiadaan perbedaan pengangguran regional yang signifikan ($BF_{10} = 0.28$ mendukung tidak adanya perbedaan) meskipun ada kesenjangan Indeks Pembangunan Manusia (IPM) menunjukkan bahwa pengangguran didorong terutama oleh ketidakcocokan keterampilan, komposisi industri, dan dinamika pasar tenaga kerja lokal daripada geografi regional yang luas. Temuan ini menyiratkan bahwa program pengangguran regional yang seragam kemungkinan tidak efisien; sebaliknya, kebijakan harus menargetkan provinsi tertentu berdasarkan faktor pengangguran lokal, misalnya, memperluas pelatihan vokasi di provinsi yang bergantung pada manufaktur seperti Jawa Barat (TPT = 6.77%) sambil mendukung diversifikasi pertanian di provinsi yang bergantung pada sumber daya dengan pengangguran struktural. Demikian pula, kurangnya perbedaan regional yang signifikan dalam pertumbuhan ekonomi ($BF_{10} = 0.55$) tampaknya didorong oleh heterogenitas ekstrem di dalam wilayah, terutama pertumbuhan *boom-and-bust* sebesar 20,80% di Papua Barat akibat ekstraksi sumber daya alam, menyoroti ketidakberlanjutan jalur pengembangan yang bergantung pada sumber daya dan kebutuhan strategi diversifikasi ekonomi di provinsi-provinsi kaya sumber daya untuk mencapai pertumbuhan yang stabil dan inklusif daripada ekspansi volatil yang menguntungkan investor eksternal.

4. KESIMPULAN

Perbandingan metodologis komprehensif antara pendekatan *Bayesian* dan *frequentist* dalam menganalisis ketimpangan sosioekonomi regional di Indonesia menunjukkan bahwa pemilihan metode sangat bergantung pada karakteristik data, dan tidak ada pendekatan yang secara universal optimal untuk semua indikator. Ketika asumsi parametrik terpenuhi, F-ANOVA memberikan daya statistik maksimum dan interpretasi yang familiar, tetapi ketika asumsi dilanggar seperti pada ketidaknormalan IPM dan heteroskedastisitas kemiskinan alternatif nonparametrik termasuk *Kruskal-Wallis*, permutasi, dan uji median memastikan inferensi yang valid dan tahan terhadap penyimpangan distribusi. ANOVA Bayesian menawarkan keunggulan unik melalui kuantifikasi bukti kontinu melalui faktor Bayes untuk IPM menunjukkan bukti yang meyakinkan; untuk kemiskinan menunjukkan bukti anekdot) dan kemampuan untuk mendukung hipotesis nol secara afirmatif. Untuk variabel pengangguran memberikan bukti moderat untuk tidak adanya perbedaan regional), mengatasi keterbatasan nilai p frekuentis yang hanya menolak atau gagal menolak tanpa mengkuantifikasi kekuatan atau arah bukti.

Pembuat kebijakan harus memprioritaskan investasi pengembangan manusia yang intensif di wilayah timur yang tertinggal, terutama provinsi Papua di mana defisit IPM melebihi 30 poin, sambil secara bersamaan mengatasi ketimpangan pendapatan di

wilayah metropolitan Jawa-Bali melalui pajak progresif, penegakan upah minimum, dan perluasan perlindungan sosial untuk memastikan manfaat pertumbuhan yang inklusif rather than terkonsentrasi. Batasan studi termasuk desain *cross-sectional* yang menghalangi inferensi kausal tentang proses perkembangan temporal, penggunaan pengelompokan regional tiga kategori yang luas yang menyembunyikan heterogenitas substansial di dalam wilayah, terutama dalam kategori "wilayah lain" yang beragam mencakup 21 provinsi.

DAFTAR PUSTAKA

- [1] H. Hill, B. P. Resosudarmo, and Y. Vidyattama, "Indonesia's changing economic geography," *Bull. Indones. Econ. Stud.*, vol. 44, no. 3, pp. 407–435, 2008, doi: 10.1080/00074910802395344.
- [2] Badan Pusat Statistik, "Indonesia's human development index (HDI) in 2025 reached 75.90." 2025. [Online]. Available: <https://www.bps.go.id/en/pressrelease/2025/11/05/2480>
- [3] Badan Pusat Statistik, *Statistical Yearbook of Indonesia 2025*. Jakarta: BPS-Statistics Indonesia, 2025.
- [4] D. C. Montgomery, *Design and analysis of experiments*, 10th ed. Hoboken: Wiley, 2020.
- [5] M. W. Fagerland and L. Sandvik, "Performance of five two-sample location tests for skewed distributions with unequal variances," *Contemp. Clin. Trials*, vol. 30, no. 5, pp. 490–496, 2009, doi: 10.1016/j.cct.2009.06.007.
- [6] W. H. Kruskal and W. A. Wallis, "Use of ranks in one-criterion variance analysis," *J. Am. Stat. Assoc.*, vol. 47, no. 260, pp. 583–621, 1952, doi: 10.1080/01621459.1952.10483441.
- [7] M. D. Ernst, "Permutation methods: A basis for exact inference," *Stat. Sci.*, vol. 19, no. 4, pp. 676–685, 2004, doi: 10.1214/088342304000000396.
- [8] A. M. Mood, "On the asymptotic efficiency of certain nonparametric two-sample tests," *Ann. Math. Stat.*, vol. 25, no. 3, pp. 514–522, 1954, doi: 10.1214/aoms/1177728719.
- [9] J. N. Rouder, R. D. Morey, P. L. Speckman, and J. M. Province, "Default Bayes factors for ANOVA designs," *J. Math. Psychol.*, vol. 56, no. 5, pp. 356–374, 2012, doi: 10.1016/j.jmp.2012.08.001.
- [10] J. van Doorn *et al.*, "The JASP guidelines for conducting and reporting a Bayesian analysis," *Psychon. Bull. & Rev.*, vol. 28, no. 3, pp. 813–826, 2021, doi: 10.3758/s13423-020-01798-5.
- [11] E.-J. Wagenmakers *et al.*, "Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications," *Psychon. Bull. & Rev.*, vol. 25, no. 1, pp. 35–57, 2018, doi: 10.3758/s13423-017-1343-3.
- [12] M. I. Pongge, E. M. Lawalu, M. Masri, and M. P. J. Gamatara, "Foreign investment, human development, and regional inequality in Indonesia," *J. Econ. Res. Soc. Sci.*, vol. 9, no. 2, 2025, doi: <https://doi.org/10.18196/jerss.v9i2.27522>.
- [13] CELIOS, "Indonesia Inequality Report 2024." Center for Equality and Livelihood Studies, 2024. [Online]. Available: https://celios.co.id/wp-content/uploads/2024/09/Indonesia_Inequality_Report_2024_EN.pdf
- [14] V. den Bergh *et al.*, "A tutorial on conducting and interpreting a Bayesian ANOVA in JASP," *Annee. Psychol.*, vol. 121, no. 1, pp. 73–96, 2021, doi: 10.3917/anpsy1.201.0073.

- [15] E. Barzizza, N. Biasetton, and R. Ceccato, "Multi-aspect permutation tests for model selection," *Expert Syst. Appl.*, vol. 238, p. 121764, 2023, doi: 10.1111/exsy.13492.
- [16] D. W. Zimmerman, "A comparative study of the efficiency of the Wilcoxon test, Student's t test, and the alternatives to Student's t test," *Percept. Mot. Skills*, vol. 114, no. 3, pp. 700–716, 2012.
- [17] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, no. 3--4, pp. 591–611, 1965, doi: 10.1093/biomet/52.3-4.591.
- [18] N. M. Razali and Y. B. Wah, "Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests," *J. Stat. Model. Anal.*, 2011.
- [19] D. W. Zimmerman, "A note on preliminary tests of equality of variances," *Br. J. Math. Stat. Psychol.*, vol. 57, no. 1, pp. 173–181, 2004, doi: 10.1348/000711004849222.
- [20] L. M. Lix, H. J. Keselman, and J. C. Keselman, "Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance F test," *Rev. Educ. Res.*, vol. 66, no. 4, pp. 579–619, 1996, doi: 10.3102/00346543066004579.
- [21] R. A. Fisher, *Statistical methods for research workers*. Oliver and Boyd, 1925.
- [22] B. . Welch, "The Generalization of 'Student's' Problem when Several Different Population Variances are Involved," *Biometrika*, vol. 34, no. 1/2, pp. 28–35, 1947.
- [23] P. Good, *Permutation tests: A practical guide to resampling methods for testing hypotheses*, 2nd ed., vol. 2. New York: Springer, 2000.
- [24] K. J. Berry, J. E. Johnston, and P. W. Mielke, "Permutation methods," *Wiley Interdiscip. Rev. Comput. Stat.*, vol. 3, no. 6, pp. 527–542, 2011, doi: <https://doi.org/10.1002/wics.177>.
- [25] S. Siegel, *Statistik Nonparametrik untuk Ilmu-ilmu Sosial*. Jakarta: PT Gramedia Pustaka, 1997.
- [26] J. W. Tukey, "Comparing individual means in the analysis of variance.," *Biometrics*, vol. 5, no. 2, pp. 99–114, 1949, doi: <https://doi.org/10.2307/3001913>.
- [27] O. J. Dunn, "Multiple comparisons using rank sums," *Technometrics*, vol. 6, no. 3, pp. 241–252, 1964, doi: <https://doi.org/10.1080/00401706.1964.10490181>.
- [28] P. H. Westfall, W. O. Johnson, and J. M. Utts, "A Bayesian perspective on the Bonferroni adjustment," *Biometrika*, vol. 84, no. 2, pp. 419–427, 1997, doi: <https://doi.org/10.1093/biomet/84.2.419>.
- [29] J. Cohen, *Statistical power analysis for the behavioral sciences*, 2nd ed. Lawrence Erlbaum Associates., 1988.
- [30] M. Tomczak and E. Tomczak, "The need to report effect size estimates revisited. An overview of some recommended measures of effect size," *Trends Sport Sci.*, vol. 1, no. 21, pp. 19–25, 2014.