ESTIMASI PARAMETER REGRESI PROBIT ORDINAL DENGAN METODE *BAYESIAN* MENGGUNAKAN ALGORITMA *GIBBS SAMPLING* PADA INDEKS PEMBANGUNAN MANUSIA DI JAWA TENGAH

SKRIPSI

Diajukan untuk Memenuhi Sebagian Syarat Guna Memperoleh
Gelar Sarjana Matematika
dalam Ilmu Matematika

Oleh : **NELIS SA'ADAH**
NIM : 2108046021

PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI WALISONGO
SEMARANG
2025

PERNYATAAN KEASLIAN

Yang bertanda tangan di bawah ini :

Nama	:	Nelis Sa'adah
NIM	:	2108046021
Jurusan/Program Studi	:	Matematika/ Matematika

menyatakan bahwa skripsi yang berjudul :

ESTIMASI PARAMETER REGRESI PROBIT ORDINAL DENGAN METODE *BAYESIAN* MENGGUNAKAN ALGORITMA *GIBBS SAMPLING* PADA INDEKS PEMBANGUNAN MANUSIA DI JAWA TENGAH

secara keseluruhan adalah hasil penelitian/karya saya sendiri, kecuali bagian tertentu yang dirujuk sumbernya.

Semarang, 20 Juni 2025
Pembuat pernyataan,

Nelis Sa'adah
NIM : 2108046021

KEMENTERIAN AGAMA R.I.
UNIVERSITAS ISLAM NEGERI WALISONGO
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Prof. Dr. Hamka (Kampus II) Ngaliyan Semarang
Telp. 024-7601295 Fax. 7615387

PENGESAHAN

Naskah skripsi berikut ini :

Judul : **ESTIMASI PARAMETER REGRESI PROBIT ORDINAL DENGAN METODE *BAYESIAN* MENGGUNAKAN ALGORITMA *GIBBS SAMPLING* PADA INDEKS PEMBANGUNAN MANUSIA DI JAWA TENGAH**

Penulis : Nelis Sa'adah

NIM : 2108046021

Jurusan : Matematika

Telah diujikan dalam sidang *tugas akhir* oleh Dewan Penguji Fakultas Sains dan Teknologi UIN Walisongo dan dapat diterima sebagai salah satu syarat memperoleh gelar sarjana dalam Ilmu Matematika.

Semarang, 26 Juni 2025

DEWAN PENGUJI

Penguji I,

Emy Siswanah, M.Sc.
NIP: 198702022011012014

Penguji II,

Eva Khoirun Nisa, M.Si.
NIP: 198701022019032010

Penguji III,

Agus Wayan Yulianto, M.Sc.
NIP: 198907162019031007

Penguji IV,

Zulaikha, M.Si.
NIP: 199204092019032027

Pembimbing,

Eva Khoirun Nisa, M.Si.
NIP: 198701022019032010

NOTA DINAS

Semarang, 17 Juni 2025

Yth. Ketua Program Studi Matematika
Fakultas Sains dan Teknologi
UIN Walisongo Semarang

Assalamu'alaikum warahmatullahi wabarakatuh

Dengan ini diberitahukan bahwa saya telah melakukan bimbingan, arahan dan koreksi naskah skripsi dengan:

Judul	:	ESTIMASI PARAMETER REGRESI PROBIT ORDINAL DENGAN METODE *BAYESIAN* MENGGUNAKAN ALGORITMA *GIBBS SAMPLING* PADA INDEKS PEMBANGUNAN MANUSIA DI JAWA TENGAH
Nama	:	Nelis Sa'adah
NIM	:	2108046021
Jurusan	:	Matematika

Saya memandang bahwa naskah skripsi tersebut sudah dapat diajukan kepada Fakultas Sains dan Teknologi UIN Walisongo untuk diujikan dalam Sidang Munaqasyah.

Wassalamu'alaikum warahmatullahi wabarakatuh

Pembimbing,

Eva Khoirun Nisa, M.Si.
NIP : 198701022019032010

ABSTRAK

Penelitian ini bertujuan untuk menganalisis estimasi parameter regresi probit ordinal dengan pendekatan *Bayesian* menggunakan algoritma *Gibbs Sampling* pada Indeks Pembangunan Manusia (IPM) di Provinsi Jawa Tengah pada tahun 2024. Variabel prediktor yang digunakan dalam penelitian ini yaitu Angka Harapan Hidup (AHH), Harapan Lama Sekolah (HLS), Rata-rata Lama Sekolah (RLS), Tingkat Pengangguran Terbuka (TPT) dan Tingkat Partisipasi Angkatan Kerja (TPAK), dengan IPM sebagai variabel respon yang dikategorikan ke dalam tiga kategori. Estimasi parameter dilakukan melalui algoritma *Gibbs Sampling* dalam kerangka *Markov Chain Monte Carlo* (MCMC). Hasil estimasi parameter menunjukkan bahwa HLS dan RLS berpengaruh terhadap kategori IPM. Evaluasi konvergensi menggunakan *trace plot* menunjukkan parameter yang signifikan sudah mencapai distribusi stasioner. Uji *Pearson Chi-Square* menghasilkan nilai 0,60 yang mengindikasikan model sesuai dengan data. Model ini memiliki nilai koefisien determinasi *McFadden* R^2 sebesar 0,8620, yang menunjukkan kemampuan model dalam menjelaskan perubahan dari variabel HLS dan RLS terhadap variabel IPM sangat besar. Dengan demikian, algoritma *Gibbs Sampling* mampu menghasilkan estimasi parameter yang stabil dan konsisten. Penelitian ini menyimpulkan bahwa peningkatan kualitas pendidikan, khususnya HLS dan RLS, berkontribusi signifikan dalam meningkatkan IPM di Provinsi Jawa Tengah.

Kata kunci: Regresi Probit Ordinal, *Bayesian*, algoritma *Gibbs Sampling*, Indeks Pembangunan Manusia (IPM), *Markov Chain Monte Carlo* (MCMC).

KATA PENGANTAR

Puji syukur kami panjatkan ke hadirat Allah SWT yang telah melimpahkan rahmat dan karunia-Nya sehingga kami dapat menyelesaikan penulisan Skripsi ini dengan judul "Estimasi Parameter Regresi Probit Ordinal dengan Metode *Bayesian* Menggunakan Algoritma *Gibbs Sampling* pada Indeks Pembangunan Manusia di Jawa Tengah". Sholawat serta salam senantiasa tercurahkan kepada junjungan Nabi Besar Muhammad SAW, suri teladan seluruh umat manusia, yang ajarannya senantiasa menjadi cahaya penerang dalam kehidupan dan syafa'atnya selalu dinantikan di hari akhir kelak.

Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Matematika (S.Mat.) pada Program Studi Matematika Fakultas Sains dan Teknologi, Universitas Islam Negeri Walisongo Semarang. Proses penyusunan Skripsi ini tidak lepas dari doa, bantuan, bimbingan, motivasi dan peran dari banyak pihak. Sehingga penulis mengucapkan terima kasih kepada:

1. Allah SWT yang dengan kasih sayang-Nya telah memberikan kesehatan, kesabaran, serta kelancaran dalam setiap proses penulis hingga Skripsi ini dapat terselesaikan;
2. Bapak Prof. Nizar, M.Ag., selaku Rektor UIN Walisongo Semarang;
3. Ibu Any Muanalifah, M.Si, Ph.D., selaku Ketua Program Studi Matematika;
4. Bapak Prihadi Kurniawan, M.Sc., selaku Sekretaris Program Studi Matematika;
5. Ibu Eva Khoirun Nisa, M.Si., selaku Dosen Pembimbing yang telah meluangkan waktu, memberikan bimbingan, arahan dan masukan yang berharga dari awal hingga akhir penulisan Skripsi ini;

6. Cinta pertama dan panutanku, Bapak Slamet Rudiono. Beliau memang tidak sempat merasakan pendidikan sampai bangku perkuliahan, namun dengan keyakinan sepenuh hati beliau mampu mendidik penulis, memberikan motivasi di setiap keberhasilan dan kegagalan penulis serta memberikan dukungan baik moral maupun material hingga penulis mampu menyelesaikan Skripsi ini;

7. Pintu surgaku, Ibu Yanti. Beliau sangat berperan penting dalam menyelesaikan program studi penulis, beliau juga memang tidak sempat merasakan pendidikan sampai di bangku perkuliahan, tapi semangat, motivasi serta doa tulus dan cinta yang luar biasa selalu beliau berikan kepada putri satu-satunya;

8. Adik tersayang Muhammad Irfan Kurniawan, yang senantiasa menjadi sumber semangat dan inspirasi dalam setiap langkah penulis. Terima kasih atas canda tawa yang selalu menghibur penulis dikala sedih dan dukungan yang tidak ada habisnya;

9. Sahabat baik penulis, Della Setya Rahma dan Siti Hidayatun Nisyak. Terima kasih karena selalu hadir dalam setiap proses penyusunan Skripsi ini, memberikan motivasi disetiap langkah, serta menjadi tempat berbagi cerita dan keluh kesah penulis dalam menyelesaikan Skripsi ini;

10. Teman dekat penulis, Titisari Azizah Wijayanti, S.Mat., Tiara Fitrianingtyas, Ely Chabibah, Ayu Imtiyaz Sari, Aeni Nurjannah, Siti Nur Arafah. Terima kasih atas segala dukungan, kebersamaan, semangat, serta tawa yang telah menjadi penyemangat penulis selama perkuliahan sampai akhirnya menyelesaikan Skripsi ini;

11. Teman seperjalanan penulis selama Kuliah Kerja Nyata (KKN), Amirul Hidayah, Nur Hidayah, dan Siti Nur Chumairoh. Terima kasih karena kehadiran kalian telah

memberikan warna dan semangat tersendiri dalam setiap kegiatan, serta menjadi bagian dari kenangan berharga penulis yang tidak akan terlupakan;

12. Seluruh teman-teman seperjuangan Program Studi Matematika kelas A angkatan 2021 yang telah memberikan dukungan dan semangat satu sama lain serta meberikan pengalaman yang berharga yang akan penulis kenang;

13. Keponakan *online* Abe Cekut, Adik Ritsuki dan Mas Natsuki. Terima kasih telah menghibur penulis dengan video-video *random* dan lucunya dikala penulis mengalami kesedihan dan kesulitan dalam menyelesaikan Skripsi ini;

14. Semua pihak yang tidak dapat penulis sebutkan satu persatu yang telah memberikan kontribusi hingga selesainya Skripsi ini.

15. *And last but not least*, untuk diri saya sendiri, Nelis Sa'adah. Terima kasih sudah bertahan sejauh ini. Terima kasih tetap memilih berusaha dan merayakan dirimu sendiri sampai di titik ini. Meskipun belum sebersinar yang lain dan sering kali merasa putus asa atas kegagalan usahanya, namun terima kasih tetap menjadi manusia yang selalu mau berusaha dan tidak lelah mencoba. Terima kasih sudah berhasil di atas keraguan orang lain. Terima kasih karena memutuskan tidak menyerah sesulit apapun proses penyusunan Skripsi ini dan telah menyelesaikannya sebaik dan semaksimal mungkin. Ini merupakan pencapaian yang patut dirayakan untuk diri sendiri. Berbahagialah selalu dimanapun berada, Nelis. Apapun kurang dan lebihmu mari merayakan diri sendiri.

Semoga kebaikan semuanya menjadi amal ibadah yang diterima dan mendapat pahala yang berlimpah dari Allah SWT. Aamiin.

Penyusunan Skripsi ini bukanlah suatu hal yang mudah, namun dengan dukungan dari semua pihak yang telah disebutkan di atas, kami dapat menyelesaikannya. Semoga Skripsi ini dapat

memberikan manfaat bagi pengembangan ilmu pengetahuan dan menjadi sumbangan kecil bagi peningkatan kualitas hidup masyarakat, khususnya untuk mengetahui faktor apa saja yang memengaruhi Indeks Pembangunan Manusia di Provinsi Jawa Tengah. Atas segala kekurangan dan kelemahan dalam Skripsi ini penulis mengharapkan saran dan kritik yang membangun. Semoga karya tulis yang sederhana ini dapat menjadi bacaan yang bermanfaat dan dapat dikembangkan bagi peneliti-peneliti selanjutnya.

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN

Halaman

BAB I
PENDAHULUAN

A. Latar Belakang Masalah

Paradigma pembangunan yang dikenal sebagai pembangunan manusia menempatkan manusia atau masyarakat sebagai tujuan utama dan tujuan akhir dari semua kegiatan pembangunan. Sasaran kegiatan tersebut meliputi memperoleh kontrol atas sumber daya (pendapatan untuk mencapai hidup layak), meningkatkan derajat kesehatan (usia hidup panjang dan sehat) dan meningkatkan pendidikan (Sangkereng et al., 2019). Salah satu aspek penting pembangunan adalah bahwa manusia dianggap sebagai subjek pembangunan dan bahwa pembangunan dilakukan dengan tujuan untuk membantu manusia atau masyarakat. Keberhasilan dalam pembangunan manusia dapat dilihat dari kualitas serta kemampuan masyarakat dalam mengelola dan memanfaatkan sumber daya alam secara optimal (Tyas, 2015). Indeks Pembangunan Manusia (IPM) merupakan suatu indikator yang digunakan untuk mengukur pembangunan manusia yang menjadi salah satu fokus pembangunan dari pemerintah (Riani, 2006).

Indeks Pembangunan Manusia (IPM) digunakan untuk mengukur kinerja pemerintah dalam menerapkan kebijakan ekonomi yang didasarkan pada tiga indikator yaitu indeks kesehatan, indeks pendidikan dan standar hidup layak (Kuncoro, 2004). Indikator kesehatan diukur melalui Angka Harapan Hidup (AHH), indikator pendidikan diukur dengan Harapan Lama Sekolah (HLS) dan Rata-rata Lama Sekolah (RLS), sementara pengukuran standar hidup layak dilakukan melalui Pengeluaran Per Kapita (PPP) (Siswati, 2018). Dengan adanya Tingkat Pengangguran Terbuka (TPT) yang rendah dan Tingkat Partisipasi Angkatan Kerja (TPAK) yang tinggi, dapat meningkatkan Pengeluaran Per Kapita (PPP) sehingga mampu berkontribusi pada peningkatan standar hidup layak (UNDP, 2019). Di Provinsi Jawa

Tengah sebagaimana terjadi di banyak daerah lain, IPM merupakan indikator yang sangat penting untuk menilai efektivitas program pembangunan dan kesejahteraan masyarakat (Yektiningsih, 2018). Pada tahun 2024 IPM Provinsi Jawa Tengah mencapai 73,87, meningkat 0,48 poin atau tumbuh 0,65% dibandingkan capaian tahun 2023 sebesar 70,39 (BPS, 2024). Meskipun demikian, IPM Provinsi Jawa Tengah masih menjadi yang terendah diantara enam provinsi yang ada di Pulau Jawa dan masih dibawah IPM Indonesia pada tahun 2024 yaitu sebesar 75,02 (BPS, 2024). Ini menunjukkan gambaran mengenai tantangan yang dihadapi daerah tersebut dalam mencapai pembangunan manusia yang berkualitas.

Tantangan yang dimaksud dalam mencapai pembangunan manusia yang berkualitas di Provinsi Jawa Tengah dapat dilihat dari data Federasi Serikat Guru Indonesia (FSGI) dari Januari - Juli 2024 mencatat ada sebanyak 5 kasus kekerasan berat yang dialami oleh anak sekolah. Beberapa jenis kekerasan yang terjadi mulai dari kekerasan fisik hingga kekerasan seksual. Oleh sebab itu, sepanjang tahun 2024 masih banyak anak putus sekolah yang memerlukan penanganan khusus dari pemerintah. Salah satunya berdasarkan data Balai Besar Penjaminan Mutu Pendidikan (BBPMP) Provinsi Jawa Tengah tercatat sebanyak 303 anak di Surakarta diketahui sebagai Anak Tidak Sekolah (ATS) dan Anak Putus Sekolah (APS). Selain alasan ekonomi, perundungan adalah faktor lain yang mendorong anak-anak untuk tidak melanjutkan sekolah. Berdasarkan kondisi tersebut, diperlukan suatu penelitian untuk mengetahui apa saja faktor yang memengaruhi Indeks Pembangunan Manusia (IPM) di Provinsi Jawa Tengah pada tahun 2024.

Indeks Pembangunan Manusia (IPM) terbagi menjadi empat kategori menurut Badan Pusat Statistik yaitu rendah, sedang, tinggi dan sangat tinggi (Pamungkas dan Widiyanto, 2022). Keempat kategori tersebut akan menjadi variabel respon dalam penelitian ini. Maka dari itu, diperlukan suatu metode statistik yang dapat menganalisis faktor yang memengaruhi IPM dengan

variabel respon berupa kategorik, salah satunya menggunakan regresi probit ordinal. Analisis regresi yang disebut regresi probit ini digunakan untuk menunjukkan bagaimana keterkaitan antara variabel respon dengan skala pengukuran dikotomus (biner) dan variabel prediktor yang skala pengukurannya bersifat dikotomus, polikotomus atau kontinu (Tinungki, 2010). Regresi probit dapat diklasifikasikan menjadi tiga yaitu regresi probit biner, regresi probit multinomial dan regresi probit ordinal (Septadianti, 2016). Berbeda dengan regresi probit biner, yang hanya dapat digunakan untuk variabel dengan dua hasil (misalnya, "iya" atau "tidak"), regresi probit ordinal dapat menangani variabel respon yang kategorinya lebih dari dua (Greene, 2018). Selain itu, regresi probit multinomial meskipun mampu menangani data kategorikal dengan lebih dari dua kategori, tidak mempertimbangkan adanya urutan di antara kategori-kategori tersebut, yang justru menjadi fokus utama dalam regresi probit ordinal (Agresti, 2018). Oleh karena itu, regresi probit ordinal menjadi pilihan yang lebih tepat dan fleksibel untuk analisis data kategorikal yang melibatkan banyak kategori dengan urutan atau peringkat yang jelas.

Terdapat beberapa metode pendugaan titik yang dapat digunakan untuk mengestimasi parameter regresi probit ordinal, yaitu metode momen, *Maximum Likelihood Estimation* (MLE), dan metode *Bayesian*. Merujuk pada penelitian yang dilakukan oleh Fikhri et al. (2014), nilai *Mean Square Error* (MSE) yang dihasilkan oleh metode MLE lebih besar jika dibandingkan dengan metode *Bayesian*. Dengan demikian, metode *Bayesian* menghasilkan estimasi parameter μ yang lebih konsisten pada distribusi *Poisson* dibandingkan metode *MLE*. Selain itu, (Nurlaila et al., 2013) melakukan pengujian dengan membandingkan metode MLE dan metode *Bayes* dalam pendugaan parameter distribusi eksponensial. Hasilnya diperoleh metode *Bayes* dengan perluasan distribusi *Prior Jeffrey* lebih efektif karena nilai bias dan MSE yang lebih kecil dibandingkan dengan metode MLE. Pengetahuan subjektif (*prior*) tentang distribusi peluang dari parameter yang tidak diketahui digabungkan dengan data sampel

dalam pendekatan *Bayes*. Distribusi posterior dihasilkan dengan menggabungkan *prior* dan *likelihood* menggunakan teorema *Bayes* (Diana, 2016). Distribusi *posterior* diperoleh dengan melibatkan suatu proses pengintegralan yang kompleks, sehingga proses tersebut didekati dengan proses pengintegralan *Monte Carlo* dengan sifat *Markov Chain*.

Markov Chain Monte Carlo (MCMC) memungkinkan untuk memperoleh sampel dari distribusi *posterior* yang sangat kompleks dan multidimensi. Terdapat beberapa algoritma yang digunakan dalam metode MCMC seperti *Metropolish Hasting* dan *Gibbs Sampling* (Anifa et al., 2012). Banyak distribusi kondisional penuh yang digunakan dalam *Gibbs Sampling* merupakan distribusi yang umum dan telah diketahui teknik pengambilan sampelnya, seperti distribusi normal, beta, dan gamma, sehingga memudahkan proses implementasi serta membuat pengambilan sampel lebih mudah dan efisien dibandingkan algoritma *Metropolis-Hastings* yang memerlukan spesifikasi *proposal distribution* (Fox, 2012). Menurut Robert (2004), *Gibbs Sampling* sering kali lebih stabil dalam hal konvergensi dibandingkan dengan *Metropolish Hasting* karena tidak memerlukan pemilihan parameter langkah yang tepat. *Gibbs Sampling* juga telah terbukti kesuksesannya dalam meringkas *high-dimensional posterior distribution* yang sulit dilakukan oleh metode lainnya (Albert, 1992).

Berdasarkan hasil penelitian dari Ginting (2023) dengan mengidentifikasi pengaruh Angka Harapan Hidup (AHH) dan Harapan Lama sekolah (HLS) terhadap Indeks Pembangunan Manusia (IPM) di Kabupaten Langkat dari tahun 2013 hingga 2022, hasilnya menunjukkan bahwa variabel AHH dan HLS berpengaruh terhadap IPM secara simultan (bersama-sama). Sementara itu, Syahid (2022) melakukan analisis faktor-faktor yang memengaruhi IPM di Provinsi Papua dan diperoleh kesimpulan bahwa secara simultan Rata-rata Lama Sekolah (RLS), rata-rata Pengeluaran Per Kapita (PPP), dan Tingkat Pengangguran Terbuka (TPT) berpengaruh terhadap IPM di setiap kabupaten/kota di

Provinsi Papua dari tahun 2018 hingga 2020. Selain itu, dalam penelitian (Faelassuffa, 2021), dengan adanya produktivitas yang baik akan meningkatkan nilai jual tenaga kerja, sehingga Tingkat Partisipasi Angkatan Kerja (TPAK) memiliki pengaruh terhadap IPM. Oleh karena dari penelitian sebelumnya diperoleh faktor-faktor yang diduga memengaruhi Indeks Pembangunan Manusia, sehingga faktor-faktor tersebut akan digunakan sebagai variabel prediktor pada penelitian ini.

Berdasarkan latar belakang di atas, peneliti ingin mengkaji mengenai faktor yang memengaruhi IPM menggunakan regresi probit ordinal dengan metode estimasi parameter *Bayesian* karena dalam penelitian sebelumnya metode *Bayesian* lebih konsisten dibandingkan dengan metode MLE dan juga menggunakan algoritma *Gibbs Sampling* karena kemudahan implementasinya dibandingkan dengan *Metropolish Hasting*, terutama ketika distribusi kondisional bersyarat dari parameter dapat dihitung secara langsung. Dengan diketahuinya faktor pengaruh IPM, maka harapannya dapat menjadi evaluasi pemerintah Provinsi Jawa Tengah dalam meningkatkan IPM sehingga tidak lagi menjadi provinsi di Pulau Jawa dengan IPM terendah (BPS, 2024).

B. Rumusan Masalah

Rumusan masalah yang akan menjadi fokus dalam penelitian ini sebagai berikut:

1. Bagaimana hasil estimasi parameter regresi probit ordinal dengan metode *Bayesian* menggunakan algoritma *Gibbs Sampling*?
2. Apa saja faktor yang memengaruhi Indeks Pembangunan Manusia Provinsi Jawa Tengah tahun 2024 dengan regresi probit ordinal metode *Bayesian* menggunakan algoritma *Gibbs Sampling*?

C. Tujuan Penelitian

Berdasarkan permasalahan yang telah dirumuskan, maka tujuan penelitian ini sebagai berikut:

1. Untuk mengetahui hasil estimasi parameter regresi probit ordinal dengan metode *Bayesian* menggunakan algoritma *Gibbs Sampling*.
2. Untuk mengetahui faktor apa saja yang memengaruhi Indeks Pembangunan Manusia Provinsi Jawa Tengah tahun 2024 dengan regresi probit ordinal metode *Bayesian* menggunakan algoritma *Gibbs Sampling*.

D. Manfaat Penelitian

Manfaat yang ingin dicapai melalui penelitian ini adalah sebagai berikut:

1. Kontribusi terhadap perkembangan ilmu pengetahuan, khususnya dalam bidang statistika *Bayesian* dan analisis data multidimensi, dengan mengaplikasikan metode *Bayesian* serta algoritma *Gibbs Sampling* pada estimasi parameter regresi probit ordinal.
2. Peningkatan pemahaman mengenai hubungan antara Indeks Pembangunan Manusia dengan faktor-faktor yang memengaruhinya di Provinsi Jawa Tengah pada tahun 2024 melalui analisis regresi probit ordinal.
3. Hasil penelitian dapat menghasilkan informasi yang berharga bagi pemerintah Provinsi Jawa Tengah maupun pihak terkait untuk lebih memahami dinamika pembangunan manusia serta faktor-faktor yang perlu diperhatikan dalam perencanaan pembangunan daerah.
4. Implikasi kebijakan yang dapat dihasilkan dari pemahaman lebih mendalam terhadap faktor-faktor yang memengaruhi

Indeks Pembangunan Manusia di Provinsi Jawa Tengah, sehingga dapat meningkatkan efektivitas program pembangunan yang ada dan mengarahkan kebijakan yang lebih tepat sasaran untuk meningkatkan kualitas hidup masyarakat setempat.

E. Batasan Masalah

Adapun batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Penelitian ini hanya memusatkan pada estimasi parameter regresi probit ordinal untuk Indeks Pembangunan Manusia Provinsi Jawa Tengah tahun 2024.

2. Penggunaan metode *Bayesian* dalam penelitian ini dibatasi pada algoritma *Gibbs Sampling* sebagai teknik utama dalam mengestimasi parameter regresi.

3. Data yang dianaliss dalam penelitian ini terbatas pada data sekunder yang tersedia di Provinsi Jawa Tengah pada tahun 2024, yang meliputi variabel-variabel yang relevan dengan Indeks Pembangunan Manusia seperti Angka Harapan Hidup, Harapan Lama Sekolah, Rata-rata Lama Sekolah, Tingkat Pengangguran Terbuka dan Tingkat Partisipasi Angkatan Kerja.

4. Penelitian ini tidak mempertimbangkan faktor-faktor eksternal yang mungkin memengaruhi Indeks Pembangunan Manusia di Provinsi Jawa Tengah, seperti faktor geografis, politik, atau ekonomi yang lebih luas.

BAB II
LANDASAN PUSTAKA

A. Kajian Teori

1. Multikolinieritas

Multikolinieritas adalah keadaan di dalam regresi yang hubungan antara variabel prediktornya sangat kuat (Sujarweni, 2015). Uji multikolinieritas bertujuan untuk menguji apakah dalam model regresi terdapat korelasi yang tinggi antar variabel prediktor (Ghozali, 2017). Pada penelitian ini, variabel prediktor yang diuji meliputi Angka Harapan Hidup (AHH), Harapan Lama Sekolah (HLS), Rata-rata Lama Sekolah (RLS), Tingkat Pengangguran Terbuka (TPT), dan Tingkat Partisipasi Angkatan Kerja (TPAK). Adanya hubungan korelatif antar variabel prediktor dapat memengaruhi keakuratan estimasi parameter regresi, sehingga menghasilkan error yang cukup besar. Maka dari itu, multikolinieritas tidak diperbolehkan dalam analisis regresi. Identifikasi adanya multikolinieritas dapat dilakukan dengan menghitung nilai *Variance Inflation Factor* (VIF). Nilai VIF untuk variabel x_j dapat dihitung sebagai berikut (Gujarati, 2004):

$$VIF_j = \frac{1}{1 - R^2_{McF}} \tag{2.1}$$

untuk $j = 1, 2, ..., p$

dimana R^2_{McF} adalah nilai koefisien determinasi *McFadden* dari model regresi yang perhitungannya sesuai dengan Persamaan (2.66) yang mengasumsikan salah satu variabel prediktor x_j sebagai respon, dengan sisanya berperan sebagai prediktor. Multikolinieritas terjadi ketika nilai VIFnya lebih dari 10 (Hocking, 1996). Pembentukan variabel interaksi bisa menjadi salah satu cara untuk menangani multikolinieritas dalam model regresi, terutama ketika dua variabel prediktor sangat berkorelasi, namun

memiliki hubungan non-linier atau efek gabungan yang signifikan terhadap variabel prediktor.

2. Regresi Probit

Bliss (1934) yang pertama kali memperkenalkan regresi probit, menyatakan bahwa probit berasal dari istilah statistik *probability unit* yang artinya model regresi probit berhubungan dengan unit-unit probabilitas (Agresti, 2007). Regresi probit merupakan metode regresi yang digunakan untuk mengkaji variabel respon yang bersifat kategorik dan variabel prediktor yang bersifat kategorik, numerik, atau keduanya (Gujarati, 2004).

Model probit dikenal juga sebagai model normit karena menggunakan pendekatan Fungsi Distribusi Kumulatif atau *Cumulative Distribution Function* (CDF) dari distribusi normal (Haslinda, 2020). Pendekatan melalui CDF diterapkan sebagai solusi terhadap kelemahan yang terdapat pada model probabilitas linier. Kelemahan tersebut terletak pada potensi nilai y yang melebihi batas jangkauan variabel respon. Model probit mengasumsikan bahwa peluang terjadinya suatu keberhasilan ditentukan oleh variabel laten yang tidak dapat diamati secara langsung, dengan batas nilai kritis tertentu yang menentukan kategori hasil. Pemodelan regresi probit dapat dilihat dengan mengawali model (Greene, 2012) sebagai berikut :

$$y^* = x^\top \beta + \varepsilon \tag{2.2}$$

dimana:

- y^* : variabel laten kontinu yang tidak teramati
- β : vektor parameter koefisien dengan $\beta = \begin{bmatrix}\beta_0 & \beta_1 & \beta_2 & \dots & \beta_p\end{bmatrix}^\top$
- x : vektor variabel prediktor dengan $x = \begin{bmatrix}1 & x_{1i} & x_{2i} & \dots & x_{pi}\end{bmatrix}^\top$

- ε : error yang diasumsikan berdistribusi normal dengan *mean* 0 dan standar deviasi 1 ($\mathcal{N}(0,1)$)

Meskipun y^* tidak teramati secara langsung, pengkategorian y^* dapat dilakukan dengan menetapkan *threshold* (ambang batas) tertentu, misalnya γ. Oleh karena itu, variabel y^* dapat dikategorikan sebagaimana dijelaskan oleh (Greene, 2012) sebagai berikut:

$$y = \begin{cases} 0, & \text{jika } y^* \leq \gamma \\ 1, & \text{jika } y^* > \gamma \end{cases} \tag{2.3}$$

Probabilitas untuk $y = 0$ menunjukkan kemungkinan bahwa suatu kejadian dikategorikan sebagai "gagal".

$$\begin{aligned} P(y=0) = P(y^* \leq \gamma) &= P(x^\top \beta + \varepsilon \leq \gamma) \\ &= P(\varepsilon \leq \gamma - x^\top \beta) \\ &= \Phi(\gamma - x^\top \beta) \end{aligned} \tag{2.4}$$

Probabilitas untuk y = 1 menunjukkan kemungkinan bahwa suatu kejadian dikategorikan sebagai "sukses".

$$\begin{aligned} P(y=1) = P(y^* > \gamma) &= 1 - P(y^* \leq \gamma) \\ &= 1 - P(x^\top \beta + \varepsilon \leq \gamma) \\ &= 1 - P(\varepsilon \leq \gamma - x^\top \beta) \\ &= 1 - \Phi(\gamma - x^\top \beta) \end{aligned} \tag{2.5}$$

dengan $\Phi(z) = \Phi(\gamma - x^\top \beta)$ dan $z = \gamma - x^\top \beta$ adalah CDF dari distribusi normal standar. CDF ini merepresentasikan peluang kumulatif, yaitu probabilitas bahwa suatu variabel *random* yang mengikuti distribusi normal standar ($Z \sim \mathcal{N}(0,1)$) memiliki nilai kurang dari atau sama dengan z.

Secara matematis, CDF normal standar dituliskan sebagai berikut:

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt \tag{2.6}$$

Artinya, $\Phi(z)$ memberikan nilai peluang bahwa $Z \leq z$, atau:

$$\Phi(z) = P(Z \leq z)$$

Dalam model regresi probit, nilai $z = \gamma - x^\top \beta$ merepresentasikan *threshold* dari variabel laten y^*, yang tidak teramati, namun diasumsikan mengikuti distribusi normal standar. Oleh karena itu, fungsi $\Phi(\gamma - x^\top \beta)$ digunakan untuk menghitung probabilitas bahwa respons y jatuh ke dalam kategori tertentu. Dengan demikian, fungsi CDF berperan penting dalam menghubungkan prediktor x dengan probabilitas kejadian pada variabel respon melalui distribusi normal standar.

3. Regresi Probit Ordinal

Model regresi probit ordinal berasal dari pengembangan model regresi probit biner untuk menangani kasus dimana kategori dari variabel respon lebih dari dua, misalnya tiga kategori yaitu rendah, sedang, tinggi, dan sangt tinggi. Regresi probit ordinal merupakan metode analisis untuk mengidentifikasi hubungan antara variabel respon berskala ordinal dengan satu atau lebih variabel prediktor yang dapat berupa dikotomus, polikotomus, atau kontinu (Wardani, 2020).

Prinsip dasar pengukuran probabilitas dalam regresi probit ordinal tetap menggunakan pendekatan biner dengan membandingkan setiap kategori dengan kategori referensi, meskipun regresi ini memiliki variabel respon dengan lebih dari dua kategori. Kategori referensi dalam model regresi probit ordinal digunakan untuk menghindari kesulitan mengidentifikasi model dan juga memberikan interpretasi yang jelas tentang koefisien untuk kategori lainnya. Oleh karena itu, meskipun

menggunakan variabel respon multikategori, model probit ordinal tetap mempertahankan struktur dasar dari regresi probit biner (Long, 2006).

Pada regresi probit ordinal, Variabel y dalam model (2.2) adalah variabel respon teramati dengan memberikan *threshold* yaitu $\gamma_1 < \gamma_2 < \ldots < \gamma_{J-1}$ pada y^* dengan j = (1,2,...,J). Jadi variabel respon y akan membentuk j kategori dengan penjelasan sebagai berikut (Greene, 2008):

$$y = \begin{cases} 1, & \text{jika } y^* \leq \gamma_1 \\ 2, & \text{jika } \gamma_1 < y^* \leq \gamma_2 \\ 3, & \text{jika } \gamma_2 < y^* \leq \gamma_3 \\ \vdots & \\ J, & \text{jika } y^* > \gamma_{J-1} \end{cases} \tag{2.7}$$

sehingga diperoleh model sebagai berikut:

$$P(y = 1) = \Phi(\gamma_1 - x^\top \beta) \tag{2.8}$$

$$P(y = 2) = \Phi(\gamma_2 - x^\top \beta) - \Phi(\gamma_1 - x^\top \beta) \tag{2.9}$$

$$P(y = 3) = \Phi(\gamma_3 - x^\top \beta) - \Phi(\gamma_2 - x^\top \beta) \tag{2.10}$$

$$\vdots$$

$$P(y = J) = \Phi(\gamma_J - x^\top \beta) - \Phi(\gamma_{J-1} - x^\top \beta) \tag{2.11}$$

$$P(y = J) = 1 - \Phi(\gamma_{J-1} - x^\top \beta) \tag{2.12}$$

dimana $y = 1$ untuk kategori terendah dan $y = J$ untuk kategori tertinggi.

4. Metode Pendugaan *Bayesian*

Pendugaan merupakan proses statistik yang digunakan untuk menghasilkan sebuah hasil dugaan *(estimate)* dari suatu parameter. Penduga statistik seperti rata-rata, persentase, variansi, dan lainnya digunakan untuk mengestimasi sebuah parameter (Harinaldi, 2012). Metode *Bayesian* menghasilkan distribusi *posterior* dengan menggabungkan fungsi *likelihood* dengan distribusi *prior* (Berger, 1985).

Dalam metode pendugaan *Bayesian*, data *prior* harus diketahui terlebih dahulu distribusinya sebelum merumuskan distribusi *posterior*nya (Irwanti et al., 2012). Selanjutnya distribusi *prior* dan *posterior* dapat membantu menyelesaikan bagian yang lebih kompleks dari sebuah solusi. Metode *Bayesian* adalah suatu pendekatan untuk estimasi parameter atau pengujian hipotesis yang didasarkan pada Teorema *Bayes*. Metode *Bayesian* menggunakan Teorema *Bayes* untuk memperbarui keyakinan tentang parameter θ berdasarkan data yang diamati y. Teorema *Bayes* dinyatakan sebagai berikut (Kusrini, 2009):

$$P(\theta \mid y) = \frac{P(y \mid \theta)P(\theta)}{P(y)} \tag{2.13}$$

dimana:

- $P(\theta \mid y)$: probabilitas *posterior* dari parameter θ setelah mengamati data y
- $P(y \mid \theta)$: fungsi *likelihood* dari data y jika parameter θ diketahui
- $P(\theta)$: probabilitas *prior* dari parameter θ
- $P(y)$: probabilitas *evidence* atau *marginal likelihood* dari data y

Dalam konteks regresi ordinal *Bayesian*, parameter model θ terdiri dari koefisien regresi β dan *threshold* γ, sehingga $\theta = (\beta, \gamma) = (\beta_1, \beta_2, \ldots, \beta_p; \gamma_1, \gamma_2, \ldots, \gamma_{J-1})$. Maka, bentuk Teorema *Bayes* menurut Gelman et al. (2013) menjadi:

$$P(\beta, \gamma \mid y) = \frac{P(y \mid \beta, \gamma)P(\beta, \gamma)}{P(y)} \tag{2.14}$$

Setelah data diamati, Teorema *Bayes* digunakan untuk memperbarui distribusi *prior* dari parameter (β, γ) menjadi distribusi *posterior* $P(\beta, \gamma \mid y)$. Dengan demikian, proses pendugaan *Bayesian* terdiri dari menggabungkan informasi sebelumnya (*prior*) dengan bukti empiris dari data melalui fungsi *likelihood*, sehingga menghasilkan distribusi *posterior*. Proses pendugaan *Bayesian* sebagai berikut:

1. *Prior Distribution* (Distribusi *Prior*)

   Distribusi *Prior* adalah distribusi probabilitas yang menggambarkan keyakinan awal terhadap suatu parameter sebelum data diamati dan digunakan sebagai titik awal dalam pendekatan inferensi *Bayesian* (Ainul et al., 2018). Sebelum merumuskan distribusi *posterior*, ditentukan distribusi *prior*nya terlebih dahulu. Dalam pendekatan *Bayesian*, terdapat berbagai jenis *prior*, antara lain *prior* konjugat dan *prior* non-konjugat, *prior* informatif atau non-informatif dan *pseudo prior* (Syafitri et al., 2021).

   Menurut Box dan Tiao (1973) *prior* konjugat mengacu pada acuan analisis model terutama dalam pembentukan fungsi *likelihood*nya, sehingga dalam penentuan *prior* konjugat selalu dipikirkan mengenai penentuan pola dari distribusi *prior* yang mempunyai bentuk konjugat dengan fungsi densitas peluang pembangun *likelihood*nya. Sedangkan *prior* non-konjugat, yaitu apabila pemberian *prior* pada suatu model tidak mengindahkan pada pola pembentuk fungsi *likelihood*nya. Penetapan *prior* informatif

maupun non-informatif bergantung pada ada tidaknya informasi sebelumnya yang relevan dengan karakteristik distribusi data (Lancaster, 2004). *Prior* informatif mengacu pada pemberian parameter dari distribusi *prior* yang telah dipilih baik *prior* konjugat atau tidak. Sementara itu, *prior* non-informatif merupakan distribusi yang dipilih tanpa mengacu pada data sebelumnya dan tidak memberikan informasi khusus mengenai parameter θ. *Pseudo prior* adalah *prior* dalam model *Bayesian* yang nilainya ditentukan dengan meniru atau menyamakan hasil dari metode *frequentist* seperti estimasi *Maximum Likelihood Estimation* (MLE) dan variansinya (Carlin dan Chib, 1995).

Prior untuk koefisien regresi β menggunakan distribusi normal (Gelman et al., 2013)

$$\beta \sim \mathcal{N}(0, \sigma^2) \tag{2.15}$$

dengan *mean* 0 dan varians σ^2.

Prior untuk *threshold* γ umumnya juga mengikuti distribusi normal (Gelman et al., 2013)

$$\gamma \sim \mathcal{N}(0, \sigma^2) \tag{2.16}$$

dengan *threshold* $\gamma_1 < \gamma_2 < \ldots < \gamma_{J-1}$ dimana j = (1,2,...,J).

2. *Likelihood*

Menurut Bain dan Engelhardt (1992), fungsi *likelihood* merupakan fungsi dari parameter θ dan dinotasikan dengan $L(\theta)$. *Likelihood* menentukan bagaimana data yang diamati memengaruhi estimasi parameter dan membentuk dasar proses inferensi *posterior*. Inferensi *posterior* adalah proses pengambilan kesimpulan tentang parameter atau hipotesis berdasarkan distribusi *posterior* dalam pendekatan *Bayesian* (Kruschke, 2014). Tanpa *likelihood*, *prior* tidak bisa

diperbarui dan tidak bisa membuat kesimpulan yang didukung oleh data.

Fungsi *likelihood* menggambarkan tingkat kemungkinan munculnya data yang diamati bergantung pada nilai parameter yang belum diketahui. Bentuk umum dari fungsi *likelihood* sebagai berikut (Casella dan Berger, 2002):

$$L(\theta \mid y) = P(y \mid \theta) \tag{2.17}$$

$$L(\beta, \gamma | y, x) = P(y | x; \beta, \gamma) \tag{2.18}$$

$$L(\beta, \gamma | y, x) = P(y = J | x; \beta, \gamma) \tag{2.19}$$

dimana:

- $L(\theta \mid y)$: fungsi *likelihood* dari parameter θ berdasarkan data y
- $P(y \mid \theta)$: probabilitas data y, dengan parameter θ tetap

Sehingga, fungsi *likelihood* berdasarkan probabilitas (2.11) sebagai berikut (Greene, 2018):

$$L(\beta, \gamma | y, x) = \Phi(\gamma_J - x^\top \beta) - \Phi(\gamma_{J-1} - x^\top \beta) \tag{2.20}$$

dimana:

- $\Phi(\cdot)$: fungsi distribusi kumulatif normal standar
- $x^\top \beta$: kombinasi linier prediktor
- γ_J : *threshold* atas untuk kategori j
- γ_{J-1} : *threshold* bawah untuk kategori j

Untuk mempermudah proses analisis dan perhitungan, fungsi *likelihood* biasanya ditransformasikan menjadi

log-likelihood melalui logaritma natural (ln) (Taboga, 2021). Untuk membentuk fungsi *ln-likelihood* terhadap parameter β dan γ, diperlukan turunan pertama dari fungsi *likelihood*.

- Turunan pertama *ln-likelihood* terhadap parameter β

  Berdasarkan fungsi *likelihood* yang terdapat pada Persamaan (2.20), maka bentuk fungsi *ln-likelihood* sebagai berikut:

$$L(\beta, \gamma | y, x) = \ln \Big[\Phi(\gamma_J - x^\top \beta) - \Phi(\gamma_{J-1} - x^\top \beta) \Big] \tag{2.21}$$

  sehingga turunan parsial *ln-likelihood* terhadap parameter β dengan menggunakan aturan rantai diberikan oleh $\frac{\partial L}{\partial \beta} = \frac{1}{f(\beta)} \cdot \frac{\partial f(\beta)}{\partial \beta}$ dengan $f(\beta) = \Phi(\gamma_J - x^\top \beta) - \Phi(\gamma_{J-1} - x^\top \beta)$ maka (Casella dan Berger, 2002):

$$\begin{aligned} \frac{\partial L}{\partial \beta} &= \frac{1}{\Phi(\gamma_J - x^\top \beta) - \Phi(\gamma_{J-1} - x^\top \beta)} \\ &\quad \cdot \frac{\partial}{\partial \beta} \Big[\Phi(\gamma_J - x^\top \beta) - \Phi(\gamma_{J-1} - x^\top \beta) \Big] \\ &= \frac{1}{\Phi(\gamma_J - x^\top \beta) - \Phi(\gamma_{J-1} - x^\top \beta)} \\ &\quad \cdot \left(\frac{\partial \Phi(\gamma_J - x^\top \beta)}{\partial \beta} - \frac{\partial \Phi(\gamma_{J-1} - x^\top \beta)}{\partial \beta} \right) \end{aligned}$$

  Turunan dari $\Phi(\gamma_J - x^\top \beta)$ dan $\Phi(\gamma_{J-1} - x^\top \beta)$ dihitung menggunakan aturan rantai dengan γ dianggap konstan dan turunan dari $-x^\top \beta$ terhadap β adalah $-x$ (Bishop, 2006). Menurut (Casella dan Berger, 2002) diketahui bahwa turunan dari CDF (*Cumulative Distribution Function*) normal standar yang dilambangkan dengan $\Phi(z)$ adalah PDF (*Probability Density Function*) normal standar yang dilambangkan

$\phi(z)$. Secara matematis dapat dituliskan sebagai:

$$\frac{\partial}{\partial z}\Phi(z) = \phi(z)$$

dengan $z = (z_1, z_2)$ dimana $z_1 = \gamma_J - x^\top\beta$ dan $z_2 = \gamma_{J-1} - x^\top\beta$, maka hasil turunan dapat dinyatakan sebagai berikut (Casella dan Berger, 2021):

$$\begin{aligned}\frac{\partial\Phi(\gamma_J - x^\top\beta)}{\partial\beta} &= \frac{\partial\Phi(\gamma_J - x^\top\beta)}{\partial\gamma_J - x^\top\beta} \cdot \frac{\partial\gamma_J - x^\top\beta}{\partial\beta} \\ &= \phi(\gamma_J - x^\top\beta) \cdot \frac{\partial}{\partial\beta}(\gamma_J - x^\top\beta) \\ &= \phi(\gamma_J - x^\top\beta) \cdot (-x) \\ &= -\phi(\gamma_J - x^\top\beta) \cdot x\end{aligned} \tag{2.22}$$

$$\begin{aligned}\frac{\partial\Phi(\gamma_{J-1} - x^\top\beta)}{\partial\beta} &= \frac{\partial\Phi(\gamma_{J-1} - x^\top\beta)}{\partial\gamma_{J-1} - x^\top\beta} \cdot \frac{\partial\gamma_{J-1} - x^\top\beta}{\partial\beta} \\ &= \phi(\gamma_{J-1} - x^\top\beta) \cdot \frac{\partial}{\partial\beta}(\gamma_{J-1} - x^\top\beta) \\ &= \phi(\gamma_{J-1} - x^\top\beta) \cdot (-x) \\ &= -\phi(\gamma_{J-1} - x^\top\beta) \cdot x\end{aligned} \tag{2.23}$$

Sehingga, turunan *ln-likelihood* terhadap parameter β menjadi:

$$\begin{aligned}\frac{\partial L}{\partial \beta} &= \frac{1}{\Phi(\gamma_J - x^\top\beta) - \Phi(\gamma_{J-1} - x^\top\beta)} \\ &\quad \cdot \Big(-\phi(\gamma_J - x^\top\beta)\cdot x + \phi(\gamma_{J-1} - x^\top\beta)\cdot x\Big) \\ &= \frac{\phi(\gamma_{J-1} - x^\top\beta) - \phi(\gamma_J - x^\top\beta)}{\Phi(\gamma_J - x^\top\beta) - \Phi(\gamma_{J-1} - x^\top\beta)} \cdot x\end{aligned} \tag{2.24}$$

- Turunan pertama *ln-likelihood* terhadap parameter γ

  Untuk memperoleh turunan pertama dari fungsi *ln-likelihood* terhadap parameter γ, dilakukan turunan parsial terhadap batas atas γ_J dan batas bawah γ_{J-1} (Greene, 2012). Fungsi *ln-likelihood* berdasarkan fungsi *likelihood* yang bergantung pada parameter γ, seperti ditunjukkan pada Persamaan (2.21). Maka, turunan parsial terhadap γ_J dan γ_{J-1} adalah:

  a. Turunan parsial terhadap γ_J

     Penurunan dilakukan dengan menerapkan aturan rantai. Bentuk turunan parsial dari fungsi *ln-likelihood* terhadap parameter γ_J dapat dituliskan sebagai berikut:

$$\begin{aligned}\frac{\partial L}{\partial \gamma_J} &= \frac{\partial}{\partial \gamma_J} \ln\Big[\Phi(\gamma_J - x^\top\beta) - \Phi(\gamma_{J-1} - x^\top\beta)\Big] \\ &= \frac{1}{\Phi(\gamma_J - x^\top\beta) - \Phi(\gamma_{J-1} - x^\top\beta)} \\ &\quad \cdot \frac{\partial}{\partial \gamma_J}\Big[\Phi(\gamma_J - x^\top\beta) - \Phi(\gamma_{J-1} - x^\top\beta)\Big] \\ &= \frac{1}{\Phi(\gamma_J - x^\top\beta) - \Phi(\gamma_{J-1} - x^\top\beta)} \\ &\quad \cdot \left(\frac{\partial \Phi(\gamma_J - x^\top\beta)}{\partial \gamma_J} - \frac{\partial \Phi(\gamma_{J-1} - x^\top\beta)}{\partial \gamma_J}\right)\end{aligned} \tag{2.25}$$

Namun karena $\Phi(\gamma_{J-1} - x^\top\beta)$ tidak bergantung pada γ_J, maka:

$$\frac{\partial\Phi(\gamma_{J-1} - x^\top\beta)}{\partial\gamma_J} = 0$$

Selanjutnya, gunakan aturan rantai untuk menurunkan $\Phi(\gamma_J - x^\top\beta)$ terhadap γ_J. Dengan demikian, turunan parsial dari $\Phi(\gamma_J - x^\top\beta)$ terhadap γ_J adalah (Casella dan Berger, 2021):

$$\begin{aligned}\frac{\partial\Phi(\gamma_J - x^\top\beta)}{\partial\gamma_J} &= \frac{\partial\Phi(\gamma_J - x^\top\beta)}{\partial(\gamma_J - x^\top\beta)} \cdot \frac{\partial(\gamma_J - x^\top\beta)}{\partial\gamma_J} \\ &= \phi(\gamma_J - x^\top\beta) \cdot \frac{\partial}{\partial\gamma_J}(\gamma_J - x^\top\beta) \\ &= \phi(\gamma_J - x^\top\beta) \cdot 1 \\ &= \phi(\gamma_J - x^\top\beta)\end{aligned} \tag{2.26}$$

Sehingga, turunan *ln-likelihood* terhadap γ_J menjadi:

$$\frac{\partial L}{\partial\gamma_J} = \frac{\phi(\gamma_J - x^\top\beta)}{\Phi(\gamma_J - x^\top\beta) - \Phi(\gamma_{J-1} - x^\top\beta)} \tag{2.27}$$

b. Turunan parsial terhadap γ_{J-1}

Sama seperti sebelumnya, penurunan dilakukan dengan menggunakan aturan rantai. Turunan parsial dari fungsi *ln-likelihood* terhadap parameter γ_{J-1} dapat dihitung sebagai berikut:

$$\begin{aligned}\frac{\partial L}{\partial \gamma_{J-1}} &= \frac{\partial}{\partial \gamma_{J-1}} \ln \left[\Phi(\gamma_J - x^\top\beta) - \Phi(\gamma_{J-1} - x^\top\beta)\right] \\ &= \frac{1}{\Phi(\gamma_J - x^\top\beta) - \Phi(\gamma_{J-1} - x^\top\beta)} \\ &\cdot \frac{\partial}{\partial \gamma_{J-1}} \left[\Phi(\gamma_J - x^\top\beta) - \Phi(\gamma_{J-1} - x^\top\beta)\right] \\ &= \frac{1}{\Phi(\gamma_J - x^\top\beta) - \Phi(\gamma_{J-1} - x^\top\beta)} \\ &\cdot \left(\frac{\partial \Phi(\gamma_J - x^\top\beta)}{\partial \gamma_{J-1}} - \frac{\partial \Phi(\gamma_{J-1} - x^\top\beta)}{\partial \gamma_{J-1}}\right)\end{aligned} \tag{2.28}$$

Namun karena $\Phi(\gamma_J - x^\top\beta)$ tidak bergantung pada γ_{J-1}, maka:

$$\frac{\partial \Phi(\gamma_J - x^\top\beta)}{\partial \gamma_{J-1}} = 0$$

Selanjutnya, gunakan aturan rantai untuk menurunkan $\Phi(\gamma_{J-1} - x^\top\beta)$ terhadap γ_{J-1}. Dengan demikian, turunan parsial dari $\Phi(b)$ terhadap γ_{J-1} adalah (Casella dan Berger, 2021):

$$\begin{aligned}\frac{\partial \Phi(\gamma_{J-1} - x^\top\beta)}{\partial \gamma_{J-1}} &= \frac{\partial \Phi(\gamma_{J-1} - x^\top\beta)}{\partial(\gamma_{J-1} - x^\top\beta)} \cdot \frac{\partial(\gamma_{J-1} - x^\top\beta)}{\partial \gamma_{J-1}} \\ &= \phi(\gamma_{J-1} - x^\top\beta) \cdot \frac{\partial}{\partial \gamma_{J-1}}(\gamma_{J-1} - x^\top\beta) \\ &= \phi(\gamma_{J-1} - x^\top\beta) \cdot 1 \\ &= \phi(\gamma_{J-1} - x^\top\beta)\end{aligned} \tag{2.29}$$

Sehingga, turunan *ln-likelihood* terhadap γ_{J-1} menjadi:

$$\frac{\partial L}{\partial \gamma_{J-1}} = -\frac{\phi(\gamma_{J-1} - x^\top \beta)}{\Phi(\gamma_J - x^\top \beta) - \Phi(\gamma_{J-1} - x^\top \beta)} \tag{2.30}$$

3. *Posterior Distribution* (Distribusi *Posterior*)

Setelah mendefinisikan *prior* dan mengetahui fungsi *likelihood*nya, pendekatan *Bayesian* untuk model ini dapat dilanjutkan dengan menghitung distribusi *posterior* dari parameter model menggunakan teorema *Bayes*. Distribusi *posterior* merupakan fungsi densitas bersyarat parameter β dan γ jika diketahui nilai observasi dari prediktor x dan variabel respon y sehingga dapat dituliskan sebagai berikut (Ainul et al., 2018):

$$P(\beta, \gamma|y, x) = \frac{P(y|x; \beta, \gamma).P(\beta, \gamma)}{P(y|x)} \tag{2.31}$$

$$P(\beta, \gamma|y, x) = \frac{P(y|x; \beta, \gamma).P(\beta).P(\gamma)}{P(y|x)} \tag{2.32}$$

dimana:

- $P(\beta, \gamma|y, x)$: probabilitas *posterior*
- $P(y|x; \beta, \gamma)$: fungsi *likelihood*
- $P(\beta)$: probabilitas *prior* untuk β
- $P(\gamma)$: probabilitas *prior* untuk γ
- $P(y|x)$: *marginal likelihood*

Marginal likelihood $P(y|x)$ diperlukan secara teoritis dalam perhitungan *posterior*, tetapi untuk pembangkitan nilai atau penarikan sampel misalnya dengan algoritma *Gibbs Sampling* tidak perlu menghitungnya secara eksplisit. Hal ini karena *marginal likelihood* hanya bertindak sebagai konstanta dan tidak memengaruhi hubungan antara data

dan parameter dalam perhitungan relatif (Bishop, 2006). *Marginal likelihood* $P(y|x)$ akan hilang dalam perbandingan rasio karena dalam perhitungan distribusi *posterior* secara proporsional, yang menjadi fokus adalah hubungan antara *likelihood* dan *prior*, bukan nilai absolut dari distribusi *posterior* itu sendiri. Bentuk proporsional dari distribusi *posterior* adalah (Soejoeti dan Soebanar, 1988):

$$P(\beta, \gamma|y, x) \propto P(y|x; \beta, \gamma).P(\beta).P(\gamma) \tag{2.33}$$

Distribusi *posterior* ini memberikan dasar untuk menghitung estimator dan interval kepercayaan dari parameter yang belum diketahui dalam pendekatan *Bayesian*. Namun, distribusi *posterior* terlalu kompleks untuk dihitung atau diintegrasikan secara analitik, sehingga perlu dilakukan teknik penarikan sampel *Markov Chain Monte Carlo* (MCMC) dengan algoritma *Gibbs Sampling* untuk menarik sampel dari *posterior* secara tidak langsung (Kruschke, 2014).

5. *Markov Chain Monte Carlo* (MCMC)

Markov Chain Monte Carlo (MCMC) merupakan teknik simulasi yang digunakan untuk memperoleh distribusi *posterior* dalam analisis *Bayesian* yang kompleks dan sulit dihitung secara langsung (Ntzoufras, 2009). Untuk mendapatkan distribusi *posterior* yang stasioner (stabil), MCMC menggunakan simulasi *Monte Carlo* secara iteratif untuk menghasilkan data parameter yang sesuai dengan proses *Markov Chain*. Berbeda dengan teknik simulasi langsung yang tidak dapat menghasilkan sampel dari distribusi *posterior* yang kompleks, MCMC memiliki karakteristik yang umum dan fleksibel.

Salah satu cara untuk mengevaluasi konvergensi hasil estimasi parameter *posterior* adalah dengan menganalisis pola dari *trace plot*. *Trace plot* adalah grafik yang menunjukkan perubahan

nilai parameter hasil simulasi MCMC terhadap sejumlah iterasi. Iterasi awal dalam proses simulasi MCMC dikategorikan sebagai *burn-in period* dan sengaja diabaikan karena nilai-nilai parameter yang dihasilkan pada tahap tersebut belum mencerminkan distribusi *posterior* yang sebenarnya. Estimasi *burn-in period* dapat dilakukan dengan meninjau *trace plot* dari hasil simulasi parameter atau fungsi tertentu dari parameter terhadap urutan iterasi (Syafitri et al., 2021). Jika *trace plot* menunjukkan pola yang acak dan tidak ada *trend* sistematik, maka parameter sudah stabil dan mendekati distribusi *posterior*. Sebaliknya jika *trace plot* menunjukkan fluktuasi yang besar atau *trend* yang jelas maka rantai MCMC belum stabil dan dibutuhkan lebih banyak iterasi. Salah satu algoritma dalam MCMC yang digunakan untuk menghasilkan sampel dari distribusi *posterior* yang kompleks dan sulit dihitung secara langsung adalah *Gibbs Sampling* (Gelman et al., 2004).

6. Algoritma *Gibbs Sampling*

Gibbs Sampling merupakan salah satu metode untuk menghasilkan variabel *random* dari distribusi *marginal* tanpa perlu menghitung fungsi kepadatan distribusinya secara eksplisit (Casella dan George, 1992). Proses ini bekerja secara iteratif dengan memanfaatkan prinsip rantai Markov, dimana setiap parameter dalam model diestimasi berdasarkan nilai parameter lain yang diperoleh dari iterasi sebelumnya.

Algoritma *Gibbs Sampling* digunakan untuk mendekati distribusi *posterior* dari parameter $\theta = (\beta_1, \beta_2, ..., \beta_p; \gamma_1, \gamma_2, ...\gamma_{J-1})$ berdasarkan data yang diamati y. Dalam metode *Gibbs Sampling*, pengambilan sampel secara langsung dari distribusi *posterior* bersama seluruh parameter sering kali tidak memungkinkan karena bentuk distribusinya yang kompleks. Sebagai alternatif, proses dilakukan dengan menarik sampel secara berurutan dari distribusi kondisional masing-masing parameter (Gelfand dan Smith, 1990). Distribusi kondisional adalah distribusi probabilitas suatu variabel *random*,

dengan asumsi bahwa nilai variabel dan parameter lain sudah diketahui. Distribusi *posterior* kondisional dari parameter β adalah (Gelman et al., 2013):

$$P(\beta|y, x, \gamma) \sim \mathcal{N}(\mu_\beta, \Sigma_\beta) \tag{2.34}$$

μ_β dan Σ_β merupakan parameter distribusi *posterior* yang dihitung berdasarkan fungsi *likelihood* dari data dan distribusi *prior* terhadap β yang mengikuti distribusi normal multivariat dengan:

- μ_β : *mean* dari distribusi *posterior* kondisional β
- Σ_β : kovarians dari distribusi *posterior* kondisional β

Distribusi normal multivariat adalah distribusi probabilitas yang menggeneralisasi distribusi normal satu variabel ke lebih dari satu variabel *random* yang saling berkorelasi (Anderson, (2003).

Setelah membangkitkan nilai dari parameter β, selanjutnya dihitung distribusi *posterior* kondisional dari parameter γ yaitu (Gelman et al., 2013):

$$P(\gamma|y, x, \beta) \sim \mathcal{N}(\mu_\gamma, \sigma^2_\gamma) \tag{2.35}$$

μ_γ dan σ^2_γ merupakan parameter yang dihitung dari data dan distribusi *posterior* kondisional yang mengikuti distribusi normal univariat dengan:

- μ_γ : *mean* dari distribusi *posterior* kondisional γ
- σ^2_γ : varians dari distribusi *posterior* kondisional γ

Distribusi normal univariat adalah distribusi probabilitas kontinu untuk satu variabel *random*, yang ditentukan oleh dua parameter yaitu μ dan σ^2 (Casella dan Berger, 2002).

Menurut Mahfiyah dan Agoestanto (2021), langkah-langkah algoritma *Gibbs Sampling* adalah:

1. Menentukan nilai awal untuk β dan γ yaitu $\left\{\beta^{(0)}, \gamma^{(0)}\right\}$.

2. Membangkitkan sampel $y^{*(t)}$

   Dalam model *probit ordinal*, variabel respon y tidak dimodelkan secara langsung, melainkan dianggap sebagai hasil kategorisasi dari variabel laten kontinu y^* sebagaimana dinyatakan dalam Persamaan (2.7) (McCullagh, 1980). Oleh karena itu, dalam implementasi *Gibbs Sampling*, sampel y^* perlu disimulasikan dari distribusi kondisionalnya pada setiap iterasi ke-t.

   Agar sampel y^* yang dihasilkan konsisten dengan kategori y, maka y^* harus diambil dari distribusi normal yang dipangkas (*truncated normal*) sesuai dengan batas bawah dan batas atas berdasarkan nilai *threshold* γ. Distribusi *truncated normal* memungkinkan pengambilan sampel yang valid secara probabilistik untuk variabel laten kontinu yang tidak teramati secara langsung (Albert dan Chib, 1993). Hal ini merupakan langkah penting dalam menjaga konsistensi antara data yang diamati dengan variabel laten dalam setiap iterasi *Gibbs Sampling*.

   Distribusi normal terpotong (*truncated normal*) adalah distribusi normal yang dibatasi pada interval $[a, b]$, dimana a adalah batas bawah dan b adalah batas atas (Robert, 2007). Artinya, nilai yang dihasilkan dari distribusi ini akan selalu berada dalam interval tersebut. Karena batas pemotongan ditentukan oleh threshold γ, maka intervalnya adalah $[\gamma_{J-1}, \gamma_J]$, dengan J menunjukkan kategori ke-J dari variabel respon y.

   Jika variabel laten mengikuti distribusi normal sebagai berikut:

   $$y^* \sim \mathcal{N}(\mu, \sigma^2), \tag{2.36}$$

   maka sampel y^* yang dibangkitkan untuk kategori tertentu berada dalam interval $[\gamma_{J-1}, \gamma_J]$, sehingga distribusinya

menjadi (Albert dan Chib, 1993):

$$y^* \mid (\gamma_{J-1} \leq y^* \leq \gamma_J) \sim \mathcal{TN}(\mu, \sigma^2; \gamma_{J-1}, \gamma_J). \tag{2.37}$$

Dengan demikian, fungsi kepadatan probabilitas dari y^* yang mengikuti distribusi normal adalah (Nisa dan Miasary, 2024):

$$f(y^*) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{y^* - \mu}{\sigma}\right)^2\right), \tag{2.38}$$

yang juga dapat dinyatakan dalam bentuk alternatif sebagai:

$$f(y^*) = \frac{1}{\sigma}\,\phi\left(\frac{y^* - \mu}{\sigma}\right), \tag{2.39}$$

dengan $\phi(z) = \dfrac{1}{\sqrt{2\pi}} \exp\left(-\dfrac{1}{2}z^2\right)$ adalah fungsi kepadatan dari distribusi normal standar, dan $z = \dfrac{y^* - \mu}{\sigma}$ adalah bentuk standarisasi dari y^*.

Distribusi *truncated normal* antara a dan b diperoleh dengan menormalkan distribusi normal biasa pada interval tersebut, yaitu:

$$f_{\text{trunc}}(y^*) = \frac{f(y^*)}{P(a \leq y^* \leq b)} = \frac{\dfrac{1}{\sigma}\phi\left(\dfrac{y^* - \mu}{\sigma}\right)}{\Phi\left(\dfrac{b - \mu}{\sigma}\right) - \Phi\left(\dfrac{a - \mu}{\sigma}\right)}, \quad a \leq y^* \leq b, \tag{2.40}$$

sehingga fungsi kepadatan probabilitas dari y^* yang mengikuti $\mathcal{TN}(\mu, \sigma^2; \gamma_{J-1}, \gamma_J)$ dituliskan sebagai (Burkardt, 2014):

$$f_{\text{trunc}}(y^*) = \frac{1}{\sigma} \cdot \frac{\phi\left(\frac{y^* - \mu}{\sigma}\right)}{\Phi\left(\frac{\gamma_J - \mu}{\sigma}\right) - \Phi\left(\frac{\gamma_{J-1} - \mu}{\sigma}\right)}, \quad \text{untuk } y^* \in [\gamma_{J-1}, \gamma_J], \tag{2.41}$$

dimana:

- $\phi(\cdot)$ adalah PDF (*Probability Density Function*) dari distribusi normal standar.
- $\Phi(\cdot)$ adalah CDF (*Cumulative Distribution Function*) dari distribusi normal standar.

Pada setiap iterasi ke-t dalam *Gibbs Sampling*, sampel $y^{*(t)}$ perlu dibangkitkan dari distribusi *truncated normal* sebagai bagian dari langkah pembaruan. Namun, karena distribusi *truncated normal* tidak memiliki bentuk tertutup yang memungkinkan pembangkitan sampel secara langsung seperti pada distribusi normal standar serta bentuk PDF yang kompleks, maka diperlukan metode numerik. Salah satu metode yang umum digunakan adalah *Inverse Transform Sampling* (Devroye, 1986). Metode ini dimanfaatkan dalam *Gibbs Sampling* ketika distribusi kondisional yang diperoleh sulit ditangani secara langsung, seperti *truncated normal*. Dengan demikian, *Gibbs Sampling* berperan dalam menentukan distribusi kondisional penuh, sedangkan *Inverse Transform Sampling* digunakan untuk menghasilkan sampel dari distribusi tersebut. Proses pembangkitan sampel dengan *Inverse Transform Sampling* dilakukan melalui langkah-langkah berikut (Robert, 2009):

(a) Melakukan standarisasi batas bawah (a) dan batas atas (b) terhadap distrbusi normal standar:

$$\alpha = \frac{a - \mu}{\sigma}, \quad \beta = \frac{b - \mu}{\sigma} \tag{2.42}$$

maka hasil standarisasi dengan $a = \gamma_{J-1}$ dan $b = \gamma_J$ diperoleh sebagai berikut:

$$\alpha = \frac{\gamma_{J-1} - \mu}{\sigma}, \quad \beta = \frac{\gamma_J - \mu}{\sigma} \tag{2.43}$$

(b) Menghitung nilai CDF normal standar pada batas bawah dan batas atas:

$$u_1 = \Phi(\alpha), \quad u_2 = \Phi(\beta) \tag{2.44}$$

(c) Mengambil nilai acak u dari distribusi *uniform* pada interval (u_1, u_2):

$$u \sim \text{uniform}(u_1, u_2) \tag{2.45}$$

(d) mentransformasi balik untuk memperoleh nilai dari distribusi *truncated normal*:

$$y^* = \mu + \sigma \cdot \Phi^{-1}(u) \tag{2.46}$$

Setelah memperoleh sampel $y^{*(t)}$ untuk seluruh observasi pada iterasi ke-t, nilai-nilai tersebut disusun dalam bentuk vektor kolom sebagai berikut:

$$\mathbf{y}^{*(\mathbf{t})} = \begin{bmatrix} y_1^{*(t)} \\ y_2^{*(t)} \\ y_3^{*(t)} \\ \vdots \\ y_i^{*(t)} \end{bmatrix} \tag{2.47}$$

dengan i menyatakan banyaknya observasi dan t menunjukkan indeks iterasi.

3. Membangkitkan nilai $\beta^{(t)}$

Nilai $\beta^{(t)}$ dibangkitkan dari Persamaan $P(\beta|y, x, \gamma)$ untuk setiap iterasi $t = 1, 2, .., T$. Dengan diketahuinya $y^{*(t)}$,

pembangkitan nilai $\beta^{(t)}$ dari distribusi *posterior* kondisional yang mengikuti distribusi normal multivariat dilakukan seperti pada Persamaan (2.34) (Gelman et al., 2013):

$$P(\beta^{(t)}|y^{*(t)}, x, \gamma^{(t-1)}) \sim \mathcal{N}(\mu_\beta^{(t)}, \Sigma_\beta^{(t)}) \tag{2.48}$$

dengan:

- Kovarians *posterior*:

$$\begin{aligned} \Sigma_\beta^{(t)} &= \left(x^\top x + \Sigma_0^{-1}\right)^{-1} \\ &= \left(x^\top x + I^{-1}\right)^{-1} \end{aligned} \tag{2.49}$$

- *Mean posterior:*

$$\begin{aligned} \mu_\beta^{(t)} &= \Sigma_\beta^{(t)} \left(x^\top y^* + \Sigma_0^{-1}\mu\right) \\ &= \Sigma_\beta^{(t)} \left(x^\top y^* + I^{-1}\mu\right) \end{aligned} \tag{2.50}$$

diperoleh:

$$\beta^{(t)} \sim \mathcal{N}\left(\mu_\beta^{(t)}, \Sigma_\beta^{(t)}\right) \tag{2.51}$$

Supaya $\beta^{(t)}$ mudah dibangkitan, distribusi tersebut diubah secara langsung dengan menggunakan dekomposisi *Cholesky*. Dekomposisi *Cholesky* adalah teknik faktorisasi matriks yang digunakan terutama untuk matriks positif-definit dan sangat berguna dalam perhitungan *Bayesian* seperti pengambilan sampel dari distribusi normal multivariat (Gelman et al., 2013). Dekomposisi *Cholesky* memfaktorkan matriks kovarians $\Sigma_\beta^{(t)}$ menjadi $\Sigma_\beta^{(t)} = LL^\top$, dimana L adalah matriks segitiga bawah (Golub dan Van Loan, 2013). Dengan hasil dekomposisi ini, kita dapat membangkitkan $\beta^{(t)}$ melalui transformasi:

$$\beta^{(t)} = \mu_{\beta}^{(t)} + L \cdot z \tag{2.52}$$

dimana:

- $\mu_{\beta}^{(t)}$: *mean posterior*
- L : hasil dekomposisi *Cholesky* dari $\Sigma_{\beta}^{(t)}$ (matriks segitiga bawah)
- z : vektor acak dari $\mathcal{N}(0, I)$

Bentuk umum dari matriks segitiga bawah L adalah:

$$L = \begin{bmatrix} \ell_{11} & 0 & 0 & \cdots & 0 \\ \ell_{21} & \ell_{22} & 0 & \cdots & 0 \\ \ell_{31} & \ell_{32} & \ell_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \ell_{i1} & \ell_{i2} & \ell_{i3} & \cdots & \ell_{ij} \end{bmatrix} \tag{2.53}$$

Untuk menghitung elemen-elemen matriks $L = [\ell_{ij}]$, dengan i menyatakan indeks baris dan j menyatakan indeks kolom dari matriks segitiga bawah L pada dekomposisi *Cholesky*, digunakan rumus umum sebagai berikut (Rue dan Held, 2005):

- Untuk elemen diagonal utama:

$$\ell_{ii} = \sqrt{\Sigma_{\beta ii}^{(t)} - \sum_{k=1}^{i-1} \ell_{ik}^2} \tag{2.54}$$

  k adalah indeks penjumlahan pada operasi $\sum$, yang digunakan untuk menjumlahkan kuadrat elemen-elemen dari baris ke-i pada kolom 1 hingga $i-1$ dari matriks L.

- Untuk elemen di bawah diagonal utama:

$$\ell_{ij} = \frac{1}{\ell_{jj}} \left(\Sigma^{(t)}_{\beta ij} - \sum_{k=1}^{j-1} \ell_{ik} \ell_{jk} \right) \tag{2.55}$$

k adalah indeks penjumlahan pada operasi $\sum$, yang digunakan untuk menjumlahkan hasil kali elemen ℓ_{ik} dan ℓ_{jk}, dengan k bervariasi dari 1 hingga $j-1$ dari matriks L.

dimana:

- elemen diagonal utama : elemen yang indeks baris dan kolomnya sama ($i = j$)
- elemen di bawah diagonal utama : elemen dengan indeks baris lebih besar darip indeks kolom ($i > j$)

Proses dekomposisi *Cholesky* ini dilakukan secara bertahap untuk menghitung setiap elemen $\ell_{11}, \ell_{21}, \ldots, \ell_{ij}$. Setelah seluruh elemen diagonal utama dan di bawah diagonal utama berhasil dihitung, maka matriks segitiga bawah L dapat dibentuk secara lengkap. Setelah itu, nilai $\beta^{(t)}$ dapat dibangkitkan dari Persamaan (2.52) dengan z adalah vektor variabel *random* yang mengikuti distribusi normal standar multivariat.

4. Membangkitkan nilai $\gamma^{(t)}$

   Setelah $\beta^{(t)}$ diketahui, langkah selanjutnya adalah menentukan distribusi *posterior* kondisional dari $\gamma^{(t)}$ seperti pada Persamaan (2.35):

$$\gamma^{(t)} \mid y^*, x, \beta^{(t)} \sim \mathcal{N}(\mu^{(t)}_{\gamma}, \sigma^{2\,(t)}_{\gamma}) \tag{2.56}$$

   Untuk setiap *threshold* γ, dihitung nilai maksimum dan minimum dari batas tersebut. Untuk memisahkan kategori

J dan $J+1$ digunakan nilai dari y^* (Albert dan Chib, 1993):

$$a = \max\{y^{*(t)} \mid y = J\}, \quad b = \min\{y^{*(t)} \mid y = J+1\} \tag{2.57}$$

lalu diambil sampel $\gamma^{(t)}$ dari distribusi *unifom*:

$$\gamma^{(t)} \sim \text{uniform}(a, b) \tag{2.58}$$

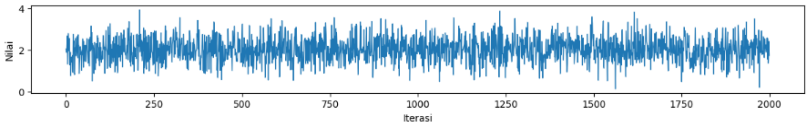
ulangi langkah tersebut untuk $j = 1, 2, ..., J-1$. Sehingga diperoleh:

$$\gamma^{(t)} = \left(\gamma_1^{(t)}, \gamma_2^{(t)}, \gamma_2^{(t)}, \ldots, \gamma_{J-1}^{(t)}\right) \tag{2.59}$$

Setelah membangkitkan nilai $\gamma^{(t)}$ dari nilai $\beta^{(t)}$ dan $y^{*(t)}$, maka proses iterasi dilanjutkan dengan kembali memperbarui variabel laten $y^{*(t+1)}$ menggunakan nilai terkini $\beta^{(t)}$ dan $\gamma^{(t)}$. Selanjutnya, parameter regresi $\beta^{(t+1)}$ diperbarui berdasarkan nilai $y^{*(t+1)}$ yang baru, diikuti dengan pembaruan parameter ambang $\gamma^{(t+1)}$. Siklus ini terus berulang hingga iterasi mencapai konvergen.

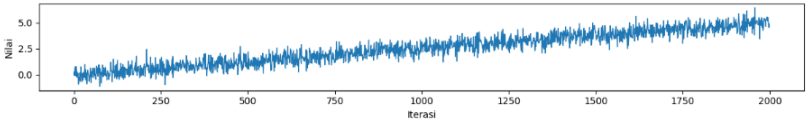
5. Memeriksa konvergensi menggunakan *trace plot*

   Pemeriksaan konvergensi menggunakan *trace plot* diperlukan untuk memastikan bahwa rantai MCMC telah mencapai distribusi stasioner. Jika *trace plot* menunjukkkan pola yang acak dan tidak ada *trend* sistematik, maka parameter sudah stabil dan mendekati distribusi *posterior* (Gelman et al., 2013).



Gambar 2.1. *Trace Plot* dengan Iterasi Konvergen

Gambar 2.1. *trace plot* tersebut menunjukkan bahwa iterasi sudah konvergen karena tidak ada *trend* naik atau *trend* turun.



Gambar 2.2. *Trace Plot* dengan Iterasi Tidak Konvergen

Gambar 2.2. *trace plot* menunjukkan iterasi yang tidak konvergen karena terdapat *trend* naik.

7. Uji Signifikansi Parameter Secara Simultan

Uji simultan terhadap koefisien model dilakukan untuk mengetahui signifikansi koefisien regresi β maupun *threshold* γ terhadap variabel respon secara simultan atau bersama-sama, dengan perumusan hipotesis sebagai berikut:

$H_0 : \beta_1 = \beta_2 = ... = \beta_p = 0$ dan $\gamma_1 = \gamma_2 =, ... = \gamma_{J-1} = 0$

H_1 : terdapat setidaknya satu $\beta_m \neq 0$ atau $\gamma_j \neq 0$, dimana $m = 1, 2, 3, ..., p$ dan $j = 1, 2, 3, ..., J - 1$

Statistik uji G^2 yang dikenal sebagai *Likelihood Ratio Test* (LRT), digunakan untuk menguji signifikansi parameter model secara simultan, dan dirumuskan sebagai berikut (Hosmer dan Lemeshow, 1989):

$$G^2 = -2[logL(\hat{\beta}_0, \hat{\gamma_0}) - logL(\hat{\beta}, \hat{\gamma})] \tag{2.60}$$

dimana:

$$\log L(\hat{\beta}, \hat{\gamma}) = \log \left[\Phi(\gamma_J - x^\top \beta) - \Phi(\gamma_{J-1} - x^\top \beta\right] \tag{2.61}$$

$$\log L(\hat{\beta}_0, \hat{\gamma}_0) = \log \left[\Phi(\gamma_J) - \Phi(\gamma_{-1})\right] \tag{2.62}$$

dengan $L(\hat{\beta}, \hat{\gamma})$ adalah fungsi *likelihood* untuk model yang menggunakan variabel prediktor dan $L(\hat{\beta}_0, \hat{\gamma_0})$ adalah fungsi *likelihood* yang diasumsikan $\beta = 0$ atau tidak ada pengaruh dari variabel prediktor.

Daerah kritisnya : Tolak H_0 jika nilai $G^2 > \chi^2_{\alpha,v}$ atau $P - value < \alpha$. Statistik uji G^2 mengikuti distribusi chi-kuadrat (χ^2) dengan tingkat signifikansi α dan derajat kebebasan v, dengan $v = p$ yaitu jumlah koefisien regresi β yang diuji.

8. Uji Signifikansi Parameter Secara Parsial

Apabila hasil uji signifikansi simultan menunjukkan penolakan terhadap H_0, maka dilanjutkan dengan pengujian secara parsial. Uji ini bertujuan untuk mengevaluasi pengaruh masing-masing parameter β dan γ secara terpisah, dengan hipotesis yang dirumuskan sebagai berikut:

Untuk koefisien regresi β,

$H_0 : \beta_1 = \beta_2 = ... = \beta_p = 0$

H_1 : terdapat setidaknya satu $\beta_m \neq 0$

Untuk *threshold* γ,

$H_0 : \gamma_1 = \gamma_2 = ... = \gamma_{J-1} = 0$

H_1 : terdapat setidaknya satu $\gamma_j \neq 0$

Uji signifikansi parameter secara parsial dilakukan dengan menggunakan statistik uji *Wald* Agresti (2013), sebagai berikut:

$$W = \left(\frac{\hat{\beta}}{\mathrm{SE}(\hat{\beta})} \right)^2 \tag{2.63}$$

dengan:

$$SE(\hat{\beta}) = \sqrt{\mathrm{var}(\hat{\beta})} \tag{2.64}$$

dimana:

- W : statistik uji $Wald$

- SE : standard error dari $\hat{\beta}$

Keputusan untuk menolak H_0 diambil ketika nilai statistik uji berada di luar batas kritis, yaitu $|W| > Z_{\alpha/2}$, atau saat P-value $< \alpha$.

9. Uji Kesesuaian Model

Uji kesesuaian model digunakan untuk mengevaluasi apakah model regresi probit ordinal simultan yang diperoleh telah sesuai, dalam arti tidak terdapat perbedaan yang signifikan antara data observasi dengan hasil prediksi model. Hipotesis yang digunakan dalam pengujian ini adalah sebagai berikut (Maslihah dan Nisa, 2024):

H_0 : model yang dihipotesiskan fit dengan data, artinya tidak terdapat perbedaan antara hasil observasi dengan kemungkinan hasil prediksi model

H_1 : model yang dihipotesiskan tidak fit dengan data, artinya terdapat perbedaan antara hasil observasi dengan kemungkinan hasil prediksi model

Salah satu pendekatan yang umum digunakan untuk menguji kesesuaian ini adalah statistik uji *Pearson*. Statistik uji *Pearson* $\hat{C}$ diformulasikan (Hosmer dan Lemeshow, 1989):

$$\hat{C} = \sum_{k=1}^{g} \frac{(o_k - n_k\,\bar{\pi}_k)^2}{n_k\,\bar{\pi}_k(1-\bar{\pi}_k)} \tag{2.65}$$

dengan:

$$o_k = \sum_{i \in G_k} \mathbb{I}(y_i = j)$$

$$\bar{\pi}_k = \frac{1}{n_k} \sum_{i \in G_k} \hat{\pi}_{ij}$$

dimana:

- o_k: jumlah observasi pada kelompok ke-k dengan nilai aktual

$y_i = j$, dihitung menggunakan fungsi indikator $\mathbb{I}(y_i = j)$ (bernilai 1 jika benar, 0 jika tidak)

- $\bar{\pi}_k$: *mean* probabilitas prediksi bahwa $y_i = j$ dalam kelompok ke-k, di mana $\hat{\pi}_{ij}$ adalah probabilitas prediksi bahwa observasi ke-i berada pada kategori ke-j
- n_k: banyaknya observasi dalam kelompok ke-k
- g: jumlah kelompok

H_0 ditolak jika nilai statistik uji $\hat{C}$ lebih besar daripada nilai kritis $\chi^2_{(\alpha,g-2)}$ atau $P - value < \alpha$ yang menyatakan bahwa model tidak fit artinya terdapat perbedaan antara hasil observasi dengan kemungkinan hasil prediksi model (Liu dan Zhang, 2017).

10. Nilai Koefisien Determinasi *McFadden* (R^2_{McF})

Untuk mengetahui seberapa besar pengaruh variabel prediktor terhadap variabel respon digunakan koefisien determinasi (*R-squared*). Koefisien determinasi merupakan pengujian untuk mengetahui persentase perubahan variabel respon (y) yang disebabkan oleh variabel prediktor (x) (Sujarweni, 2015). Nilai *R-squared* yang tinggi mengindikasikan bahwa variabel prediktor dalam model memiliki kemampuan yang baik dalam menjelaskan variasi yang terjadi pada variabel respon. Sebaliknya, nilai *R-squared* yang rendah menunjukkan bahwa kontribusi variabel prediktor dalam menjelaskan variabilitas respon relatif lemah. Untuk regresi linier menggunakan koefisien determinasi *R-squared*, sedangkan untuk regresi probit menggunakan *McFadden R-Squared*. Nilai *McFadden R-Squared* berkisar antara 0 dan 1 dan dapat dirumuskan sebagai berikut (Ghozali, 2017):

- Jika nilai *McFadden R-Squared* semakin mendekati 1, maka semakin besar kemampuan model dalam menjelaskan perubahan dari variabel prediktor terhadap variabel respon.

- Jika nilai *McFadden R-Squared* semakin mendekati 0, berarti semakin kecil kemampuan model dalam menjelaskan perubahan dari variabel prediktor terhadap variabel respon.

Nilai *McFadden R-Squared* dapat dihitung menggunakan Persamaan berikut (Drapper dan Smithh, 1992):

$$R^2_{\text{McF}} = 1 - \ln\left(\frac{L_{\text{M}}}{L_0}\right) \tag{2.66}$$

dimana :

- L_M adalah nilai estimasi fungsi *likelihood* dari model yang melibatkan variabel prediktor
- L_0 adalah fungsi *likelihood* yang diasumsikan $\beta = 0$ atau tidak ada pengaruh dari variabel prediktor
- R^2_{McF} adalah nilai koefisien determinasi *McFadden*

11. Indeks Pembangunan Manusia (IPM)

Konsep pembangunan manusia berfokus pada manusia sebagai tujuan utama pembangunan, bukan semata sebagai sarana untuk mencapai tujuan pembangunan lainnya (Bonaraja, 2021). Menurut UNDP (*United Nations Development Programe*), pembangunan manusia merupakan suatu proses untuk memperbanyak pilihan-pilihan yang dimiliki oleh manusia. Dimana pilihan-pilihan tersebut terdiri dari tiga komponen dasar, yaitu untuk berumur panjang dan sehat, untuk memiliki ilmu pengetahuan, dan yang ketiga untuk mempunyai akses terhadap sumber daya yang dibutuhkan sehingga dapat menjalani kehidupan yang layak (HDR, 2014). Oleh karena itu, Indeks Pembangunan Manusia (IPM) didefinisikan sebagai suatu indeks komposit yang dirancang untuk mengukur capaian pembangunan manusia dengan mengacu pada beberapa komponen dasar kualitas hidup manusia. Indeks

kesehatan, indeks pendidikan dan standar hidup layak adalah tiga komponen utama dalam mengukur taraf kesejahteraan manusia.

Ketiga komponen utama penyusun IPM memiliki cakupan makna yang luas karena masing-masing dipengaruhi oleh berbagai faktor yang saling terkait. Pengukuran indeks kesehatan dalam IPM dapat dilakukan menggunakan indikator Angka Harapan Hidup (AHH) saat lahir. Nilai AHH diperoleh berdasarkan dua komponen utama, yaitu jumlah Anak Lahir Hidup (ALH) dan Anak Masih Hidup (AMH). Sementara itu, pengukuran indeks pendidikan mencakup dua indikator, yaitu Harapan Lama Sekolah (HLS) dan Rata-rata Lama Sekolah (RLS). HLS dihitung untuk penduduk usia 7 tahun ke atas dengan asumsi bahwa kemungkinan anak untuk terus melanjutkan pendidikan setara dengan peluang bersekolah penduduk lain pada usia yang sama saat ini. RLS adalah jumlah tahun yang dihabiskan oleh penduduk usia 15 tahun ke atas untuk bersekolah secara formal. Adapun untuk mengukur indeks standar hidup layak digunakan indikator Tingkat Pengangguran Terbuka (TPT) dan Tingkat Partisipasi Angkatan Kerja (TPAK) yang mewakili pencapaian pembangunan untuk kehidupan yang layak (Maumere, 2016).

B. Kajian Pustaka

1. Penelitian-Penelitian Terdahulu

Penelitian terdahulu terkait analisis Indeks Pembangunan Manusia (IPM) yang dijadikan dasar dalam pemilihan variabel-variabel yang akan digunakan sebagai faktor-faktor yang memengaruhi IPM sebagai berikut:

a). Penelitian yang dilakukan oleh Ginting (2023) dengan menganalisis pengaruh Angka Harapan Hidup (AHH) dan Harapan Lama sekolah (HLS) terhadap IPM di Kabupaten Langkat dari tahun 2013 sampai 2022, hasilnya menunjukkan bahwa variabel AHH dan HLS berpengaruh terhadap IPM secara simultan.

b). Syahid (2022) melakukan analisis faktor-faktor yang memengaruhi IPM di Provinsi Papua. Diperoleh kesimpulan bahwa secara simultan, variabel Rata-rata Lama Sekolah (RLS), Rata-rata Pengeluaran Per Kapita (PPP) dan Tingkat Pengangguran Terbuka (TPT) terbukti berpengaruh terhadap IPM di seluruh kabupaten dan kota di Provinsi Papua selama periode 2018 hingga 2020.

c). Tingkat Partisipasi Angkatan Kerja (TPAK) dikaji sebagai faktor yang memengaruhi Indeks Pembangunan Manusia (IPM), dengan kesimpulan bahwa produktivitas yang tinggi dapat meningkatkan daya saing tenaga kerja di pasar, sehingga TPAK diperkirakan berpengaruh terhadap IPM (Faelassuffa, 2021).

d). Manurung dan Hutabarat (2021) melakukan pengujian pengaruh angka Harapan Lama Sekolah (HLS), Rata-rata Lama Sekolah (RLS), Pengeluaran Per Kapita (PPP) terhadap Indeks Pembangunan Manusia di Indonesia pada tahun 2019 sampai 2020. Dari penelitian tersebut diperoleh kesimpulan bahwa angka HLS, RLS dan PPP berpengaruh terhadap IPM dengan kontribusi tertinggi adalah variabel HLS diikuti RLS sedangkan kontribusi paling kecil dari variabel PPP.

Penelitian ini memiliki persamaan dengan penelitian-penelitian sebelumnya dalam hal penggunaan variabel prediktor yang diduga berpengaruh terhadap Indeks Pembangunan Manusia (IPM), yaitu Angka Harapan Hidup (AHH), Harapan Lama Sekolah (HLS), Rata-rata Lama Sekolah (RLS), Tingkat Pengangguran Terbuka (TPT), dan Tingkat Partisipasi Angkatan Kerja (TPAK). Perbedaannya terletak pada pendekatan yang digunakan, dimana variabel-variabel tersebut dipilih berdasarkan hasil signifikan dari penelitian terdahulu, dengan harapan memberikan kontribusi yang relevan dalam konteks wilayah dan periode yang dianalisis dalam penelitian ini.

BAB III
METODOLOGI PENELITIAN

A. Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari BPS Provinsi Jawa Tengah mengenai Indeks Pembangunan Manusia (IPM) beserta faktor-faktor yang diduga memengaruhinya pada tahun 2024. Data IPM sebagai variabel respon sedangkan data Angka Harapan Hidup (AHH), Harapan Lama Sekolah (HLS), Rata-rata Lama Sekolah (RLS), Tingkat Pengangguran Terbuka (TPT) dan Tingkat Partisipasi Angkatan Kerja (TPAK) sebagai variabel prediktor.

B. Variabel Penelitian

Berdasarkan pada penelitian terdahulu yang telah dibahas dalam bab sebelumnya, penelitian ini menggunakan sejumlah variabel yang diduga berpengaruh terhadap Indeks Pembangunan Manusia (IPM). Variabel yang digunakan terdiri atas satu variabel respon yang dikategorikan ke dalam empat kategori serta lima variabel prediktor. Rincian susunan variabel yang dianalisis dalam konteks IPM di Provinsi Jawa Tengah pada tahun 2024 disajikan pada Tabel 3.1.

Tabel 3.1. Variabel Penelitian

Variabel	Keterangan	Satuan
y	Indeks Pembangunan Manusia (IPM)	Persen
x_1	Angka Harapan Hidup (AHH)	Tahun
x_2	Harapan Lama Sekolah (HLS)	Tahun
x_3	Rata-rata Lama Sekolah (RLS)	Tahun
x_4	Tingkat Pengangguran Terbuka (TPT)	Persen
x_5	Tingkat Partisipasi Angkatan Kerja (TPAK)	Persen

Penjelasan mengenai variabel-variabel yang digunakan dalam penelitian ini diuraikan sebagai berikut:

1. Indeks Pembangunan Manusia (IPM)

   Indeks Pembangunan Manusia (IPM) merupakan indeks komposit yang mengukur pembangunan manusia berdasarkan tiga indikator, yaitu indikator kesehatan, indikator pendidikan dan standar hidup layak (Kuncoro, 2004). Berdasarkan rentang nilainya, IPM diklasifikasikan menjadi empat kategori, yaitu:

   Tabel 3.2. Nilai IPM dan Kategorinya

Nilai IPM	**Kategori**
IPM < 60	Rendah
$60 \leq$ IPM < 70	Sedang
$70 \leq$ IPM < 80	Tinggi
IPM ≥ 80	Sangat Tinggi

2. Angka Harapan Hidup (AHH)

   Angka Harapan Hidup (AHH) menunjukkan umur rata-rata yang dicapai seseorang dalam kondisi mortalitas yang sesuai di masyarakatnya. AHH yang rendah di suatu daerah menunjukkan bahwa pembangunan kesehatan belum berhasil, sementara AHH yang tinggi menunjukkan bahwa pembangunan kesehatan di daerah tersebut berhasil (Ginting, 2023). AHH adalah standar untuk mengukur rata-rata usia hidup masyarakat atau kelompok. Bayi dengan AHH tinggi memiliki peluang hidup yang lebih panjang, yang dapat berdampak positif pada rata-rata harapan hidup penduduk secara keseluruhan. Perhitungan Angka Harapan Hidup (AHH) menggunakan pendekatan demografis melalui tabel kehidupan (*life table*) yang didasarkan pada data jumlah penduduk dan kematian menurut kelompok umur. Dari data ini dihitung probabilitas bertahan hidup pada

setiap usia, lalu dihitung total tahun hidup seluruh individu dalam suatu kohor atau kelompok individu. AHH diperoleh dengan membagi total tahun hidup tersebut dengan jumlah penduduk pada usia nol tahun.

3. Harapan Lama Sekolah (HLS)

   Harapan Lama Sekolah (HLS) adalah angka partisipasi sekolah berdasarkan usia. Indikator ini menunjukkan berapa lama seorang anak diharapkan belajar (dalam tahun) pada usia tertentu di masa depan (Herdiansyah, 2020). Angka ini diperoleh dengan membagi total jumlah penduduk yang bersekolah pada usia tertentu pada tahun tertentu dengan jumlah total penduduk yang bersekolah pada usia tersebut pada tahun tersebut.

4. Rata-rata Lama Sekolah (RLS)

   Jumlah tahun rata-rata yang dihabiskan oleh individu untuk menyelesaikan seluruh jenjang pendidikan formal yang pernah mereka ikuti dikenal sebagai Rata-rata Lama Sekolah (RLS) (Gaol et al., 2024). Angka ini menunjukkan jumlah tahun yang dihabiskan oleh masyarakat usia 15 tahun ke atas untuk bersekolah secara formal. RLS mendeskripsikan bahwa semakin tinggi tingkat pendidikan masyarakat suatu wilayah, semakin baik pula IPMnya. RLS diartikan sebagai estimasi rata-rata lamanya waktu yang dihabiskan oleh seseorang untuk menempuh pendidikan formal yang diukur dalam satuan tahun.

5. Tingkat Pengangguran Terbuka (TPT)

   Setiap negara pasti akan mengalami pengangguran baik dalam bentuk sosial dan ekonomi (Sukirno, 2016). Pengangguran terbuka adalah kondisi dimana seseorang belum bekerja tetapi sedang berusaha memperoleh pekerjaan. Pengangguran terjadi karena pertumbuhan angkatan kerja lebih besar dari lapangan pekerjaan yang tersedia. Salah satu indikator penting dalam bidang

ketenagakerjaan yaitu tingkat pengangguran yang dapat digunakan untuk menentukan seberapa besar jumlah pekerja yang dapat diserap oleh lapangan kerja yang tersedia (Rambe et al., 2019).

6. Tingkat Partisipasi Angkatan Kerja (TPAK)

   Tingkat Partisipasi Angkatan Kerja (TPAK) menjadi salah satu indikator kesejahteraan dalam penilaian Indeks Pembangunan Manusia (IPM). TPAK dapat digunakan untuk mengetahui berapa banyak penduduk yang memiliki peluang untuk bekerja. Jumlah angkatan kerja yang tinggi dapat menyebabkan potensi penduduk yang ingin bekerja juga tinggi. Hal ini telah menunjukkan bahwa perekonomian suatu daerah berkembang dengan baik (Sari dan Sugiharti, 2022).

C. Metode Analisis Data

Metode analisis data yang digunakan dalam penelitian ini bertujuan untuk mengidentifikasi faktor-faktor yang memengaruhi variabel respon dengan menerapkan model regresi probit ordinal berbasis pendekatan *Bayesian* melalui Algoritma *Gibbs Sampling*. Tahapan penelitian yang akan dilakukan sebagai berikut:

1. Mengumpulkan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) Provinsi Jawa Tengah.

2. Melakukan pengujian terhadap asumsi tidak adanya multikolinieritas antar variabel prediktor. Jika terdapat multikolinieritas, maka dilakukan penanganan untuk mengatasi multikolinieritas dengan pembentukan variabel interaksi. Jika tidak terdapat multikolinieritas, maka dapat dilanjutkan dengan analisis regresi probit ordinal.

3. Estimasi parameter regresi probit ordinal dengan metode *Bayesian* menggunakan algoritma *Gibbs Sampling* dan

pembentukan model lengkap dengan semua variabel prediktor berupa Angka Harapan Hidup (AHH), Harapan Lama Sekolah (HLS), Rata-rata Lama Sekolah (RLS), Tingkat Pengangguran Terbuka (TPT) dan Tingkat Partisipasi Angkatan Kerja (TPAK).

4. Melakukan uji signifikansi parameter secara simultan dengan *Likelihood Ratio Test* dan parsial dengan uji *Wald*. Jika hasil uji simultan menunjukkan bahwa parameter berpengaruh terhadap model, maka dilanjutkan dengan uji parsial untuk mengevaluasi pengaruh setiap parameter. Namun, jika tidak signifikan secara simultan, maka uji parsial tidak dilanjutkan. Apabila dalam uji parsial terdapat parameter yang tidak signifikan, maka dilakukan evaluasi ulang terhadap multikolinieritas dengan hanya mempertimbangkan variabel-variabel yang signifikan.

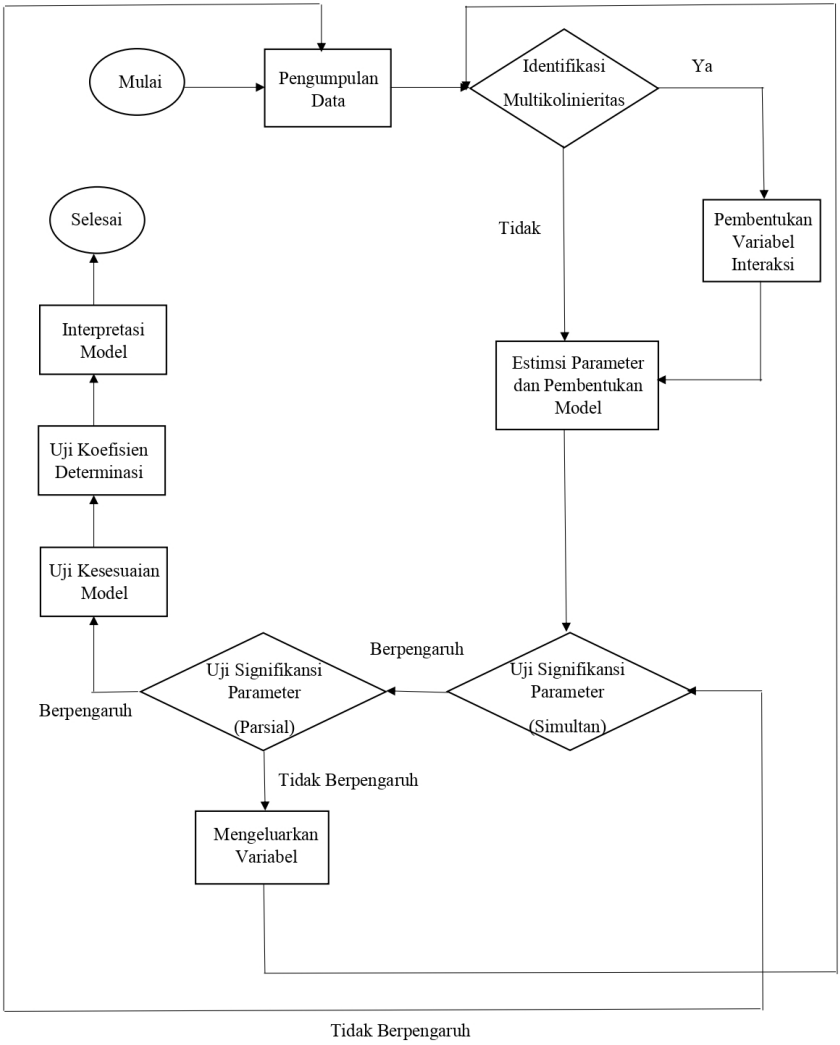
5. Untuk memastikan apakah model yang dibentuk telah sesuai, dilakukan uji *Pearson* yang bertujuan mengidentifikasi adanya perbedaan antara nilai observasi dan hasil prediksi dari model.

6. Melakukan uji koefisien determinasi *McFadden* untuk mengetahui seberapa besar pengaruh variabel prediktor terhadap variabel respon.

7. Interpretasi hasil estimasi model regresi probit ordinal

D. Alur Penelitian

Langkah-langkah penelitian disajikan secara ringkas dalam diagram alir penelitian pada Gambar 3.1.



Gambar 3.1. Diagram Alir Penelitian

BAB IV
HASIL PENELITIAN DAN PEMBAHASAN

Pada bab ini akan dibahas mengenai estimasi parameter regresi probit ordinal dengan metode *Bayesian* menggunakan algoritma *Gibbs Sampling* pada Indeks Pembangunan Manusia (IPM) di Jawa Tengah pada tahun 2024. Setelah mendapatkan model dari hasil estimasi parameternya, maka dapat ditentukan faktor apa saja yang memengaruhi IPM di daerah tersebut. Berikut merupakan hasil penelitian dan pembahasannya.

A. Identifikasi Multikolinieritas

Dalam analisis statistik yang berkaitan dengan pembentukan model, penting untuk memastikan terpenuhinya asumsi bebas multikolinieritas. Pengujian terhadap multikolinieritas dilakukan untuk mendeteksi adanya hubungan yang kuat antar variabel prediktor yang dapat menyebabkan hasil analisis menjadi bias atau tidak akurat. Untuk mengidentifikasi multikolinieritas pada penelitian ini dilakukan dengan cara melihat nilai VIF (*Variance Inflation Factor*) berdasarkan Lampiran 3 yang disajikan pada Tabel 4.1.

Tabel 4.1. Nilai VIF Berdasarkan Hasil Uji Multikolinieritas

Variabel	**VIF**
x_1	4,368856
x_2	8,943120
x_3	9,125712
x_4	2,499289
x_5	2,188210

Berdasarkan Tabel 4.1, seluruh variabel prediktor memiliki nilai VIF kurang dari 10, sehingga dapat disimpulkan bahwa tidak terjadi multikolinieritas antar variabel prediktor dalam

penelitian ini. Tidak ditemukannya interkolerasi antar variabel prediktor menunjukkan bahwa asumsi bebas multikolinieritas telah terpenuhi. Oleh karena itu, analisis dapat dilanjutkan pada tahap estimasi parameter model lengkap dengan seluruh variabel prediktor. Estimasi parameter dilakukan dengan menghitung nilai *posterior* dari parameter regresi β dan *threshold* γ melalui iterasi *Gibbs Sampling*.

B. Estimasi Parameter Regresi Probit Ordinal dengan Metode *Bayesian Gibbs Sampling*

Proses estimasi parameter regresi probit ordinal dengan pendekatan *Bayesian* diawali dengan membangkitkan sampel variabel laten y^* dari distribusi *posterior* kondisionalnya. Sampel y^* ini diperlukan sebagai prasyarat untuk mengestimasi parameter regresi β dan *threshold* γ pada setiap iterasi. Pendekatan ini mengikuti kerangka kerja *Gibbs Sampling*, dimana setiap parameter diperbarui secara bergantian berdasarkan distribusi *posterior* penuhnya. Berikut merupakan langkah-langkah pembangkitan sampel pada iterasi pertama pada algoritma *Gibbs Sampling*.

1. Menentukan nilai awal $\beta^{(0)}$ dan $\gamma^{(0)}$

Untuk model regresi probit ordinal dengan tiga kategori pada variabel respon ($y = 1, 2, 3$), diperlukan dua *threshold*, yaitu γ_1 dan γ_2, yang berfungsi untuk memisahkan ketiga kategori tersebut. Kedua nilai *threshold* ini harus memenuhi syarat:

$$\gamma_1 < \gamma_2 \tag{4.1}$$

Kategori dari variabel respon y ditentukan berdasarkan nilai dari variabel laten y^* sebagaimana ditunjukkan pada Persamaan (2.7) yaitu:

$$\begin{cases} y = 1 & \text{jika } y^* \leq \gamma_1 \\ y = 2 & \text{jika } \gamma_1 < y^* \leq \gamma_2 \\ y = 3 & \text{jika } y^* > \gamma_2 \end{cases} \tag{4.2}$$

Sebagai langkah awal dalam proses estimasi parameter menggunakan *Gibbs Sampling*, diperlukan penetapan nilai awal untuk parameter-parameter model. Nilai awal untuk parameter β yang terdiri dari enam parameter yaitu satu *intercept* dan lima variabel prediktor, diasumsikan bernilai nol:

$$\beta^{(0)} = \left(\beta_0^{(0)}, \beta_1^{(0)}, \beta_2^{(0)}, \beta_3^{(0)}, \beta_4^{(0)}, \beta_5^{(0)}\right) = (0,0,0,0,0,0) \tag{4.3}$$

Sementara itu, nilai awal untuk parameter *threshold* γ ditentukan sebagai berikut:

$$\gamma_1^{(0)} = 0, \quad \gamma_2^{(0)} = 1 \tag{4.4}$$

Berdasarkan Persamaan (4.2), maka pengkategorian variabel respon y menjadi:

$$\begin{cases} y = 1 & \text{jika } y^* \leq 0 \\ y = 2 & \text{jika } 0 < y^* \leq 1 \\ y = 3 & \text{jika } y^* > 1 \end{cases} \tag{4.5}$$

Penetapan nilai-nilai awal ini dilakukan dengan asumsi bahwa variabel laten y^* mengikuti distribusi normal standar, dengan *mean* 0 dan standar deviasi 1.

2. Membangkitkan sampel $y^{*(1)}$

Sebelum membangkitkan nilai $\beta^{(1)}$ dan $\gamma^{(1)}$, proses iterasi diawali dengan membangkitkan sampel $y^{*(1)}$. Seperti yang sudah dijelaskan sebelumnya, regresi probit ordinal memiliki dua *threshold* yaitu γ_1 dan γ_2 yang memisahkan ketiga kategori berdasarkan Persamaan (4.5), sehingga:

- Jika $y = 1$, maka $y^{*(1)} \in (-\infty, 0]$
- Jika $y = 2$, maka $y^{*(1)} \in (0, 1]$
- Jika $y = 3$, maka $y^{*(1)} \in (1, +\infty)$

Untuk menghasilkan nilai $y^{*(1)}$ yang sesuai dengan kategori y, dibutuhkan nilai *mean* μ dari distribusi normal standar yang menjadi dasar pembangkitan $y^{*(1)}$. Nilai μ dihitung berdasarkan kombinasi linier antara vektor parameter regresi $\beta^{(1)}$ dengan vektor prediktor x, yang dapat dinyatakan sebagai:

$$\mu = x^\top \beta^{(1)} \tag{4.6}$$

Namun, karena nilai $\beta^{(1)}$ pada iterasi pertama belum tersedia, maka dalam perhitungan μ digunakan nilai awal yang telah ditetapkan sebelumnya, yaitu $\beta^{(0)}$. Secara eksplisit, perhitungan nilai μ untuk iterasi pertama sebagai berikut:

$$\begin{aligned}
\mu &= x^\top \beta^{(0)} \\
&= \begin{bmatrix} 1 & x_{i1} & x_{i2} & x_{i3} & x_{i4} & x_{i5} \end{bmatrix} \begin{bmatrix} \beta_0^{(0)} \\ \beta_1^{(0)} \\ \beta_2^{(0)} \\ \beta_3^{(0)} \\ \beta_4^{(0)} \\ \beta_5^{(0)} \end{bmatrix} \\
&= 1 \cdot \beta_0^{(0)} + x_{i1} \cdot \beta_1^{(0)} + x_{i2} \cdot \beta_2^{(0)} + x_{i3} \cdot \beta_3^{(0)} + x_{i4} \cdot \beta_4^{(0)} + x_{i5} \cdot \beta_5^{(0)} \\
&= 0 + x_{i1} \cdot 0 + x_{i2} \cdot 0 + x_{i3} \cdot 0 + x_{i4} \cdot 0 + x_{i5} \cdot 0 \\
\mu &= 0
\end{aligned} \tag{4.7}$$

Perhitungan ini dilakukan untuk setiap observasi $i = 1, 2, \ldots, 35$ dan hasil $\mu = 0$ digunakan untuk membangkitkan nilai $y^{*(1)}$ dari distribusi *truncated normal*.

Untuk iterasi pertama, sampel $y^{*(1)}$ dibangkitkan berdasarkan kategori y dengan menggunakan pendekatan *Inverse Transform*

Sampling. Dalam proses ini digunakan CDF dari distribusi normal standar, dengan asumsi bahwa $\sigma = 1$, dan *threshold* yang digunakan seperti pada Persamaan (4.4) dengan *mean* pada Persamaan (4.7). Berdasarkan keterangan tersebut, pembangkitan sampel $y^{*(1)}$ dilakukan melalui langkah-langkah berikut:

a). Jika $y = 1$, maka nilai $y_1^{*(1)}$ diasumsikan berasal dari distribusi *truncated normal* $\mathcal{TN}(0, 1; -\infty, 0)$, yakni distribusi normal standar dengan $\mu = 0$, $\sigma = 1$, serta batas bawah $\gamma_{J-1} = -\infty$ dan batas atas $\gamma_J = 0$. Maka, langkah-langkah pembangkitan $y_1^{*(1)}$ menggunakan pendekatan *Inverse Transform Sampling* dilakukan sebagai berikut:

- Melakukan standarisasi batas bawah (γ_{J-1}) dan batas atas (γ_J) terhadap distribusi normal standar:

$$\alpha = \frac{\gamma_{J-1} - \mu}{\sigma} = \frac{-\infty - 0}{1} = -\infty$$
$$\beta = \frac{\gamma_J - \mu}{\sigma} = \frac{0 - 0}{1} = 0$$

- Menghitung nilai CDF normal standar pada batas bawah dan atas:

$$u_1 = \Phi(\alpha) = \Phi(-\infty) = 0$$
$$u_2 = \Phi(\beta) = \Phi(0) = 0,5$$

- Mengambil nilai acak u dari distribusi *uniform* pada interval $(u_1, u_2) = (0; 0,5)$:

$$u \sim \text{uniform}(0; 0,5)$$
$$u \approx 0,349$$

- Melakukan transformasi balik untuk memperoleh nilai dari distribusi *truncated normal*:

$$\begin{aligned} y_1^{*(1)} &= \mu + \sigma \cdot \Phi^{-1}(u) \\ &= 0 + 1 \cdot \Phi^{-1}(0,349) \\ &= \Phi^{-1}(0,349) \approx -0,39 \end{aligned} \tag{4.8}$$

b). Jika $y = 2$, maka nilai $y_2^{*(1)}$ berasal dari distribusi *truncated normal* $\mathcal{TN}(0,1;\,0,1)$, yaitu distribusi normal standar dengan $\mu = 0$, $\sigma = 1$, serta batas bawah $\gamma_{J-1} = 0$ dan batas atas $\gamma_J = 1$. Maka, langkah-langkah pembangkitan $y_2^{*(1)}$ menggunakan pendekatan *Inverse Transform Sampling* dilakukan sebagai berikut:

- Melakukan standarisasi batas bawah (γ_{J-1}) dan batas atas (γ_J) terhadap distribusi normal standar:

$$\alpha = \frac{\gamma_{J-1} - \mu}{\sigma} = \frac{0-0}{1} = 0$$
$$\beta = \frac{\gamma_J - \mu}{\sigma} = \frac{1-0}{1} = 1$$

- Menghitung nilai CDF normal standar pada batas bawah dan atas:

$$u_1 = \Phi(\alpha) = \Phi(0) = 0,5$$
$$u_2 = \Phi(\beta) = \Phi(1) \approx 0,841$$

- Mengambil nilai acak u dari distribusi *uniform* pada interval $(u_1, u_2) = (0,5; 0,841)$:

$$u \sim \text{uniform}(0,5; 0,841)$$
$$u \approx 0,792$$

- Melakukan transformasi balik untuk memperoleh nilai dari distribusi *truncated normal*:

$$\begin{aligned} y_2^{*(1)} &= \mu + \sigma \cdot \Phi^{-1}(u) \\ &= 0 + 1 \cdot \Phi^{-1}(0,792) \\ &= \Phi^{-1}(0,792) \approx 0,812 \end{aligned} \tag{4.9}$$

c). Jika $y = 3$, maka nilai $y_3^{*(1)}$ berasal dari distribusi *truncated normal* $\mathcal{TN}(0,1;\, 1,\infty)$, yaitu distribusi normal standar dengan $\mu = 0$, $\sigma = 1$, serta batas bawah $\gamma_{J-1} = 1$ dan batas atas $\gamma_J = \infty$. Maka, langkah-langkah pembangkitan $y_3^{*(1)}$ menggunakan pendekatan *Inverse Transform Sampling* dilakukan sebagai berikut:

- Melakukan standarisasi batas bawah (γ_{J-1}) dan batas atas (γ_J) terhadap distribusi normal standar:

$$\alpha = \frac{\gamma_{J-1} - \mu}{\sigma} = \frac{1-0}{1} = 1$$
$$\beta = \frac{\gamma_J - \mu}{\sigma} = \frac{\infty - 0}{1} = \infty$$

- Menghitung nilai CDF normal standar pada batas bawah dan atas:

$$u_1 = \Phi(\alpha) = \Phi(1) \approx 0,841$$
$$u_2 = \Phi(\beta) = \Phi(\infty) = 1$$

- Mengambil nilai acak u dari distribusi *uniform* pada interval $(u_1, u_2) = (0,841; 1)$:

$$u \sim \text{uniform}(0,841; 1)$$
$$u \approx 0,890$$

- Melakukan transformasi balik untuk memperoleh nilai dari distribusi *truncated normal*:

$$\begin{aligned} y_3^{*(1)} &= \mu + \sigma \cdot \Phi^{-1}(u) \\ &= 0 + 1 \cdot \Phi^{-1}(0,890) \\ &= \Phi^{-1}(0,890) \approx 1,227 \end{aligned} \tag{4.10}$$

Berdasarkan perhitungan tersebut, diperoleh nilai $y^{*(1)}$ adalah:

$$y^{*(1)} = \begin{bmatrix} y_1^* \\ y_2^* \\ y_3^* \\ y_4^* \\ \vdots \\ y_{33}^* \\ y_{34}^* \\ y_{35}^* \end{bmatrix} = \begin{bmatrix} 0,812 \\ 0,812 \\ 0,812 \\ -0,39 \\ \vdots \\ 1,227 \\ 0,812 \\ 0,812 \end{bmatrix} \tag{4.11}$$

3. Membangkitkan nilai $\beta^{(1)}$

Setelah nilai variabel laten $y^{*(1)}$ berhasil dibangkitkan pada langkah sebelumnya, tahap selanjutnya dalam proses iterasi *Gibbs Sampling* adalah membangkitkan nilai parameter $\beta^{(1)}$ dari distribusi *posterior* kondisional yang mengikuti distribusi normal multivariat seperti pada Persamaan (2.48) dengan $t = 1$ untuk iterasi pertama, sehingga:

$$P(\beta^{(1)}|y^{*(1)}, x, \gamma^{(0)}) \sim \mathcal{N}(\mu_\beta^{(1)}, \Sigma_\beta^{(1)}) \tag{4.12}$$

Dari probabilitas pada Persamaan (4.12), matriks kovarians *posterior* $\Sigma_\beta^{(1)}$ dan *mean posterior* $\mu_\beta^{(1)}$ dihitung berdasarkan Persamaan (2.49) dan (2.50) sebagai berikut:

$$\Sigma_\beta^{(1)} = \left(x^\top x + I^{-1}\right)^{-1} \tag{4.13}$$

$$\mu_\beta^{(1)} = \Sigma_\beta^{(1)}(x^\top y^* + I^{-1}\mu) \tag{4.14}$$

a). Kovarians Posterior $\Sigma_{\beta}^{(1)}$

Langkah pertama untuk mengetahui nilai $\Sigma_{\beta}^{(1)}$ adalah dengan menghitung *transpose* dari matriks x yang terdiri dari satu *intercept* dan lima variabel prediktor dari 35 observasi. Matriks $x^{\top}$ dapat dihitung dari matriks x yang sudah diketahui pada Lampiran 1 sebagai berikut:

$$x = \begin{bmatrix} 1 & x_{11} & x_{12} & x_{13} & x_{14} & x_{15} \\ 1 & x_{21} & x_{22} & x_{23} & x_{24} & x_{25} \\ 1 & x_{31} & x_{32} & x_{33} & x_{34} & x_{35} \\ 1 & x_{41} & x_{42} & x_{43} & x_{44} & x_{45} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{35,1} & x_{35,2} & x_{35,3} & x_{35,4} & x_{35,5} \end{bmatrix}$$

$$x = \begin{bmatrix} 1 & 74,97 & 12,69 & 7,40 & 7,83 & 67,95 \\ 1 & 74,34 & 13,34 & 7,91 & 6,18 & 68,03 \\ 1 & 74,19 & 12,03 & 7,36 & 4,96 & 74,47 \\ 1 & 74,70 & 11,83 & 6,87 & 5,57 & 73,38 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 75,01 & 13,25 & 9,28 & 5,88 & 69,61 \end{bmatrix} \tag{4.15}$$

$$x^{\top} = \begin{bmatrix} 1 & 1 & 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & x_{31} & x_{41} & \cdots & x_{35,1} \\ x_{12} & x_{22} & x_{32} & x_{42} & \cdots & x_{35,2} \\ x_{13} & x_{23} & x_{33} & x_{43} & \cdots & x_{35,3} \\ x_{14} & x_{24} & x_{34} & x_{44} & \cdots & x_{35,4} \\ x_{15} & x_{25} & x_{35} & x_{45} & \cdots & x_{35,5} \end{bmatrix}$$

$$x^\top = \begin{bmatrix} 1 & 1 & 1 & 1 & \cdots & 1 \\ 74,97 & 74,34 & 74,19 & 74,70 & \cdots & 75,01 \\ 12,69 & 13,34 & 12,03 & 11,83 & \cdots & 13,25 \\ 7,40 & 7,91 & 7,36 & 6,87 & \cdots & 9,28 \\ 7,83 & 6,18 & 4,96 & 5,57 & \cdots & 5,88 \\ 67,95 & 68,03 & 74,47 & 73,38 & \cdots & 69,61 \end{bmatrix} \quad (4.16)$$

Kemudian menghitung perkalian dari matriks $x^\top$ yang sudah diperoleh dengan matriks x sebagai berikut:

$$x^\top x = \begin{bmatrix} 35 & 2651,93 & 458,01 & 291,71 & 154,09 & 2580,81 \\ 2651,93 & 200996,08 & 34736,27 & 22153,96 & 11654,28 & 195505,59 \\ 458,01 & 34736,27 & 6022,80 & 3857,15 & 2017,50 & 33717,10 \\ 291,71 & 22153,96 & 3857,15 & 2495,29 & 1278,94 & 21433,36 \\ 154,09 & 11654,28 & 2017,50 & 1278,94 & 738,45 & 11295,46 \\ 2580,81 & 195505,59 & 33717,10 & 21433,36 & 11295,46 & 190688,61 \end{bmatrix} \quad (4.17)$$

Karena sudah diketahui bahwa $\Sigma_0 = I$, maka Σ_0^{-1} untuk matriks x berukuran 6×6 adalah:

$$\Sigma_0^{-1} = (I_6)^{-1} = I_6 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (4.18)$$

Selanjutnya menghitung penjumlahan dari matriks $x^\top x$ dengan matriks I_6 sebagai berikut:

$$x^\top x + I_6 = \begin{bmatrix} 36 & 2651{,}93 & 458{,}01 & 291{,}71 & 154{,}09 & 2580{,}81 \\ 2651{,}93 & 200997{,}08 & 34736{,}27 & 22153{,}96 & 11654{,}28 & 195505{,}59 \\ 458{,}01 & 34736{,}27 & 6023{,}81 & 3857{,}15 & 2017{,}49 & 33717{,}10 \\ 291{,}71 & 22153{,}96 & 3857{,}15 & 2496{,}29 & 1278{,}94 & 21433{,}36 \\ 154{,}09 & 11654{,}28 & 2017{,}49 & 1278{,}94 & 739{,}45 & 11295{,}46 \\ 2580{,}81 & 195505{,}59 & 33717{,}10 & 21433{,}36 & 11295{,}46 & 190689{,}61 \end{bmatrix} \tag{4.19}$$

Setelah itu, invers dihitung dari matriks $x^\top x + I_6$ untuk memperoleh matriks kovarians *posterior* sebagai berikut:

$$\Sigma_\beta^{(1)} = (x^\top x + I_6^{-1})^{-1}$$

$$= \begin{bmatrix} 0{,}9971 & -0{,}0127 & 0{,}0019 & 0{,}0065 & -0{,}0054 & -0{,}0013 \\ -0{,}0127 & 0{,}0106 & -0{,}0262 & -0{,}0010 & -0{,}0032 & -0{,}0058 \\ 0{,}0019 & -0{,}0262 & 0{,}1978 & -0{,}0972 & -0{,}0167 & 0{,}0038 \\ 0{,}0065 & -0{,}0010 & -0{,}0972 & 0{,}0868 & 0{,}0175 & 0{,}0074 \\ -0{,}0054 & -0{,}0032 & -0{,}0167 & 0{,}0175 & 0{,}0202 & 0{,}0032 \\ -0{,}0013 & -0{,}0058 & 0{,}0038 & 0{,}0074 & 0{,}0032 & 0{,}0043 \end{bmatrix} \tag{4.20}$$

b). *Mean Posterior* $\mu_\beta^{(1)}$

Oleh karena semua nilai yang diperlukan untuk menghitung *mean posterior* sudah diperoleh dari perasamaan (4.7), (4.11), (4.16), (4.18) dan (4.20) maka perhitungan $\mu_\beta^{(1)} = \Sigma_\beta^{(1)} \left(x^\top y^{*(1)} + I^{-1}\mu\right)$ pada Persamaan (4.14) dapat diawali dengan menghitung perkalian antara matriks $x^\top$ dan nilai $y^{*(1)}$ sebagai berikut:

$$x^{\top}y^{*} = \begin{bmatrix} 26,474 \\ 2014,64819 \\ 353,47032 \\ 231,78299 \\ 113,58533 \\ 1948,54184 \end{bmatrix} \tag{4.21}$$

Selanjutnya menghitung perkalian antara matriks I_6 dengan μ sebagai berikut:

$$I^{-1}\mu = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{4.22}$$

Kemudian menghitung penjumlahan antara $x^{\top}y^{*(1)}$ dengan $I^{-1}\mu$ sebagai berikut:

$$x^{\top}y^{*(1)} + I^{-1}\mu = \begin{bmatrix} 26,474 \\ 2014,64819 \\ 353,47032 \\ 231,78299 \\ 113,58533 \\ 1948,54184 \end{bmatrix} \tag{4.23}$$

Setelah itu, menghitung perkalian antara matriks $\Sigma_{\beta}^{(1)}$ dengan $x^{\top}y^{*} + I^{-1}\mu$ untuk memperoleh vektor ***mean posterior*** sebagai berikut:

$$\mu_\beta^{(1)} = \Sigma_\beta^{(1)} \cdot (x^\top y^* + I^{-1}\mu) = \begin{bmatrix} -0,1571 \\ -0,1387 \\ 0,1612 \\ 0,3258 \\ 0,0932 \\ 0,0812 \end{bmatrix} \tag{4.24}$$

Setelah diketahui bahwa parameter $\beta^{(1)}$ mengikuti distribusi normal multivariat dengan vektor *mean posterior* $\mu_\beta^{(1)}$ dan matriks kovarians *posterior* $\Sigma_\beta^{(1)}$, yaitu

$$\beta^{(1)} \sim \mathcal{N}(\mu_\beta^{(1)}, \Sigma_\beta^{(1)}),$$

maka nilai $\beta^{(1)}$ dapat diambil secara acak dari distribusi *posterior* ini sebagai bagian dari proses pembangkitan dalam algoritma *Gibbs Sampling*. Untuk mempermudah pembangkitan sampel dari distribusi tersebut, digunakan dekomposisi *Cholesky*. Dengan hasil dekomposisi ini, nilai $\beta^{(1)}$ dapat dibangkitkan melalui Persamaan (2.52), yaitu:

$$\beta^{(1)} = \mu_\beta^{(1)} + L \cdot z \tag{4.25}$$

dengan $z \sim \mathcal{N}(0, I)$, yaitu vektor acak normal standar dan L adalah matriks segitiga bawah yang dapat dihitung dari $\Sigma_\beta^{(1)}$. Karena $\Sigma_\beta^{(1)}$ sudah diketahui dari Persamaan (4.20), maka dekomposisi *Cholesky* dapat langsung dilakukan untuk memperoleh matriks L berukuran 6×6, sebagai berikut:

$$L = \begin{bmatrix} \ell_{11} & 0 & 0 & 0 & 0 & 0 \\ \ell_{21} & \ell_{22} & 0 & 0 & 0 & 0 \\ \ell_{31} & \ell_{32} & \ell_{33} & 0 & 0 & 0 \\ \ell_{41} & \ell_{42} & \ell_{43} & \ell_{44} & 0 & 0 \\ \ell_{51} & \ell_{52} & \ell_{53} & \ell_{54} & \ell_{55} & 0 \\ \ell_{61} & \ell_{62} & \ell_{63} & \ell_{64} & \ell_{65} & \ell_{66} \end{bmatrix} \tag{4.26}$$

Untuk menghitung elemen-elemen $L = [\ell_{ij}]$ pada dekomposisi

Cholesky dari Persamaan (4.26), digunakan rumus umum pada Persamaan (2.54) untuk elemen diagonal utama dan Persamaan (2.55) untuk elemen di bawah diagonal utama. Proses dekomposisi *Cholesky* ini dilakukan secara bertahap untuk menghitung setiap elemen $\ell_{11}, \ell_{21}, \ell_{22}, \ldots, \ell_{66}$, dimulai dari baris pertama hingga baris keenam, dengan urutan perhitungan sebagai berikut:

- **Baris 1**

$$\ell_{11} = \sqrt{\Sigma_{11}^{(1)}} = \sqrt{0,9971} \approx 0,99854895$$

- **Baris 2**

$$\ell_{21} = \frac{\Sigma_{21}^{(1)}}{\ell_{11}} = \frac{-0,0127}{0,99854895} \approx -0,01271846$$

$$\begin{aligned} \ell_{22} &= \sqrt{\Sigma_{22}^{(1)} - \ell_{21}^2} \\ &= \sqrt{0,0106 - (-0,01271846)^2} \\ &= \sqrt{0,0106 - 0,0001617592} \\ &= \sqrt{0,0104382408} \approx 0,10216771 \end{aligned}$$

- **Baris 3**

$$\ell_{31} = \frac{\Sigma_{31}^{(1)}}{\ell_{11}} = \frac{0,0019}{0,99854895} \approx 0,00190276$$

$$\begin{aligned} \ell_{32} &= \frac{\Sigma_{32}^{(1)} - \ell_{31}\ell_{21}}{\ell_{22}} \\ &= \frac{-0{,}0262 - (0{,}00190276 \cdot (-0{,}01271846))}{0{,}10216771} \\ &= \frac{-0{,}0262 - (-0{,}0000242002)}{0{,}10216771} \\ &= \frac{-0{,}0261757998}{0{,}10216771} \approx -0{,}25620423 \end{aligned}$$

$$\begin{aligned}
\ell_{33} &= \sqrt{\Sigma_{33}^{(1)} - (\ell_{31}^2 + \ell_{32}^2)} \\
&= \sqrt{0{,}1978 - ((0{,}00190276)^2 + (-0{,}25620423)^2)} \\
&= \sqrt{0{,}1978 - (0{,}0000036205 + 0{,}0656406075)} \\
&= \sqrt{0{,}1978 - 0{,}065644228} \\
&= \sqrt{0{,}132155772} \approx 0{,}36353235
\end{aligned}$$

- **Baris 4**

$$\ell_{41} = \frac{\Sigma_{41}^{(1)}}{\ell_{11}} = \frac{0,0065}{0,99854895} \approx 0,00650945$$

$$\begin{aligned}
\ell_{42} &= \frac{\Sigma_{42}^{(1)} - \ell_{41}\ell_{21}}{\ell_{22}} \\
&= \frac{-0{,}0010 - (0{,}00650945 \cdot (-0{,}01271846))}{0{,}10216771} \\
&= \frac{-0{,}0010 - (-0{,}0000827902)}{0{,}10216771} \\
&= \frac{-0{,}0009172098}{0{,}10216771} \approx -0{,}00897749
\end{aligned}$$

$$\begin{aligned}
\ell_{43} &= \frac{\Sigma_{43}^{(1)} - (\ell_{41}\ell_{31} + \ell_{42}\ell_{32})}{\ell_{33}} \\
&= \frac{-0{,}0972 - \begin{pmatrix} 0{,}00650945 \cdot 0{,}00190276 \\ +\ (-0{,}00897749) \cdot (-0{,}25620423) \end{pmatrix}}{0{,}36353235} \\
&= \frac{-0{,}0972 - (0{,}0000123859 + 0{,}0023000709)}{0{,}36353235} \\
&= \frac{-0{,}0995124568}{0{,}36353235} \approx -0{,}27373756
\end{aligned}$$

$$\ell_{44} = \sqrt{\Sigma_{44}^{(1)} - (\ell_{41}^2 + \ell_{42}^2 + \ell_{43}^2)}$$

$$= \sqrt{0,0868 - \begin{pmatrix}(0,00650945)^2 + (-0,00897749)^2 \\ +\ (-0,27373756)^2\end{pmatrix}}$$

$$= \sqrt{0,0868 \begin{pmatrix}-\,0,0000423729 + 0,0000805953 \\ +\ 0,0749322518\end{pmatrix}}$$

$$= \sqrt{0,0868 - 0,07505522}$$

$$= \sqrt{0,01174478} \approx 0,10837334$$

- **Baris 5**

$$\ell_{51} = \frac{\Sigma_{51}^{(1)}}{\ell_{11}} = \frac{-0,0054}{0,99854895} \approx -0,00540785$$

$$\ell_{52} = \frac{\Sigma_{52}^{(1)} - \ell_{51}\ell_{21}}{\ell_{22}}$$

$$= \frac{-0{,}0032 - ((-0{,}00540785) \cdot (-0{,}01271846))}{0{,}10216771}$$

$$= \frac{-0{,}0032 - 0{,}0000687795}{0{,}10216771}$$

$$= \frac{-0{,}0032687795}{0{,}10216771} \approx -0{,}03199425$$

$$
\begin{aligned}
\ell_{53} &= \frac{\Sigma_{53}^{(1)} - (\ell_{51}\ell_{31} + \ell_{52}\ell_{32})}{\ell_{33}} \\
&= \frac{-0{,}0167 - \begin{pmatrix} (-0{,}00540785) \cdot 0{,}00190276 \\ + (-0{,}03199425) \cdot (-0{,}25620423) \end{pmatrix}}{0{,}36353235} \\
&= \frac{-0{,}0167 - ((-0{,}0000102898) + 0{,}0081970622)}{0{,}36353235} \\
&= \frac{-0{,}0167 - 0{,}0081867724}{0{,}36353235} \\
&= \frac{-0{,}0248867724}{0{,}36353235} \approx -0{,}06845821
\end{aligned}
$$

$$
\begin{aligned}
\ell_{54} &= \frac{\Sigma_{54}^{(1)} - (\ell_{51}\ell_{41} + \ell_{52}\ell_{42} + \ell_{53}\ell_{43})}{\ell_{44}} \\
&= \frac{0{,}0175 - \begin{pmatrix} (-0{,}00540785) \cdot 0{,}00650945 \\ + (-0{,}03199425) \cdot (-0{,}00897749) \\ + (-0{,}06845821) \cdot (-0{,}27373756) \end{pmatrix}}{0{,}10837334} \\
&= \frac{0{,}0175 - \begin{pmatrix} (-0{,}0000352021) + 0{,}0002872281 \\ + 0{,}0187395834 \end{pmatrix}}{0{,}10837334} \\
&= \frac{0{,}0175 - 0{,}0189916094}{0{,}10837334} \\
&= \frac{-0{,}0014916094}{0{,}10837334} \approx -0{,}01376361
\end{aligned}
$$

$$\begin{aligned}
\ell_{55} &= \sqrt{\Sigma_{55}^{(1)} - (\ell_{51}^2 + \ell_{52}^2 + \ell_{53}^2 + \ell_{54}^2)} \\
&= \sqrt{0{,}0202 - \begin{pmatrix}(-0{,}00540785)^2 + (-0{,}03199425)^2 \\ + (-0{,}06845821)^2 + (-0{,}01376361)^2\end{pmatrix}} \\
&= \sqrt{0{,}0202 - \begin{pmatrix}0{,}0000292448 + 0{,}001023632 \\ + 0{,}0046865265 + 0{,}000189437\end{pmatrix}} \\
&= \sqrt{0{,}0202 - 0{,}0059288403} \\
&= \sqrt{0{,}0142711597} \approx 0{,}11946196
\end{aligned}$$

- **Baris 6**

$$\ell_{61} = \frac{\Sigma_{61}^{(1)}}{\ell_{11}} = \frac{-0,0013}{0,99854895} \approx -0,00130189$$

$$\begin{aligned}
\ell_{62} &= \frac{\Sigma_{62}^{(1)} - \ell_{61}\ell_{21}}{\ell_{22}} \\
&= \frac{-0{,}0058 - ((-0{,}00130189) \cdot (-0{,}01271846))}{0{,}10216771} \\
&= \frac{-0{,}0058 - 0{,}000016558}{0{,}10216771} \\
&= \frac{-0{,}005816558}{0{,}10216771} \approx -0{,}05693147
\end{aligned}$$

$$
\begin{aligned}
\ell_{63} &= \frac{\Sigma_{63}^{(1)} - (\ell_{61}\ell_{31} + \ell_{62}\ell_{32})}{\ell_{33}} \\
&= \frac{0{,}0038 - \begin{pmatrix}(-0{,}00130189)\cdot 0{,}00190276 \\ +(-0{,}05693147)\cdot(-0{,}25620423)\end{pmatrix}}{0{,}36353235} \\
&= \frac{0{,}0038 - ((-0{,}0000024772) + 0{,}0145860834)}{0{,}36353235} \\
&= \frac{0{,}0038 - 0{,}0145836062}{0{,}36353235} \\
&= \frac{-0{,}0107836062}{0{,}36353235} \approx -0{,}0296634
\end{aligned}
$$

$$
\begin{aligned}
\ell_{64} &= \frac{\Sigma_{64}^{(1)} - (\ell_{61}\ell_{41} + \ell_{62}\ell_{42} + \ell_{63}\ell_{43})}{\ell_{44}} \\
&= \frac{0{,}0074 - \begin{pmatrix}(-0{,}00130189)\cdot 0{,}00650945 \\ +(-0{,}05693147)\cdot(-0{,}00897749) \\ +(-0{,}0296634)\cdot(-0{,}27373756)\end{pmatrix}}{0{,}10837334} \\
&= \frac{0{,}0074 - \begin{pmatrix}(-0{,}0000084746) + 0{,}0005111017 \\ +0{,}0081199867\end{pmatrix}}{0{,}10837334} \\
&= \frac{0{,}0074 - 0{,}0086226138}{0{,}10837334} \\
&= \frac{-0{,}0012226138}{0{,}10837334} \approx -0{,}01128151
\end{aligned}
$$

$$
\begin{aligned}
\ell_{65} &= \frac{\Sigma^{(1)}_{65} - (\ell_{61}\ell_{51} + \ell_{62}\ell_{52} + \ell_{63}\ell_{53} + \ell_{64}\ell_{54})}{\ell_{55}} \\
&= \frac{0{,}0032 - \begin{pmatrix} (-0{,}00130189)\cdot(-0{,}00540785) \\ +(-0{,}05693147)\cdot(-0{,}03199425) \\ +(-0{,}0296634)\cdot(-0{,}06845821) \\ +(-0{,}01128151)\cdot(-0{,}01376361) \end{pmatrix}}{0{,}11946196} \\
&= \frac{0{,}0032 - \begin{pmatrix} 0{,}0000070404 + 0{,}0018214797 \\ +0{,}0020307033 + 0{,}0001552743 \end{pmatrix}}{0{,}11946196} \\
&= \frac{0{,}0032 - 0{,}0040144977}{0{,}11946196} \\
&= \frac{-0{,}0008144977}{0{,}11946196} \approx -0{,}00681805
\end{aligned}
$$

$$
\begin{aligned}
\ell_{66} &= \sqrt{\Sigma^{(1)}_{66} - (\ell^2_{61} + \ell^2_{62} + \ell^2_{63} + \ell^2_{64} + \ell^2_{65})} \\
&= \sqrt{0{,}0043 - \begin{pmatrix} (-0{,}00130189)^2 + (-0{,}05693147)^2 \\ +(-0{,}0296634)^2 + (-0{,}01128151)^2 \\ +(-0{,}00681805)^2 \end{pmatrix}} \\
&= \sqrt{0{,}0043 - \begin{pmatrix} 0{,}0000016949 + 0{,}0032411923 \\ +0{,}0008799173 + 0{,}0001272725 \\ +0{,}0000464858 \end{pmatrix}} \\
&= \sqrt{0{,}0043 - 0{,}0042965628} \\
&= \sqrt{0{,}0000034372} \approx 0{,}00185397
\end{aligned}
$$

Setelah seluruh elemen diagonal utama dan di bawah diagonal utama, yaitu $\ell_{11}, \ell_{21}, \ell_{22}, \ldots, \ell_{66}$, berhasil dihitung melalui dekomposisi *Cholesky* terhadap matriks kovarians $\Sigma^{(1)}_{\beta}$, maka matriks segitiga bawah L dapat dibentuk secara lengkap. Matriks L diperoleh sebagai berikut:

$$L = \begin{bmatrix} 0{,}99855 & 0 & 0 & 0 & 0 & 0 \\ -0{,}01272 & 0{,}10217 & 0 & 0 & 0 & 0 \\ 0{,}00190 & -0{,}25620 & 0{,}36353 & 0 & 0 & 0 \\ 0{,}00651 & -0{,}00898 & -0{,}27374 & 0{,}10837 & 0 & 0 \\ -0{,}00541 & -0{,}03199 & -0{,}06846 & -0{,}01376 & 0{,}11946 & 0 \\ -0{,}00130 & -0{,}05693 & -0{,}02966 & -0{,}01128 & -0{,}00682 & 0{,}00185 \end{bmatrix} \tag{4.27}$$

Sebagai langkah verifikasi, matriks L dapat diuji dengan menghitung kembali $\Sigma_{\beta}^{(1)} = LL^{\top}$, kemudian dibandingkan dengan nilai $\Sigma_{\beta}^{(1)}$ yang telah diperoleh sebelumnya untuk memastikan kesesuaian.

$$L^{\top} = \begin{bmatrix} 0{,}99855 & -0{,}01272 & 0{,}00190 & 0{,}00651 & -0{,}00541 & -0{,}00130 \\ 0 & 0{,}10217 & -0{,}25620 & -0{,}00898 & -0{,}03199 & -0{,}05693 \\ 0 & 0 & 0{,}36353 & -0{,}27374 & -0{,}06846 & -0{,}02966 \\ 0 & 0 & 0 & 0{,}10837 & -0{,}01376 & -0{,}01128 \\ 0 & 0 & 0 & 0 & 0{,}11946 & -0{,}00682 \\ 0 & 0 & 0 & 0 & 0 & 0{,}00185 \end{bmatrix} \tag{4.28}$$

$$\Sigma_{\beta}^{(1)} = L.L^{\top} = \begin{bmatrix} 0{,}9971 & -0{,}0127 & 0{,}0019 & 0{,}0065 & -0{,}0054 & -0{,}0013 \\ -0{,}0127 & 0{,}0106 & -0{,}0262 & -0{,}0010 & -0{,}0032 & -0{,}0058 \\ 0{,}0019 & -0{,}0262 & 0{,}1978 & -0{,}0972 & -0{,}0167 & 0{,}0038 \\ 0{,}0065 & -0{,}0010 & -0{,}0972 & 0{,}0868 & 0{,}0175 & 0{,}0074 \\ -0{,}0054 & -0{,}0032 & -0{,}0167 & 0{,}0175 & 0{,}0202 & 0{,}0032 \\ -0{,}0013 & -0{,}0058 & 0{,}0038 & 0{,}0074 & 0{,}0032 & 0{,}0043 \end{bmatrix}$$

Setelah verifikasi menunjukkan kesesuaian, langkah selanjutnya adalah membangkitkan nilai $\beta^{(1)} = \mu_{\beta}^{(1)} + L.z$, dengan z adalah vektor variabel *random* yang mengikuti distribusi normal standar multivariat sebagai berikut:

$$z = \begin{bmatrix} -1,0856 \\ 0,9973 \\ 0,2830 \\ -1,5063 \\ -0,5786 \\ 1,6514 \end{bmatrix} \tag{4.29}$$

Setelah z diambil secara acak dari $z \sim \mathcal{N}(0, I)$, maka:

$$L \cdot z = \begin{bmatrix} -1,08304 \\ 0,10664 \\ -0,19184 \\ -0,27210 \\ -0,05961 \\ -0,04942 \end{bmatrix} \tag{4.30}$$

Dengan diperolehnya $L \cdot z$ dan $\mu_\beta^{(1)}$ dari perhitungan sebelumnya, maka nilai $\beta^{(1)}$ dapat ditentukan sebagai berikut:

$$\begin{aligned} \beta^{(1)} &= \mu_\beta^{(1)} + L \cdot z \\ &= \begin{bmatrix} -0{,}1571 \\ -0{,}1387 \\ 0{,}1612 \\ 0{,}3258 \\ 0{,}0932 \\ 0{,}0812 \end{bmatrix} + \begin{bmatrix} -1{,}08304 \\ 0{,}10664 \\ -0{,}19184 \\ -0{,}27210 \\ -0{,}05961 \\ -0{,}04942 \end{bmatrix} = \begin{bmatrix} -1{,}24014 \\ -0{,}03206 \\ -0{,}03064 \\ 0{,}05370 \\ 0{,}03359 \\ 0{,}03178 \end{bmatrix} \end{aligned} \tag{4.31}$$

4. Membangkitkan nilai $\gamma^{(1)}$

Setelah memperoleh nilai $\beta^{(1)}$, langkah selanjutnya adalah membangkitkan nilai $\gamma^{(1)}$, yaitu parameter *threshold* yang digunakan dalam klasifikasi kategori pada model regresi probit ordinal. Sebagaimana telah dijelaskan sebelumnya, pengkategorian variabel respon y berdasarkan nilai laten $y^{*(1)}$

mengikuti aturan pada Persamaan (4.2).

Dalam rangka menjaga konsistensi klasifikasi, nilai $\gamma_1^{(1)}$ harus dipilih sedemikian rupa sehingga tidak terjadi tumpang tindih antara observasi pada kategori $y = 1$ dan kategori $y = 2$. Dengan demikian, $\gamma_1^{(1)}$ dibangkitkan dari distribusi *uniform* yang dibatasi oleh nilai maksimum $y^{*(1)}$ dari kategori 1 dan nilai minimum $y^{*(1)}$ dari kategori 2 seperti pada Persamaan (2.57):

$$\gamma_1^{(1)} \sim \text{uniform}\left(\max\{y^{*(1)}|y=1\},\ \min\{y^{*(1)}|y=2\}\right) = \text{uniform}(-0,39; 0,812)$$

$$\gamma_1^{(1)} \approx 0,20 \tag{4.32}$$

Hal serupa diterapkan pada pembangkitan sampel $\gamma_2^{(1)}$ agar tidak terjadi kesalahan klasifikasi antara kategori $y = 2$ dan kategori $y = 3$. Oleh karena itu, $\gamma_2^{(1)}$ dibangkitkan dari distribusi *uniform* yang dibatasi oleh nilai maksimum $y^{*(1)}$ dari kategori 2 dan nilai minimum $y^{*(1)}$ dari kategori 3 seperti pada Persamaan (2.57):

$$\gamma_2^{(1)} \sim \text{uniform}\left(\max\{y^{*(1)}|y=2\},\ \min\{y^{*(1)}|y=3\}\right) = \text{uniform}(0,812; 1,227)$$

$$\gamma_2^{(1)} \approx 1,03 \tag{4.33}$$

Sebagai hasil dari proses pembangkitan ini, diperoleh nilai *threshold* dari iterasi pertama sebagai berikut:

$$\gamma^{(1)} = (0,20,\ 1,03) \tag{4.34}$$

Setelah memperoleh nilai parameter $\beta^{(1)}$ dan $\gamma^{(1)}$, langkah selanjutnya adalah membangkitkan kembali nilai variabel laten $y^{*(2)}$ dengan menggunakan $\beta^{(1)}$ dan $\gamma^{(1)}$. Proses ini kemudian dilakukan secara berulang hingga iterasi mencapai konvergen. Hasil uji konvergensi menggunakan *trace plot* sebagaimana Lampiran 19, terlihat bahwa *trace plot* tidak ada *trend* naik

atau turun yang signifikan, sehingga kondisi konvergen tercapai. Hasil estimasi parameter β dan γ berdasaran *trace plot* dengan menggunakan seluruh variabel prediktor disajikan pada Lampiran 12. Estimasi dilakukan dengan mengikuti langkah-langkah yang sama seperti pada iterasi pertama. Adapun nilai estimator dari parameter $\hat{\beta}$ dan $\hat{\gamma}$ diperoleh sebagai berikut:

Estimator parameter β

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \\ \hat{\beta}_5 \end{bmatrix} = \begin{bmatrix} 0,3462 \\ 1,5675 \\ 1,6243 \\ -0,7141 \\ 0,0732 \end{bmatrix} \tag{4.35}$$

Estimator parameter γ

$$\hat{\gamma} = \begin{bmatrix} \hat{\gamma_1} \\ \hat{\gamma_2} \end{bmatrix} = \begin{bmatrix} -4,8202 \\ 5,4702 \end{bmatrix} \tag{4.36}$$

Setelah memperoleh estimasi awal dari parameter model, yaitu koefisien regresi β dan *threshold* γ, langkah selanjutnya adalah melakukan pengujian signifikansi parameter terhadap model tersebut. Pengujian ini bertujuan untuk mengevaluasi apakah parameter-parameter yang telah diestimasi secara statistik signifikan dalam menjelaskan hubungan antara variabel prediktor dan variabel respon ordinal.

C. Uji Signifikansi Parameter Model Regresi Probit Ordinal dengan Metode *Bayesian Gibbs Sampling*

1. Uji Signifikansi Parameter Secara Simultan

Uji simultan dilakukan untuk mengetahui apakah seluruh parameter dalam model secara bersama-sama berpengaruh signifikan terhadap variabel respon. Langkah awal dilakukan dengan menghitung nilai *log-likelihood* dari model penuh dan model terbatas, sebagaimana diperoleh melalui proses estimasi

parameter pada Lampiran 6 sebagai berikut:

- $\log L(\hat{\beta}, \hat{\gamma}) = -2,1667$ (model penuh)
- $\log L(\hat{\beta_0}, \hat{\gamma_0}) = -161,1810$ (model terbatas)

Hipotesis yang diuji:

$H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$ dan $\gamma_1 = \gamma_2 = \cdots = \gamma_{J-1} = 0$
H_1 : Setidaknya terdapat satu $\beta_m \neq 0$ atau $\gamma_j \neq 0$

Dalam analisis ini, digunakan statistik uji *Likelihood Ratio Test* (G^2) yang dirumuskan pada Persamaan (2.60) sebagai berikut:

$$\begin{aligned} G^2 &= -2\left[\log L(\hat{\beta_0}, \hat{\gamma_0}) - \log L(\hat{\beta}, \hat{\gamma})\right] \\ &= -2\left[-161,1810 - (-2,1667)\right] \\ &= -2 \times -159,0143 \\ &= 318,0286 \end{aligned} \tag{4.37}$$

Perhitungan ini sesuai dengan hasil pada Lampiran 6 dengan nilai *P-value* = 0. Karena nilai dari $G^2 > \chi^2_{\alpha,v}$ dengan $\alpha = 0,10$ dan $v = 5$ yaitu $318,0286 > 9,236$ dan $P - value < \alpha$ yaitu $0 < 0,10$ maka H_0 ditolak, artinya semua parameter β berpengaruh signifikan secara simultan. Setelah itu, dilakukan uji signifikansi parameter secara parsial.

2. Uji Signifikansi Parameter Secara Parsial

Untuk menguji signifikansi masing-masing parameter $\hat{\beta}$ secara parsial, digunakan uji *Wald*. Uji ini memerlukan estimator parameter regresi $\hat{\beta}$ dan nilai *Standard Error* (SE) dari $\hat{\beta}$. Estimator parameter $\hat{\beta}$ diperoleh berdasarkan Persamaan (4.35), sedangkan nilai SE dan *P-value* masing-masing koefisien tersedia pada Lampiran 7, sebagai berikut:

Tabel 4.2. Nilai Estimator,*Standard Error (SE)* dan *P-value* untuk Uji *Wald* Model Lengkap

Estimator	Nilai Estimator	SE	P-value
$\hat{\beta}_1$	0,3462	0,7816	0,6579
$\hat{\beta}_2$	1,5675	0,7449	0,0353
$\hat{\beta}_3$	1,6243	0,8153	0,0463
$\hat{\beta}_4$	-0,7141	0,5403	0,1862
$\hat{\beta}_5$	0,0732	0,5550	0,8951

Hipotesis yang diuji adalah sebagai berikut:
$H_0 : \beta_1 = \beta_2 = \ldots = \beta_p = 0$
H_1 : Terdapat setidaknya satu $\beta_m \neq 0$
Statistik uji *Wald* dihitung berdasarkan Persamaan (2.63) berikut:

$$W = \left(\frac{\hat{\beta}}{\mathrm{SE}(\hat{\beta})} \right)^2$$

Hasil perhitungan statistik uji untuk setiap parameter $\hat{\beta}$ sebagai berikut:

1). Parameter $\hat{\beta}_1$:

$$W_1 = \left(\frac{0,3462}{0,7816} \right)^2 = 0,1963 \tag{4.38}$$

2). Parameter $\hat{\beta}_2$:

$$W_2 = \left(\frac{1,5675}{0,7449} \right)^2 = 4,4235 \tag{4.39}$$

3). Parameter $\hat{\beta}_3$:

$$W_3 = \left(\frac{1,6243}{0,8153} \right)^2 = 3,9688 \tag{4.40}$$

4). Parameter $\hat{\beta}_4$:

$$W_4 = \left(\frac{-0,7141}{0,5403}\right)^2 = 1,7468 \tag{4.41}$$

5). Parameter $\hat{\beta}_5$:

$$W_5 = \left(\frac{0,0732}{0,5550}\right)^2 = 0,0174 \tag{4.42}$$

Penentuan keputusan dilakukan dengan membandingkan nilai statistik uji terhadap batas kritis dari distribusi normal standar pada tingkat signifikansi $\alpha = 0,10$, yaitu $Z_{0,10/2} = Z_{0,05} = 1,645$ sesuai dengan perhitungan pada Lampiran 14. Interpretasi hasil uji *Wald* sebagai berikut:

1.) $W_1 = 0,1963 < 1,645$ dan $P - value = 0,6579 > 0,10$ sehingga H_0 gagal ditolak, artinya parameter $\hat{\beta}_1$ tidak berpengaruh signifikan.

2.) $W_2 = 4,4235 > 1,645$ dan $P - value = 0,0353 < 0,10$ sehingga H_0 ditolak, artinya parameter $\hat{\beta}_2$ berpengaruh signifikan.

3.) $W_3 = 3,9688 > 1,645$ dan $P - value = 0,0463 < 0,10$ sehingga H_0 ditolak, artinya parameter $\hat{\beta}_3$ berpengaruh signifikan.

4.) $W_4 = 1,7468 > 1,645$ tetapi $P - value = 0,1862 > 0,10$ sehingga H_0 gagal ditolak, artinya parameter $\hat{\beta}_4$ tidak berpengaruh signifikan.

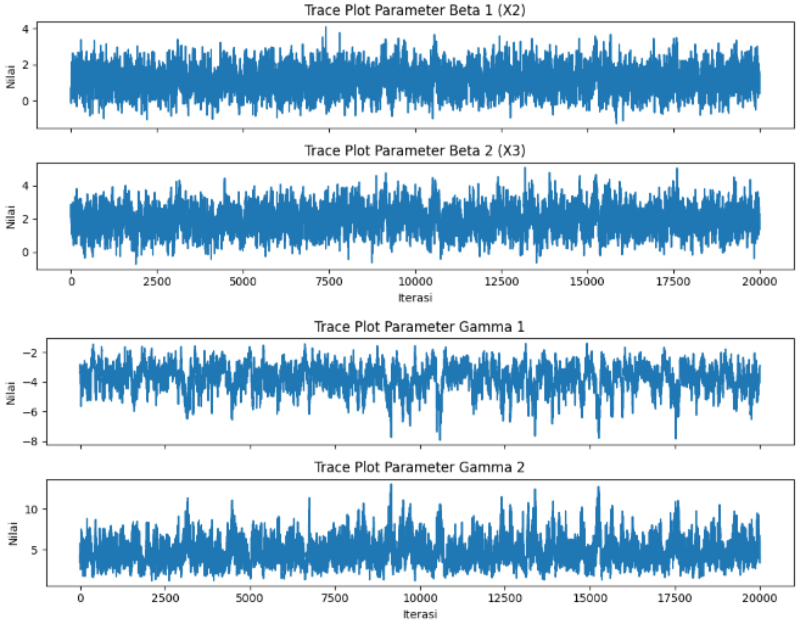
5.) $W_5 = 0,0174 < 1,645$ dan $P - value = 0,8951 > 0,10$ sehingga H_0 gagal ditolak, artinya parameter $\hat{\beta}_5$ tidak berpengaruh signifikan.

Dengan demikian, berdasarkan uji signifikansi parsial, hanya parameter $\hat{\beta}_2$ dan $\hat{\beta}_3$ yang berpengaruh signifikan terhadap variabel respon pada tingkat signifikansi 10%.

Setelah dilakukan perhitungan uji *Wald* terhadap masing-masing parameter regresi $\hat{\beta}$, diperoleh bahwa hanya $\hat{\beta}_2$ dan $\hat{\beta}_3$ yang memiliki pengaruh signifikan terhadap variabel respon. Oleh karena itu, proses estimasi parameter perlu diulang dengan memasukkan hanya variabel-variabel yang signifikan tersebut. Langkah ini bertujuan untuk membentuk model regresi probit ordinal yang lebih efisien, dengan struktur model yang lebih sederhana namun tetap mampu menangkap hubungan yang relevan antara variabel prediktor dan respon. Model hasil penyederhanaan ini diharapkan menjadi model terbaik yang dapat memberikan estimasi parameter yang lebih stabil dan interpretasi yang lebih jelas.

D. Estimasi Parameter Model Regresi Probit dengan β_2 dan β_3

Berdasarkan hasil uji signifikansi parameter menggunakan uji *Wald*, diperoleh bahwa hanya parameter $\hat{\beta}_2$ dan $\hat{\beta}_3$ yang memberikan pengaruh signifikan terhadap variabel respon. Oleh karena itu, proses pemodelan ulang dilakukan dengan mempertimbangkan kedua parameter tersebut, guna memperoleh model regresi probit ordinal yang lebih optimal dan efisien. Proses estimasi dilakukan mengikuti langkah-langkah yang sama seperti pada estimasi model lengkap dengan melibatkan lima variabel prediktor. Estimasi ini dilaksanakan secara iteratif hingga rantai Markov mencapai kondisi konvergen. Berikut adalah hasil uji konvergensi menggunakan *trace plot* sebagaimana Lampiran 13.



Gambar 4.1. *Trace Plot* Model Terbaik

Berdasarkan Gambar 4.2. terlihat bahwa *trace plot* tidak terdapat *trend* naik maupun turun yang konsisten, sehingga kondisi konvergen tercapai. Maka, hasil estimasi parameter β dan γ dari *trace plot* tersebut dengan 25.000 iterasi dan 5.000 *burn-in*, sebagaimana ditunjukkan pada Lampiran 8 sebagai berikut:

Estimator Parameter β

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} 1,1656 \\ 1,9306 \end{bmatrix} \tag{4.43}$$

Estimator Parameter γ

$$\hat{\gamma} = \begin{bmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \end{bmatrix} = \begin{bmatrix} -3,7632 \\ 4,8245 \end{bmatrix} \tag{4.44}$$

Nilai parameter $\hat{\beta}$ mengindikasikan seberapa kuat hubungan

antara variabel prediktor dengan peluang suatu observasi masuk ke dalam kategori tertentu pada variabel respon, sedangkan nilai γ berperan sebagai *threshold* pada variabel laten y^* yang memisahkan antar kategori ordinal. Selanjutnya, dilakukan pengujian signifikansi terhadap parameter $\hat{\beta}$ dari model untuk memastikan kontribusi statistik masing-masing prediktor terhadap model secara keseluruhan.

E. Uji Signifikansi Parameter Model Regresi Probit dengan β_2 dan β_3

1. Uji Signifikansi Parameter Secara Simultan

Uji simultan dengan *Likelihood Ratio Test* (G^2) dihitung berdasarkan perbedaan antara *log-likelihood* model terbatas (model null) yang hanya memuat *intercept* dan *threshold*, serta model penuh (model terbaik) yang mencakup seluruh parameter yang signifikan. Nilai *log-Likelihood* tersebut diperoleh melalui proses estimasi parameter pada Lampiran 9 sebagai berikut:

- $\log L(\hat{\beta}, \hat{\gamma}) = -3,0765$ (model terbaik)
- $\log L(\hat{\beta}_0, \hat{\gamma_0}) = -22,2944$ (model null)

Berdasarkan nilai *log-Likelihood* tersebut, diperoleh nilai statistik G^2 sebagai berikut:

$$\begin{aligned} G^2 &= -2\left[\log L(\hat{\beta}_0, \hat{\gamma_0}) - \log L(\hat{\beta}, \hat{\gamma})\right] \\ &= -2\left[-22,2944 - (-3,0765)\right] \\ &= -2 \times (-19,2179) \\ &= 38,4358 \end{aligned} \tag{4.45}$$

Hasil perhitungan tersebut sesuai dengan yang ditampilkan pada Lampiran 9, dengan *P-value* = 0. Karena nilai G^2 yang diperoleh yaitu $38,4862$ lebih besar dari nilai kritis $\chi^2_{0,10,2} = 4,605$, maka H_0 ditolak. Sehingga, dapat diketahui bahwa secara simultan,

parameter β yang digunakan dalam model memberikan pengaruh terhadap variabel respon.

2. Uji Signifikansi Parameter Secara Parsial

Sebelum melakukan uji signifikansi parameter menggunakan statistik uji *Wald*, nilai *Standard Error* (SE) dan P-value yang digunakan dalam perhitungan ini diperoleh dari hasil estimasi parameter model terbaik yang ditunjukkan pada Lampiran 8, sebagaimana dirangkum pada Tabel 4.3 berikut:

Tabel 4.3. Nilai Estimator, *Standard Error (SE)* dan *P-value* untuk Uji *Wald* Terbaik

Estimator	Nilai Estimator	SE	P-value
$\hat{\beta}_2$	1,1656	0,6694	0,0816
$\hat{\beta}_3$	1,9306	0,7472	0,0098

Berdasarkan nilai tersebut, perhitungan statistik uji *Wald* untuk setiap parameter $\hat{\beta}$ dilakukan sebagai berikut:

1). Parameter $\hat{\beta}_2$

$$W_2 = \left(\frac{\hat{\beta}_2}{\mathrm{SE}(\hat{\beta}_2)}\right)^2 = \left(\frac{1,1656}{0,6694}\right)^2 = 3,029 \qquad (4.46)$$

Karena nilai $W_2 > Z_{\alpha/2} = 1.645$ dan P-value $= 0,0816 < \alpha = 0,10$, maka H_0 ditolak. Dengan demikian, parameter $\hat{\beta}_2$ berpengaruh signifikan secara parsial terhadap model.

2). Parameter $\hat{\beta}_3$

$$W_3 = \left(\frac{\hat{\beta}_3}{\mathrm{SE}(\hat{\beta}_3)}\right)^2 = \left(\frac{1,9306}{0,7472}\right)^2 = 6,679 \qquad (4.47)$$

Karena nilai $W_3 > Z_{\alpha/2} = 1,645$ dan P-value $= 0,0098 < \alpha = 0,10$, maka H_0 ditolak. Hal ini menunjukkan bahwa parameter $\hat{\beta}_3$ juga berpengaruh signifikan secara parsial terhadap model.

Berdasarkan hasil pengujian parameter β secara parsial, diperoleh bahwa $\hat{\beta}_2$ dan $\hat{\beta}_3$ signifikan, sehingga variabel Harapan Lama Sekolah (HLS) dan Rata-rata Lama Sekolah (RLS) berpengaruh terhadap Indeks Pembangunan Manusia (IPM) di Provinsi Jawa Tengah tahun 2024.

F. Uji Kesesuaian Model

Untuk menguji apakah model yang dibangun sesuai dengan data, dilakukan uji *Pearson* $\hat{C}$. Hasil uji pada Lampiran 10 menunjukkan bahwa nilai statistik uji *Pearson* $\hat{C}$ yang diperoleh adalah 0,60 dan $\chi^2_{(\alpha,g-2)} = \chi^2_{(0,10,3-2)} = 2,7055$. Karena $\hat{C} < \chi^2_{(\alpha,g-2)} = 0,60 < 2,7055$ atau $P - value > \alpha = 0,7413 > 0,10$, maka model fit. Artinya, tidak terdapat perbedaan antara hasil observasi dengan kemungkinan hasil prediksi. Dengan kata lain, prediksi model tidak berbeda secara signifikan dari data aktual.

Ini menunjukkan bahwa model memiliki kemampuan representasi yang baik dan dapat dipercaya dalam menjelaskan keterkaitan antara variabel prediktor dengan variabel respon ordinal. Keberhasilan *Gibbs Sampling* dalam memberikan estimasi parameter yang konvergen dan hasil uji kesesuaian yang baik mengindikasikan bahwa model yang telah dibangun dinyatakan memenuhi syarat kelayakan untuk digunakan pada tahap analisis berikutnya.

G. Uji Koefisien Determinasi *McFadden* (R^2_{McF})

Dalam analisis regresi probit ordinal dengan pendekatan *Bayesian Gibbs Sampling*, salah satu ukuran yang penting untuk mengevaluasi sejauh mana model dapat menjelaskan variabel respon adalah *McFadden R-Squared*. Ukuran ini digunakan untuk

mengetahui seberapa besar pengaruh variabel prediktor terhadap variabel respon. Semakin mendekati nilai satu, *McFadden R-Squared* menunjukkan bahwa kemampuan model dalam menjelaskan perubahan variabel prediktor terhadap variabel respon semakin besar.

Berdasarkan hasil perhitungan pada Lampiran 11, diperoleh nilai *McFadden* R^2 sebesar 0,8620, artinya variabel Harapan Lama Sekolah (HLS) dan Rata-rata Lama Sekolah (RLS) memengaruhi Indeks Pembangunan Manusia (IPM) sebesar 86,20% dan 13,80% dijelaskan oleh variabel lain yang tidak dimasukkan dalam model. Karena nilai 0,8620 mendekati 1, maka kemampuan model dalam menjelaskan perubahan dari variabel HLS dan RLS terhadap variabel IPM sangat besar.

H. Interpretasi Hasil Estimasi Model Regresi Probit Ordinal

Berdasarkan hasil estimasi model regresi probit ordinal, diperoleh bahwa variabel Harapan Lama Sekolah (HLS) dan Rata-rata Lama Sekolah (RLS) berpengaruh terhadap klasifikasi Indeks Pembangunan Manusia (IPM). Model variabel laten yang terbentuk dituliskan sebagai berikut:

$$y^* = 1{,}1656\, x_2 + 1{,}9306\, x_3 + \varepsilon, \tag{4.48}$$

dengan:

$$1{,}1656\, x_2 + 1{,}9306\, x_3 = x^\top \beta \tag{4.49}$$

dimana:

- x_2 : variabel Harapan Lama Sekolah (HLS)
- x_3 : variabel Rata-rata Lama Sekolah (RLS)

variabel y^* merupakan variabel laten yang tidak teramati secara langsung, dan ε mengikuti distribusi normal standar.

Klasifikasi IPM berdasarkan nilai y^* dibagi ke dalam tiga kategori sebagai berikut:

$$\begin{aligned} y = 1: &\quad y^* \leq -3{,}7632, \\ y = 2: &\quad -3{,}7632 < y^* \leq 4{,}8245, \\ y = 3: &\quad y^* > 4{,}8245. \end{aligned} \tag{4.50}$$

Sehingga, model probabilitas dari masing-masing kategori IPM dirumuskan sebagai berikut:

$$\begin{aligned} \hat{P}(y=1) &= \Phi(\gamma_1 - x^\top \beta) = \Phi(-3{,}7632 - (1{,}1656\, x_2 + 1{,}9306\, x_3)), \\ \hat{P}(y=2) &= \Phi(\gamma_2 - x^\top \beta) - \Phi(\gamma_1 - x^\top \beta) \\ &= \Phi(4{,}8245 - (1{,}1656\, x_2 + 1{,}9306\, x_3)) - \Phi(-3{,}7632 \\ &\quad - (1{,}1656\, x_2 + 1{,}9306\, x_3)), \\ \hat{P}(y=3) &= 1 - \Phi(\gamma_2 - x^\top \beta) = 1 - \Phi(4{,}8245 - (1{,}1656\, x_2 + 1{,}9306\, x_3)) \end{aligned} \tag{4.51}$$

Ketika variabel HLS sebesar $-0,43288466$ dan variabel RLS sebesar $-0,67931612$ seperti pada data hasil standarisasi di Lampiran 6, maka nilai $x^\top \beta$ berdasaran Persamaan (4.49) dapat dihitung:

$$\begin{aligned} x^\top \beta &= 1{,}1656\, x_2 + 1{,}9306\, x_3 \\ &= 1{,}1656 \cdot (-0,43288466) + 1{,}9306 \cdot (-0,67931612) \\ &= -0,504308 + (-1,311765) \\ &= -1,816073 \end{aligned}$$

sehingga, probabilitas masing-masing kategori IPM sebagai berikut:

$$
\begin{aligned}
\hat{P}(y=1) &= \Phi(\gamma_1 - x^\top \beta) \\
&= \Phi(-3{,}7632 - (-1,816073)) \\
&= \Phi(-1,947127) \\
&= 0,0258 \\
\hat{P}(y=2) &= \Phi(\gamma_2 - x^\top \beta) - \Phi(\gamma_1 - x^\top \beta) \\
&= \Phi(4{,}8245 - (-1,816073)) - \Phi(-3{,}7632 - (-1,816073)) \\
&= \Phi(6,640573) - \Phi(-1,947127) \\
&= 1 - 0,0258 \\
&= 0,9742 \\
\hat{P}(y=3) &= 1 - \Phi(\gamma_2 - x^\top \beta) \\
&= 1 - \Phi(4{,}8245 - (-1,816073)) \\
&= 1 - \Phi(6,640573) \\
&= 1 - 1 = 0
\end{aligned}
$$

dengan $y = 1$ untuk kategori sedang, $y = 2$ untuk kategori tinggi dan $y = 3$ untuk kategori sangat tinggi.

Hasil probabilitas menunjukkan bahwa peluang suatu kabupaten/kota di Jawa Tengah berada pada kategori IPM sedang sebesar 0,0258, tinggi sebesar 0,9742, dan sangat tinggi mendekati nol. Hal ini mengindikasikan bahwa kabupaten/kota tersebut cenderung berada pada kategori IPM tinggi.

Berdasarkan hasil model regresi probit ordinal pada Persamaan (4.48), dapat disimpulkan bahwa jika Harapan Lama Sekolah (HLS) meningkat satu satuan, maka nilai variabel laten y^* akan bertambah sebesar 1,1656. Peningkatan ini akan meningkatkan peluang suatu daerah berada pada kategori IPM yang lebih tinggi. Hal yang sama berlaku untuk Rata-rata Lama Sekolah (RLS). Setiap kenaikan satu satuan pada RLS akan menaikkan variabel laten y^* sebesar 1,9306, sehingga juga meningkatkan kemungkinan daerah tersebut masuk ke kategori IPM yang lebih tinggi.

BAB V
PENUTUP

A. Kesimpulan

Berdasarkan hasil analisis yang telah dilakukan menggunakan model regresi probit ordinal dengan pendekatan *Bayesian* melalui algoritma *Gibbs Sampling*, maka dapat disimpulkan:

1. Proses estimasi parameter menghasilkan estimasi model regresi probit ordinal sebagai berikut:

$$y^* = 1{,}1656\, x_2 + 1{,}9306\, x_3 + \varepsilon,$$

dengan fungsi probabilitas tiap kategori adalah:

$$\begin{aligned}
\hat{P}(y=1) &= \Phi(\gamma_1 - x^\top \beta) = \Phi(-3{,}7632 - (1{,}1656\, x_2 + 1{,}9306\, x_3)),\\
\hat{P}(y=2) &= \Phi(\gamma_2 - x^\top \beta) - \Phi(\gamma_1 - x^\top \beta)\\
&= \Phi(4{,}8245 - (1{,}1656\, x_2 + 1{,}9306\, x_3)) - \Phi(-3{,}7632\\
&\quad - (1{,}1656\, x_2 + 1{,}9306\, x_3)),\\
\hat{P}(y=3) &= 1 - \Phi(\gamma_2 - x^\top \beta) = 1 - \Phi(4{,}8245 - (1{,}1656\, x_2\\
&\quad + 1{,}9306\, x_3))
\end{aligned}$$

Model tersebut menunjukkan bahwa koefisien regresi untuk variabel x_2 (Harapan Lama Sekolah) dan x_3 (Rata-rata Lama Sekolah) bernilai positif. Hal ini mengindikasikan bahwa peningkatan kedua variabel tersebut berkontribusi dalam meningkatkan probabilitas suatu wilayah untuk berada pada kategori Indeks Pembangunan Manusia (IPM) yang lebih tinggi.

2. Berdasarkan hasil uji signifikansi parameter secara parsial, diperoleh bahwa faktor yang berpengaruh terhadap IPM Provinsi Jawa Tengah tahun 2024 adalah Harapan Lama Sekolah (HLS) dan Rata-rata Lama Sekolah (RLS).

B. Saran

Bagi instansi pemerintah Provinsi Jawa Tengah, khususnya yang bergerak di bidang perencanaan pembangunan daerah, disarankan untuk memberikan perhatian lebih pada peningkatan kualitas dan aksesibilitas pendidikan, mengingat Harapan Lama Sekolah (HLS) dan Rata-rata Lama Sekolah (RLS) terbukti berpengaruh signifikan terhadap Indeks Pembangunan Manusia (IPM). Penelitian ini menggunakan pendekatan *Bayesian* dengan algoritma *Gibbs Sampling* untuk estimasi parameter regresi. Kedepannya, disarankan untuk mengeksplorasi metode sampling lain seperti *Metropolis-Hastings* atau *Hamiltonian Monte Carlo*, serta membandingkan pendekatan *Bayesian* dengan pendekatan frekuentis agar hasil analisis lebih komprehensif. Selain itu, penambahan variabel lain seperti tingkat kemiskinan, pengeluaran per kapita, atau akses kesehatan dapat memberikan gambaran yang lebih menyeluruh terhadap IPM. Evaluasi konvergensi juga dapat diperkuat dengan diagnostik tambahan seperti *Gelman-Rubin* atau ACF untuk memastikan kestabilan hasil estimasi.

DAFTAR PUSTAKA

Agresti, A., 2007. *An Introduction to Categorical Data Analysis Second Edition (2nd ed.)*. New York: A John Wiley & Sons, Inc., Publication.

Agresti, A., 2013. *Categorical Data Analysis*. New York: A John Wiley & Sons, Inc., Publication.

Agresti, A., 2018. *Statistical Methods for the Social Sciences (5th ed.)*. Pearson.

Ainul, A. M., Junaidi. & Utami, I. T. 2018. Penerapan Model Analisis Regresi Linier Berganda dengan Pendekatan Bayesian pada Data Aset Bank di Indonesia. *Jurnal Keteknikan dan Sains*. 1(1): 41 - 47.

Albert, J. H. 1992. Bayesian Estimation of Normal Ogive Item Response Curves Using Gibbs Sampling. *Journal of Educational Statistics*. 17(3): 251–269.

Albert, J. H. & Chib, S. 1993. Bayesian Analysis of Binary and Polychotomous Response Data. *Journal of the American Statistical Association*. 88(422): 669–679.

Anderson, T. W., 2003. *An Introduction to Multivariate Statistical Analysis (3rd ed.)*. Wiley.

Anifa, Mukid M. A. & Rusgiyono. A. 2012. Simulasi Stokastik Menggunakan Algoritma *Gibbs Sampling*. *Jurnal Gaussian*. 1(1): 21 - 30.

Badan Pusat Statistik Provinsi Jawa Tengah. 2024. Indeks Pembangunan Manusia (IPM) Provinsi Jawa Tengah 2024.

Badan Pusat Statistik Indonesia. 2024. Indeks Pembangunan Manusia (IPM) 2024.

Bain, L. J. & Engelhardt, M. 1992. *Introduction Probability and Mathematical Statistics Second Edition*. California: Duxburry Press.

Berger, J. O. 1985. *Satatistical Decision Theory and Bayesian Analiysis*. New York: Springer.

Bishop, C. M. 2006. *Pattern Recognition and Machine Learning*. New York: Springer.

Bliss, C.I. 1934. The Method of Probits. *American Association for the Advancement of Science*. LXXVIX(2039), 38-39.

Bonaraja, A. 2021. *Pembangunan Berkelanjutan dan Indeks Pembangunan Manusia*. Yogyakarta: Penerbit Deepublish.

Box, G. E. P. & Tiao, G. C. 1973. *Bayesian Inference In Statistical Analysis*. Philippines: Addision-Wesley Publishing Company.

Burkardt, J. 2014. *The Truncated Normal Distribution*. Department of Scientific Computing, Florida State University, Tallahassee.

Casella, G., & Berger, R. L. 2002. *Statistical Inference (2nd ed.)*. Duxbury.

Casella, G., & Berger, R. L. 2021. *Statistical Inference (3rd ed.)*. Cambridge University Press.

Casella, G. & George, E.I. 1992. Explaining Gibbs Sampler. *Journal of The AmericaStatistical Association*. 46(3): 167-174.

Carlin, B. P. & Chib, S. 1995. Bayesian Model Choice via Markov Chain Monte Carlo Methods. *Journal of The Royal Statistical Society*. 57(3): 473-484.

Devroye, L. 1986. *Non-Uniform Random Variate Generation*. Springer-Verlag.

Diana, E. N. 2016. *Pendekatan Metode Bayesian untuk Kajian Estimasi Parameter Distribusi Log-Normal untuk Non-Informatif Prior*. Tugas Akhir. Surabaya: Fakultas Matematika dan Ilmu Pengetahuan Alam Institut Teknologi Sepuluh Nopember Surabaya.

Drapper, N. R. & Smith, H. 1992. *Analisis Regresi Terapan* Jakarta: Gramedia Pustaka Utama.

Faelassuffa, A. & Yuliani, E. 2021. Kajian Tingkat Partisipasi Angkatan Kerja Terhadap Indeks Pembangunan Manusia. *Jurnal Kajian Ruang*. 1(1): 49 - 61.

Fikhri, M., Yanuar, F. & Asdi, Y. 2014. Pendugaan Parameter dari Distribusi Poissson dengan Menggunakan Metode Maximum Likelihood Estimation (MLE) dan Metode Bayes . *Jurnal Matematika UNAND*. 3(4): 152 - 159.

Fox, J. P., 2012. *Bayesian Item Response Modeling: Theory and Applications*. Springer.

Gaol, R. I. L., Suharianto, J., Sianturi, R. & Siagian, Y. 2024. Pengaruh Angka Harapan Lama Sekolah, Rata-Rata Lama Sekolah dan Pengeluaran per Kapita Terhadap Indeks Pembangunan Manusia di Provinsi Sumatra Utara Tahun 2010-2022. *Jurnal Penelitian dan Pengabdian Masyarakat Indonesia*. 3(1): 470 - 478.

Gelfand, A. E. & Smith, A. F. M. 1990. Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*. 85(410): 398 – 409.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. 2004. *Bayesian Data Analysis (2nd ed.)*. Washington: Chapman & Hall.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. 2013. *Bayesian Data Analysis (3rd ed.)*. CRC Press.

Ghozali, I., & Ratmono, D. 2017. *Analisis Multivariat dan Ekonometrika (Teori, Konsep, dan Aplikasi dengan EViews 10), Edisi Kedua*. Yogyakarta: Badan Penerbit Universitas Dipenegoro.

Ginting, D. I. & Lubis, I. 2023. Pengaruh Angka Harapan Hidup dan Harapan Lama Sekolah Terhadap Indeks Pembangunan Manusia. *Jurnal Bisnis Net*. 6(2): 519 - 528.

Golub, G. H. & Van Loan, C. F. 2013. *Matrix Computations (4th ed.)*. Baltimore: Johns Hopkins University.

Greene, W. H. 2008. *Econometric Analysis (6th ed.)*. New Jersey: PretinceHall, Inc.

Greene, W. H. 2012. *Econometric Analysis (7th ed.)*. New York: Pearson Education Limited.

Greene, W. H. 2018. *Econometric Analysis (8th ed.)*. Pearson.

Gujarati, D. N. 2004. *Basic Econometric (4th ed.)*. New York: Mc Graw Hill.

Harinaldi. 2012. *Statistik untuk Teknik dan Sains*. Jakarta: PT Gramedia Pustaka Utama.

Haslinda, Nusrang, M. & Aidid, M. K. 2020. Model Regresi Probit Data Panel pada Indeks Pembangunan Manusia Kabupaten/Kota di Provinsi Sulawesi Selatan. *Journal of Statistics and its Application on Teaching and Research*. 1(2): 1 - 10.

HDR,H. 2014. *Sustaining Human Progress : Reducing Vulnerabilities and Building Resilience*.New York, United State of America: United Nations Development Programme (UNDP).

Herdiansyah, D. & Kurniati, P. S. 2020. Pembangunan Sektor Pendidikan Sebagai Penunjang Indeks Pembangunan

Manusia Di Kota Bandung. *Jurnal Agregasi : Aksi Reformasi Government dalam Demokrasi*. 8(1): 43 - 50.

Hocking, R. 1996. *Methods and Application of Linear Models*. New York: John Wiley & Sons.

Hosmer, D. W., & Lemeshow, S. 1989. *Applied Logistic Regression*.New York: John Wiley.

Irwanti, L. K, Mukid, M. A. & Rahmawati, R. 2012. Pembangkit Sampel Random Menggunakan Algoritma Metropolis-Hastings. *Jurnal Gaussian*. 1(1): 135 - 146.

Kruschke, J. K. 2014. *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan (2nd ed.)*. Academic Press

Kuncoro, M. 2004. *Otonomi Daerah dan Pembangunan Daerah: Reformasi, Perencanaan, Strategi, dan Peluang*. Jakarta: Erlangga.

Kusrini, Luthfi, E. T. 2009. *Algoritma Data Mining*. Surabaya: Andi Offset.

Lancaster. T. 2004. *An Introduction to Modern Bayesian Econometrics*. New Jersey: Wiley-Blackwell.

Liu, H. & Zhang, Z. 2017. Logistic Regression with Misclassification in Binary Outcome Variables: A. Method and Software. *Behaviormetrika* (Vol. 44).

Long, J. S., & Freese, J. 2006. *Regression Models for Categorical Dependent Variables Using Stata (2nd ed.)*. Stata Press.

Mahfiyah, A. & Agoestanto, A. 2021. Pemodelan Multilevel Survival dengan Bayesian Markov Chain Monte Carlo pada Penyakit DBD. *UNNES Journal of Mathematics*. 10(2): 1 - 11.

Manurung, E. N. & Hutabarat, F. 2021. Pengaruh Angka Harapan Lama Sekolah, Rata-Rata Lama Sekolah, Pengeluaran Per

Kapita Terhadap Indeks Pembangunan Manusia. *Jurnal Ilmiah Akuntansi Manajemen*. 4(2): 121-129.

Maslihah, S. & Nisa, E. K. 2024. Pemodelan Regresi Logistik Firth Penalized Maximum Likelihood Esimation pada Kepemilikan Tempat Usaha. *AKSIOMA: Jurnal Matematika dan Pendidikan Matematika*. 15(3): 326-339.

Maumere, D. 2016. *Pemodelan Indeks Pembangunan Manusia (IPM) Provinsi Jawa Timur dengan Menggunakan Metode Regresi Logistik Ridge*. Surabaya: Statistika FMIPA ITS.

McCullagh, P. 1980. Regression Models for Ordinal Data. *Journal of the Royal Statistical Society*. 42(2): 109-142.

Nisa, E. K. & Miasary, S. D. 2024. *Parameter Estimation and Application of Inverse Gaussian Regression*. Semarang: AIP Publishing.

Ntzoufras, I. 2014. *Bayesian Modeling Using WinBUGS*. Hoboken, NJ: John Wiley & Sons.

Nurlaila, D., Kusnandar, D. & Sulistianingsih, E. 2013. Perbandingan Metode Maximum Likelihood Estimation (MSE) dan Metode Bayes dalam Pendugaan Parameter Distribusi Eksponensial. *Buletin Ilmiah Mat. Stat. dan Terapannya (Bimaster)*. 2(1):. 51 – 56.

Pamungkas, C. A. & Widiyanto W. W. 2022. Klasifikasi Indeks Pembangunan Manusia di Indonesia Tahun 2022 dengan Support Vector Machine. *Jurnal ilmiah Sistem Informasi dan Ilmu Komputer*. 2(3): 139 - 145.

Rambe, R. C., Prihanto, P. H. & Hardiani. 2019. Analisis Faktor-Faktor yang Memengaruhi Pengangguran Terbuka di Provinsi Jambi. *Jurnal Ekonomi Sumberdaya dan Lingkungan*. 8(1): 54 - 67.

Riani. W. 2006. Pembangunan Pendidikan Sebagai Motor Penggerak IPM Jawa Barat. *Jurnal Sosial dan Pembangunan*. 22(3): 278 - 291.

Robert, C. P. & Casella, G. 2004. *Monte Carlo Statistical Methods (2nd ed.)*. Springer.

Robert, C. P. 2007. *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation (2nd ed.)*. Springer.

Robert, C. P. 2009. *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation (2nd ed.)*. Springer.

Rue, H. & Held. 2005. *Gaussian Markov Random Fields: Theory and Applications*. Chapman & Hall/CRC.

Sangkereng, W., Engka, D. S. M. & Sumual, J. I. 2019. Faktor-Faktor yang Mempengaruhi Indeks Pembangunan Manusia di Provinsi Sulawesi Utara. *Jurnal Berkala Ilmiah Efisiensi*. 19(4): 60 - 71.

Sari, R. A. & Sugiharti, R. R. 2022. Analisis Faktor-Faktor yang Mempengaruhi Tingkat Partisipasi Angkatan Kerja di Indonesia Tahun 2001-2020. *Jurnal Ilmu Ekonomi dan Pembangunan*. 5(2): 603 - 616.

Septadianti. A. T. 2016. *Pendekatan Model Probit Biner Bivariat Pada Data Penolong Kelahiran dan Partisipasi Kerja di Provinsi Papua Barat*. Tesis. Surabaya: Fakultas Matematika dan Ilmu Pengetahuan Alam Institut Teknologi Sepuluh Nopember Surabaya.

Siswati, E. & Hermawati, D. T. 2018. Analisis Indeks Pembangunan Manusia (IPM) Kabupaten Bojonegoro. *Jurnal Ilmiah Sosio Agribis*. 18(2): 93 - 114.

Soejoeti, Z. & Soebanar. 1988. *InferensiBayesian*. Jakarta: Karunika Universitas Terbuka.

Sujarweni, V. W. 2015. *Metode Penelitian: Lengkap, Praktis, dan Mudah Dipahami*. Yogyakarta: Pustaka Baru Press.

Sukirno, S. 2016. *Teori Pengantar Makroekonomi*.J akarta: PT Raja Grafindo Persada.

Syafitri, F., Goejantoro, R. & Wasono. 2021. Regresi Logistik dengan Metode Bayes untuk Pemodelan Indeks Pembangunan Manusia Kabupaten/Kota di Pulau Kalimantan. *Jurnal Eksponensial*. 12(2): 103 - 110.

Syahid, I. 2022. *Analisis Faktor-Faktor yang Memengaruhi Indeks Pembangunan Manusia di Provinsi Papua Tahun 2018 - 2020*. Jakarta: Fakultas Ekonomi dan Bisnis Universitas Islam Negeri Syarif Hidayatullah Jakarta.

Taboga, M., 2021. *Log-likelihood, Lectures on Probability Theory and Mathematical Statistics*. Kindle Direct Publishing.

Tinungki, G. M. 2010. Aplikasi Model Regresi Logit dan Probit pada Data Kategorik. *Jurnal Matematika, Statistika & Komputasi*. 6(2): 107 - 114.

Tyas, A. A. W. P. & Ikhsani, K. T. W. 2015. Sumber Daya Alam dan Sumber Daya Manusia untuk Pembangunan Ekonomi Indonesia. *Jurnal Ekonomi Sumberdaya dan Lingkungan*. 12(1): 1 - 15.

UNDP. 2019. *Human Development Report: Beyond Income, Beyond Averages, Beyond Today: Inequalities in Human Development in the 21st Century. United Nations Development Programme.*. New York.

Wardani, I. N. 2020. Estimasi Parameter Model Regresi Probit Multinomial dengan Metode Maximum Likelihood. *Jurnal Ilmiah Matematika*. 8(2): 209 - 215.

Yektiningsih, E. 2018. Analisis Indeks Pembangunan Manusia (IPM) Kabupaten Pacitan Tahun 2018. *Jurnal Ilmiah Sosio Agribis*. 18(2): 32 - 50.

Lampiran 1. Data Penelitian IPM Menurut Kabupaten/Kota di Provinsi Jawa Tengah Tahun 2024 dan Faktor yang Diduga Berpengaruh

No.	Kabupaten/ Kota	Y (Persen)	X1 (Tahun)	X2 (Tahun)	X3 (Tahun)	X4 (Persen)	X5 (Persen)
1.	Cilacap	72,55	74,97	12,69	7,40	7,83	67,95
2.	Banyumas	74,52	74,34	13,34	7,91	6,18	68,03
3.	Purbalingga	70,97	74,19	12,03	7,36	4,96	74,47
4.	Banjarnegara	69,62	74,70	11,83	6,87	5,57	73,38
5.	Kebumen	72,48	75,22	13,39	7,87	5,07	77,37
6.	Purworejo	75,16	75,64	13,55	8,65	3,89	73,72
7.	Wonosobo	70,63	74,25	11,81	6,90	4,02	74,97
8.	Magelang	72,10	74,68	12,62	7,83	3,55	77,72
9.	Boyolali	75,96	76,44	12,67	8,17	3,16	75,25
10.	Klaten	78,16	77,31	13,43	9,29	3,97	74,24
11.	Sukoharjo	79,30	78,01	13,92	10,01	3,65	68,21
12.	Wonogiri	72,54	76,82	12,61	7,68	2,40	77,75
13.	Karanganyar	78,11	77,91	13,73	9,26	3,47	74,57
14.	Sragen	75,53	76,18	12,93	7,88	3,53	74,54
15.	Grobogan	72,02	75,25	12,48	7,29	3,23	76,56
16.	Blora	71,42	74,92	12,60	7,26	3,67	73,51
17.	Rembang	72,53	74,98	12,30	7,73	2,84	74,50

No.	Kabupaten/ Kota	Y (Persen)	X1 (Tahun)	X2 (Tahun)	X3 (Tahun)	X4 (Persen)	X5 (Persen)
18.	Pati	74,10	76,56	12,98	7,82	3,87	76,75
19.	Kudus	77,21	77,07	13,28	9,35	3,19	75,53
20.	Jepara	74,32	76,21	12,86	8,27	3,34	71,94
21.	Demak	74,57	75,79	13,36	8,28	4,75	72,36
22.	Semarang	75,67	76,15	13,06	8,16	3,73	79,29
23.	Temanggung	71,86	75,94	12,62	7,53	2,35	78,06
24.	Kendal	74,34	74,73	13,00	7,74	5,01	76,85
25.	Batang	70,73	75,01	12,17	7,08	5,67	77,91
26.	Pekalongan	71,95	74,25	12,46	7,48	3,30	76,38
27.	Pemalang	68,65	74,23	12,02	6,56	6,63	72,24
28.	Tegal	71,70	74,25	12,96	7,36	7,53	73,49
29.	Brebes	70,18	74,18	12,45	6,41	8,35	71,92
30.	Magelang	82,15	77,54	14,62	11,43	4,40	67,66
31.	Surakarta	84,41	77,91	15,07	11,25	4,61	67,42
32.	Salatiga	85,72	78,27	15,46	11,48	3,86	70,72
33.	Semarang	85,24	78,23	15,57	11,05	5,82	69,88
34.	Pekalongan	77,21	74,79	12,89	9,34	4,91	76,06
35.	Tegal	77,50	75,01	13,25	9,28	5,88	69,61

Sumber: Badan Pusat Statistik Provinsi Jawa Tengah

Lampiran 2. Data Penelitian Indeks Pembangunan Manusia dengan Kategori Ordinal pada Variabel Respon

No.	Kabupaten/ Kota	Kategori Y	X1 (Tahun)	X2 (Tahun)	X3 (Tahun)	X4 (Persen)	X5 (Persen)
1.	Cilacap	2	74,97	12,69	7,40	7,83	67,95
2.	Banyumas	2	74,34	13,34	7,91	6,18	68,03
3.	Purbalingga	2	74,19	12,03	7,36	4,96	74,47
4.	Banjarnegara	1	74,70	11,83	6,87	5,57	73,38
5.	Kebumen	2	75,22	13,39	7,87	5,07	77,37
6.	Purworejo	2	75,64	13,55	8,65	3,89	73,72
7.	Wonosobo	2	74,25	11,81	6,90	4,02	74,97
8.	Magelang	2	74,68	12,62	7,83	3,55	77,72
9.	Boyolali	2	76,44	12,67	8,17	3,16	75,25
10.	Klaten	2	77,31	13,43	9,29	3,97	74,24
11.	Sukoharjo	2	78,01	13,92	10,01	3,65	68,21
12.	Wonogiri	2	76,82	12,61	7,68	2,40	77,75
13.	Karanganyar	2	77,91	13,73	9,26	3,47	74,57
14.	Sragen	2	76,18	12,93	7,88	3,53	74,54
15.	Grobogan	2	75,25	12,48	7,29	3,23	76,56
16.	Blora	2	74,92	12,60	7,26	3,67	73,51
17.	Rembang	2	74,98	12,30	7,73	2,84	74,50

No.	Kabupaten/ Kota	Kategori Y	X1 (Tahun)	X2 (Tahun)	X3 (Tahun)	X4 (Persen)	X5 (Persen)
18.	Pati	2	76,56	12,98	7,82	3,87	76,75
19.	Kudus	2	77,07	13,28	9,35	3,19	75,53
20.	Jepara	2	76,21	12,86	8,27	3,34	71,94
21.	Demak	2	75,79	13,36	8,28	4,75	72,36
22.	Semarang	2	76,15	13,06	8,16	3,73	79,29
23.	Temanggung	2	75,94	12,62	7,53	2,35	78,06
24.	Kendal	2	74,73	13,00	7,74	5,01	76,85
25.	Batang	2	75,01	12,17	7,08	5,67	77,91
26.	Pekalongan	2	74,25	12,46	7,48	3,30	76,38
27.	Pemalang	1	74,23	12,02	6,56	6,63	72,24
28.	Tegal	2	74,25	12,96	7,36	7,53	73,49
29.	Brebes	1	74,18	12,45	6,41	8,35	71,92
30.	Magelang	3	77,54	14,62	11,43	4,40	67,66
31.	Surakarta	3	77,91	15,07	11,25	4,61	67,42
32.	Salatiga	3	78,27	15,46	11,48	3,86	70,72
33.	Semarang	3	78,23	15,57	11,05	5,82	69,88
34.	Pekalongan	2	74,79	12,89	9,34	4,91	76,06
35.	Tegal	2	75,01	13,25	9,28	5,88	69,61

Lampiran 3. Uji Multikolinieritas pada Data Penelitian

```
# DataFrame
df = pd.DataFrame(data)

# Matriks X (tanpa Y)
X = df[["X1", "X2", "X3", "X4", "X5"]]
X_const = add_constant(X)

# Menghitung VIF
vif_data = pd.DataFrame()
vif_data["Variable"] = X_const.columns
vif_data["VIF"] = [variance_inflation_factor(X_const.values, i) for i in range(X_const.shape[1])]

print(vif_data)
```

	Variable	VIF
0	const	12966.389031
1	X1	4.368856
2	X2	8.943120
3	X3	9.125712
4	X4	2.499289
5	X5	2.188210

Lampiran 4. Tahap Preprocessing Data

```
# --- PREPROCESSING DATA ---
Y = df["Y"].values.astype(int)
X = df[["X1", "X2", "X3", "X4", "X5"]].values

# Standarisasi (mean = 0, std = 1)
X = (X - X.mean(axis=0)) / X.std(axis=0)
print("\nX (standardized):")
print(X)

n, p = X.shape
print(f"\nJumlah data observasi (n): {n}")
print(f"Jumlah varibel prediktor (p): {p}")
```

```
X (standardized):
[[-0.60623147 -0.43288466 -0.67931612  2.24724682 -1.74198924]
 [-1.08398     0.27765834 -0.30308923  1.12711473 -1.71790961]
 [-1.19772965 -1.15435908 -0.70882411  0.29889585  0.22050061]
 [-0.81098084 -1.37298769 -1.07029701  0.71300529 -0.10758435]
 [-0.41664872  0.33231549 -0.33259722  0.37357133  1.09338721]
 [-0.0981497   0.50721838  0.24280862 -0.42749284 -0.00524592]
 [-1.15222979 -1.39485056 -1.04816602 -0.33924     0.3709983 ]
 [-0.82614746 -0.50940467 -0.36210521 -0.65830793  1.19873559]
 [ 0.5085151  -0.45474752 -0.11128728 -0.92306643  0.45527701]
 [ 1.16826308  0.37604122  0.71493648 -0.3731834   0.15127168]
 [ 1.69909478  0.91168132  1.24608033 -0.59042114 -1.66373045]
 [ 0.79668089 -0.5203361  -0.47276018 -1.43900606  1.20776545]
 [ 1.62326168  0.70398414  0.69280549 -0.71261737  0.25060015]
 [ 0.31134904 -0.17053032 -0.32522022 -0.67188529  0.24157029]
 [-0.39389879 -0.6624447  -0.7604631  -0.87554567  0.84958095]
 [-0.64414802 -0.53126753 -0.78259409 -0.57684378 -0.06845495]
 [-0.59864816 -0.85921045 -0.43587519 -1.14030417  0.22953047]
 [ 0.59951483 -0.11587317 -0.36948221 -0.44107019  0.90677007]
 [ 0.98626364  0.21206976  0.75919847 -0.90270039  0.53955571]
 [ 0.33409897 -0.24705033 -0.0375173  -0.8008702  -0.54101769]
 [ 0.01559995  0.2995212  -0.03014031  0.15633359 -0.41459963]
 [ 0.28859911 -0.02842172 -0.11866428 -0.5361117   1.67129833]
 [ 0.1293496  -0.50940467 -0.58341515 -1.47294945  1.30107402]
 [-0.78823091 -0.0940103  -0.42849819  0.33283925  0.93686961]
```

```
 [-0.57589823 -1.00131905 -0.91538005  0.78089209  1.25592471]
 [-1.15222979 -0.68430756 -0.62030014 -0.82802492  0.79540178]
 [-1.16739641 -1.16529051 -1.29898395  1.4326053  -0.45071908]
 [-1.15222979 -0.13773603 -0.70882411  2.04358644 -0.07447486]
 [-1.20531296 -0.69523899 -1.40963891  2.60025815 -0.5470376 ]
 [ 1.34267921  1.67688147  2.29361403 -0.08127019 -1.8292779 ]
 [ 1.62326168  2.16879585  2.16082807  0.06129208 -1.90151679]
 [ 1.89626084  2.59512164  2.33049902 -0.44785887 -0.90823205]
 [ 1.8659276   2.71536738  2.01328811  0.88272228 -1.16106817]
 [-0.74273105 -0.21425604  0.75182147  0.26495246  0.69908326]
 [-0.57589823  0.17927546  0.70755948  0.92345435 -1.24233692]]

Jumlah data observasi (n): 35
Jumlah varibel prediktor (p): 5
```

Lampiran 5. Estimasi Parameter Regresi Probit Ordinal *Bayesian* dengan *Gibbs Sampling* pada IPM Provinsi Jawa Tengah Tahun 2024

```
# --- GIBBS SAMPLING ---
np.random.seed(42)
n_iter = 25000
burn = 5000
beta_samples = np.zeros((n_iter, p))
thres_samples = np.zeros((n_iter, 2))

beta = np.zeros(p)
threshold = np.array([0.0, 1.0])

def truncated_normal(mean, sd, lower, upper, size=None):
    a, b = (lower - mean) / sd, (upper - mean) / sd
    return truncnorm.rvs(a, b, loc=mean, scale=sd, size=size)

for it in range(n_iter):
    mu = X @ beta
    Y_star = np.zeros(n)
    for i in range(n):
        if Y[i] == 1:
            Y_star[i] = truncated_normal(mu[i], 1, -np.inf, threshold[0])
        elif Y[i] == 2:
            Y_star[i] = truncated_normal(mu[i], 1, threshold[0], threshold[1])
        else:
            Y_star[i] = truncated_normal(mu[i], 1, threshold[1], np.inf)

    V = np.linalg.inv(X.T @ X + np.eye(p))
    m = V @ X.T @ Y_star
```

```
    # Membangkitkan sampel beta menggunakan dekomposisi Cholesky
    L = np.linalg.cholesky(V)
    z = np.random.normal(size=len(m))
    beta = m + L @ z

    beta_samples[it] = beta

    # Membangkitkan threshold (sama seperti kode asli)
    Y_star1_max = np.max(Y_star[Y == 1])
    Y_star2_min = np.min(Y_star[Y == 2])
    Y_star2_max = np.max(Y_star[Y == 2])
    Y_star3_min = np.min(Y_star[Y == 3])

    if Y_star1_max < Y_star2_min and Y_star2_max < Y_star3_min:
        threshold[0] = np.random.uniform(Y_star1_max, Y_star2_min)
        threshold[1] = np.random.uniform(Y_star2_max, Y_star3_min)

    thres_samples[it] = threshold

beta_samples = beta_samples[burn:]
thres_samples = thres_samples[burn:]
```

```python
# --- PREDIKSI DAN PROBABILITAS ---
n_samples = beta_samples.shape[0]
y_pred_samples = np.zeros((n_samples, n))
for i in range(n_samples):
    Y_star_i = X @ beta_samples[i]
    y_pred_samples[i][Y_star_i < thres_samples[i, 0]] = 1
    y_pred_samples[i][(Y_star_i >= thres_samples[i, 0]) &
     (Y_star_i < thres_samples[i, 1])] = 2
    y_pred_samples[i][Y_star_i >= thres_samples[i, 1]] = 3

y_pred_mode = np.squeeze(mode(y_pred_samples, axis=0).mode).astype(int)

# --- ESTIMASI PARAMETER ---
mean_beta = beta_samples.mean(axis=0)
mean_thres = thres_samples.mean(axis=0)

# Cetak Beta 1, Beta 2, Beta 3, Beta 4, Beta 5
print("\nEstimasi Parameter Beta:")
for i, b in enumerate(mean_beta, start=1):
    print(f"  Beta {i}: {b:.4f}")

# Cetak gamma 1, gamma 2
print("\nEstimasi Parameter Gamma:")
for j, t in enumerate(mean_thres, start=1):
    print(f"  Gamma {j}: {t:.4f}")
```

```
Estimasi Parameter Beta:
  Beta 1: 0.3462
  Beta 2: 1.5675
  Beta 3: 1.6243
  Beta 4: -0.7141
  Beta 5: 0.0732

Estimasi Parameter Gamma:
  Gamma 1: -4.8202
  Gamma 2: 5.4702
```

Lampiran 6. Uji Signifikansi Parameter Secara Simultan dengan *Likelihood Ratio Test*

```
# --- FUNGSI LOG-LIKELIHOOD ---
def log_likelihood(y_true, y_prob):
    eps = 1e-10
    return np.sum(np.log(np.maximum(y_prob[range(len(y_true)), y_true - 1], eps)))

# --- LOG-LIKELIHOOD MODEL PENUH ---
y_prob_model = np.zeros((n, 3))
for i in range(3):
    y_prob_model[:, i] = np.mean(y_pred_samples == (i + 1), axis=0)
ll_model = log_likelihood(Y, y_prob_model)

# --- LRT ---
beta_restricted = np.zeros(p)
z_restricted = X @ beta_restricted
cp_mean = thres_samples.mean(axis=0)
y_prob_restricted = np.zeros((n, 3))
y_prob_restricted[:, 0] = (z_restricted < cp_mean[0])
y_prob_restricted[:, 1] = (z_restricted >= cp_mean[0]) & (z_restricted < cp_mean[1])
y_prob_restricted[:, 2] = (z_restricted >= cp_mean[1])
ll_restricted = log_likelihood(Y, y_prob_restricted)

lrt_stat = 2 * (ll_model - ll_restricted)
p_lrt = 1 - chi2.cdf(lrt_stat, df=p)

print(f"\nLog-Likelihood Model Penuh     : {ll_model:.4f}")
print(f"Log-Likelihood Model Terbatas  : {ll_restricted:.4f}")

print("\nUji Signifikansi Simultan (Likelihood Ratio Test (LRT)):")
print(f"  Statistik LRT = {lrt_stat:.2f}")
print(f"  p-value       = {p_lrt:.4f}")
```

```
Log-Likelihood Model Penuh     : -2.1667
Log-Likelihood Model Terbatas  : -161.1810

Uji Signifikansi Simultan (Likelihood Ratio Test (LRT)):
  Statistik LRT = 318.03
  p-value       = 0.0000
```

Lampiran 7. Uji Signifikansi Parameter Secara Parsial dengan Uji *Wald*

```
# --- WALD TEST ---
mu_beta = beta_samples.mean(axis=0)
cov_beta = np.cov(beta_samples, rowvar=False)
se_beta = np.sqrt(np.diag(cov_beta))
nama_var = ["X1", "X2", "X3", "X4", "X5"]

print("\nUji Signifikansi Parsial (Wald Test per-koefisien):")
for i in range(len(mu_beta)):
    wald_stat = (mu_beta[i] / se_beta[i])**2
    p_value = 1 - chi2.cdf(wald_stat, df=1)
    signif = "SIGNIFIKAN" if p_value < 0.10 else "Tidak Signifikan"
    print(f"  {nama_var[i]}:")
    print(f"    - Standard Error: {se_beta[i]:.4f}")
    print(f"    - Wald          : {wald_stat:.2f}")
    print(f"    - p-value       : {p_value:.4f} → {signif}")

# --- SELEKSI VARIABEL BERDASARKAN WALD TEST ---
alpha = 0.10  # tingkat signifikansi
signif_idx = [i for i in range(p) if (mu_beta[i] / se_beta[i])**2 >
              chi2.ppf(1 - alpha, df=1)]
signif_vars = [nama_var[i] for i in signif_idx]
print("\nVariabel yang Signifikan (p-value < 0.10):", signif_vars)
```

```
Uji Signifikansi Parsial (Wald Test per-koefisien):
  X1:
    - Standard Error: 0.7816
    - Wald          : 0.20
    - p-value       : 0.6579 → Tidak Signifikan
  X2:
    - Standard Error: 0.7449
    - Wald          : 4.43
    - p-value       : 0.0353 → SIGNIFIKAN
  X3:
    - Standard Error: 0.8153
    - Wald          : 3.97
    - p-value       : 0.0463 → SIGNIFIKAN
  X4:
    - Standard Error: 0.5403
    - Wald          : 1.75
    - p-value       : 0.1862 → Tidak Signifikan
  X5:
    - Standard Error: 0.5550
    - Wald          : 0.02
    - p-value       : 0.8951 → Tidak Signifikan
```

Lampiran 8. Estimasi Parameter Regresi Probit Ordinal *Bayesian* dengan *Gibbs Sampling* untuk Parameter yang Signifikan

```
# --- BANGUN ULANG MODEL DENGAN VARIABEL SIGNIFIKAN SAJA ---
X_signif = X[:, signif_idx]
p_signif = X_signif.shape[1]

# Gibbs Sampling Ulang
np.random.seed(42)
n_iter = 25000
burn =5000
beta_samples_new = np.zeros((n_iter, p_signif))
thres_samples_new = np.zeros((n_iter, 2))
beta = np.zeros(p_signif)
threshold = np.array([0.0, 1.0])

for it in range(n_iter):
    mu = X_signif @ beta
    Y_star = np.zeros(n)
    for i in range(n):
        if Y[i] == 1:
            Y_star[i] = truncated_normal(mu[i], 1, -np.inf, threshold[0])
        elif Y[i] == 2:
            Y_star[i] = truncated_normal(mu[i], 1, threshold[0], threshold[1])
        else:
            Y_star[i] = truncated_normal(mu[i], 1, threshold[1], np.inf)

    V = np.linalg.inv(X_signif.T @ X_signif + np.eye(p_signif))
    m = V @ X_signif.T @ Y_star

    L = np.linalg.cholesky(V)
    z = np.random.normal(size=p_signif)
    beta = m + L @ z
    beta_samples_new[it] = beta
```

```
    y1_max = np.max(Y_star[Y == 1])
    y2_min = np.min(Y_star[Y == 2])
    y2_max = np.max(Y_star[Y == 2])
    y3_min = np.min(Y_star[Y == 3])

    if y1_max < y2_min and y2_max < y3_min:
        threshold[0] = np.random.uniform(y1_max, y2_min)
        threshold[1] = np.random.uniform(y2_max, y3_min)

    thres_samples_new[it] = threshold

# Buang burn-in
beta_samples_new = beta_samples_new[burn:]
thres_samples_new = thres_samples_new[burn:]

# --- Evaluasi Model Baru ---
mu_beta_new = beta_samples_new.mean(axis=0)
cov_beta_new = np.cov(beta_samples_new, rowvar=False)
se_beta_new = np.sqrt(np.diag(cov_beta_new))

print("\nModel Terbaik (dengan variabel signifikan):")
for i in range(p_signif):
    wald_stat = (mu_beta_new[i] / se_beta_new[i])**2
    p_val = 1 - chi2.cdf(wald_stat, df=1)
    signif = "SIGNIFIKAN" if p_val < 0.10 else "Tidak Signifikan"
    print(f"  {signif_vars[i]}: Beta = {mu_beta_new[i]:.4f}, Standard Error =
    {se_beta_new[i]:.4f}, Wald = {wald_stat:.2f}, p-value = {p_val:.4f} → {signif}")

# Ambil hasil setelah burn-in
thres_samples_post = thres_samples_new[burn:]
```

```
# Hitung estimasi threshold
threshold_1_est = np.mean(thres_samples_post[:, 0])
threshold_2_est = np.mean(thres_samples_post[:, 1])

print("\nEstimasi Threshold:")
print(f"  - Gamma 1 : {threshold_1_est:.4f}")
print(f"  - Gamma 2 : {threshold_2_est:.4f}")
```

```
Variabel yang Signifikan (p-value < 0.10): ['X2', 'X3']

Model Terbaik (dengan variabel signifikan):
  X2: Beta = 1.1656, Standard Error = 0.6694, Wald = 3.03, p-value = 0.0816 → SIGNIFIKAN
  X3: Beta = 1.9306, Standard Error = 0.7472, Wald = 6.68, p-value = 0.0098 → SIGNIFIKAN

Estimasi Threshold:
  - Gamma 1 : -3.7632
  - Gamma 2 : 4.8245
```

Lampiran 9. Uji Signifikansi Parameter Secara Simultan untuk Model Terbaik

```
def log_likelihood(y_true, y_prob):
    eps = 1e-10  # untuk menghindari log(0)
    return np.sum(np.log(np.maximum(y_prob[range(len(y_true)), y_true - 1], eps)))

# --- Likelihood Model Null ---
p_null = np.bincount(Y)[1:] / len(Y)  # proporsi kelas
y_prob_null = np.tile(p_null, (n, 1))  # probabilitas sama untuk semua observasi

ll_null = log_likelihood(Y, y_prob_null)
print(f"Log-Likelihood Model Null      (L0): {ll_null:.4f}")

# --- Likelihood Model Terbaik ---
y_prob_model_new = np.zeros((n, 3))
for i in range(3):
    y_prob_model_new[:, i] = np.mean(y_pred_samples_new == (i + 1), axis=0)

ll_model_new = log_likelihood(Y, y_prob_model_new)
print(f"Log-Likelihood Model Terbaik   (LM): {ll_model_new:.4f}")

# --- LRT untuk model terbaik (rebuild) ---
lrt_stat_new = 2 * (ll_model_new - ll_null)
p_lrt_new = 1 - chi2.cdf(lrt_stat_new, df=p_signif)

print("\nLikelihood Ratio Test (LRT) untuk Model Terbaik:")
print(f"  Statistik LRT = {lrt_stat_new:.2f}")
print(f"  p-value       = {p_lrt_new:.4f}")
```

```
Likelihood Ratio Test (LRT) untuk Model Terbaik:
  Statistik LRT = 38.44
  p-value       = 0.0000
```

Lampiran 10. Uji Kesesuaian Model dengan *Pearson* $\hat{C}$

```
# --- PEARSON CHI-SQUARE GOODNESS-OF-FIT TEST ---
# Ambil prediksi probabilitas tiap kategori dari model terbaik
expected_freq = np.sum(y_prob_model_new, axis=0)
observed_freq = np.array([np.sum(Y == k) for k in [1, 2, 3]])

chi2_stat, p_value_chi2 = chisquare(f_obs=observed_freq, f_exp=expected_freq)

print("\nPearson Chi-Square Goodness-of-Fit:")
print(f"  Uji Pearson C : {chi2_stat:.2f}")
print(f"  p-value       : {p_value_chi2:.4f}")
```

```
Pearson Chi-Square Goodness-of-Fit:
  Uji Pearson C : 0.60
  p-value       : 0.7413
```

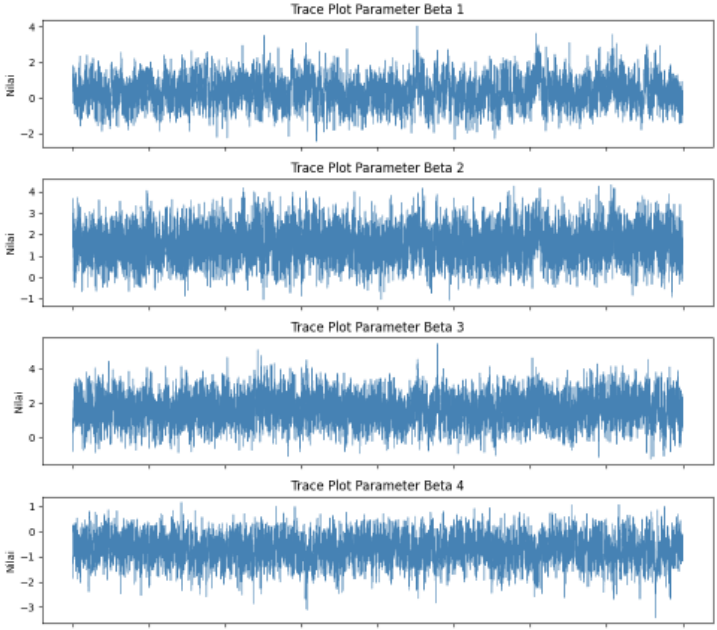
Lampiran 11. Uji Koefisien Determinasi *McFadden*

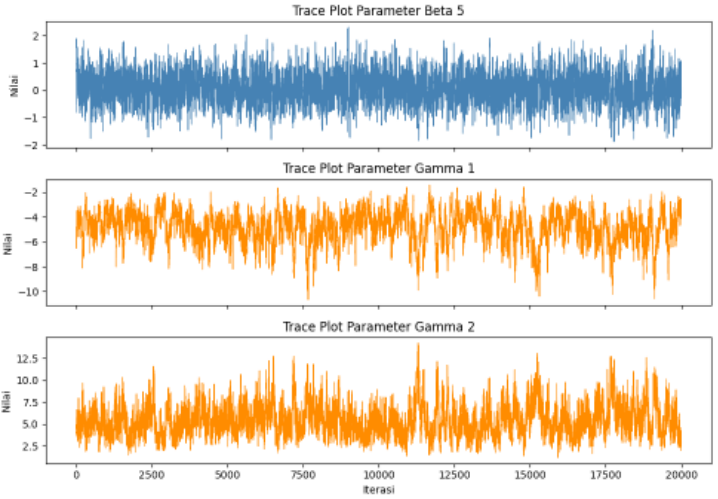
```
# --- McFADDEN R-SQUARED UNTUK MODEL TERBAIK ---

def log_likelihood(y_true, y_prob):
    eps = 1e-10  # untuk menghindari log(0)
    return np.sum(np.log(np.maximum(y_prob[range(len(y_true)), y_true - 1], eps)))

# --- Likelihood Model Null ---
p_null = np.bincount(Y)[1:] / len(Y)  # proporsi kelas
y_prob_null = np.tile(p_null, (n, 1))  # probabilitas sama untuk semua observasi

ll_null = log_likelihood(Y, y_prob_null)
print(f"Log-Likelihood Model Null      (L0): {ll_null:.4f}")

# --- Likelihood Model Terbaik ---
y_prob_model_new = np.zeros((n, 3))
for i in range(3):
    y_prob_model_new[:, i] = np.mean(y_pred_samples_new == (i + 1), axis=0)

ll_model_new = log_likelihood(Y, y_prob_model_new)
print(f"Log-Likelihood Model Terbaik   (LM): {ll_model_new:.4f}")

# --- McFadden R² ---
r2_mcfadden_new = 1 - (ll_model_new / ll_null)
print(f"McFadden R² (Model Terbaik): {r2_mcfadden_new:.4f}")
```

```
Log-Likelihood Model Null      (L0): -22.2944
Log-Likelihood Model Terbaik   (LM): -3.0765
McFadden R² (Model Terbaik): 0.8620
```

Lampiran 12. *Trace Plot* Model Penuh

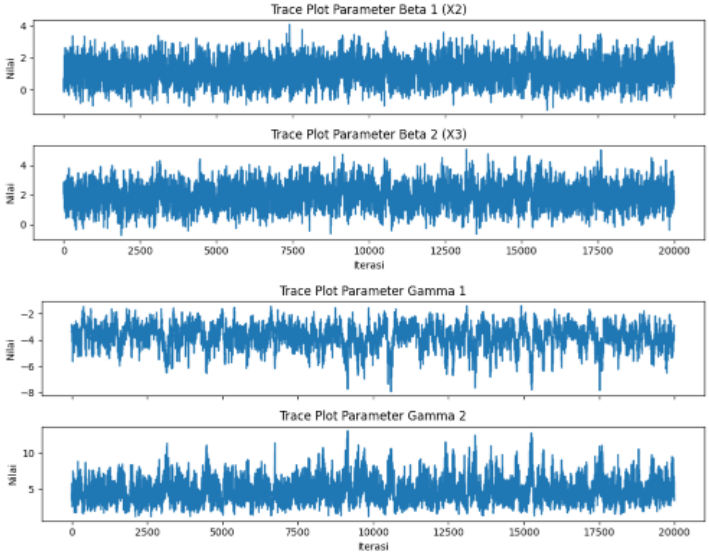
```
# --- TRACEPLOT UNTUK PARAMETER BETA DAN GAMMA ---
import matplotlib.pyplot as plt

fig, axes = plt.subplots(p + 2, 1, figsize=(10, 2.2 * (p + 2)), sharex=True)
for i in range(p):
    axes[i].plot(beta_samples[:, i], color='steelblue', linewidth=0.7)
    axes[i].set_title(f"Trace Plot Parameter Beta {i+1}")
    axes[i].set_ylabel("Nilai")

axes[p].plot(thres_samples[:, 0], color='darkorange', linewidth=0.7)
axes[p].set_title("Trace Plot Parameter Gamma 1")
axes[p].set_ylabel("Nilai")

axes[p + 1].plot(thres_samples[:, 1], color='darkorange', linewidth=0.7)
axes[p + 1].set_title("Trace Plot Parameter Gamma 2")
axes[p + 1].set_ylabel("Nilai")

plt.xlabel("Iterasi")
plt.tight_layout()
plt.show()
```


Trace Plot Parameter Beta 1
Nilai
Trace Plot Parameter Beta 2
Nilai
Trace Plot Parameter Beta 3
Nilai
Trace Plot Parameter Beta 4
Nilai


Trace Plot Parameter Beta 5
Nilai
Trace Plot Parameter Gamma 1
Nilai
Trace Plot Parameter Gamma 2
Nilai
0
2500
5000
7500
10000
12500
15000
17500
20000
Iterasi

Lampiran 13. *Trace Plot* Model Terbaik

```
# --- TRACE PLOTS UNTUK MODEL TERBAIK (HANYA VARIABEL SIGNIFIKAN) ---
fig, axes = plt.subplots(p_signif, 1, figsize=(10, 2 * p_signif), sharex=True)
for i in range(p_signif):
    axes[i].plot(beta_samples_new[:, i])
    axes[i].set_title(f"Trace Plot Parameter Beta {i+1} ({signif_vars[i]})")
    axes[i].set_ylabel("Nilai")
axes[-1].set_xlabel("Iterasi")
plt.tight_layout()
plt.show()

fig, axes = plt.subplots(2, 1, figsize=(10, 4), sharex=True)
for i in range(2):
    axes[i].plot(thres_samples_new[:, i])
    axes[i].set_title(f"Trace Plot Parameter Gamma {i+1}")
    axes[i].set_ylabel("Nilai")
axes[-1].set_xlabel("Iterasi")
plt.tight_layout()
plt.show()
```



Lampiran 14. Nilai Z pada Uji Signifikansi Parameter Secara Parsial

```
from scipy.stats import norm

# Hitung z-score untuk area di bawah kurva = 0.95
z_0_05 = norm.ppf(0.95)

print("Z_{0.05} =", round(z_0_05, 3))
```

```
Z_{0.05} = 1.645
```

DAFTAR RIWAYAT HIDUP

A. Identitas Diri

1. Nama Lengkap	:	Nelis Sa'adah
2. Tempat & Tgl. Lahir	:	Rembang, 18 Agustus 2002
3. Alamat Rumah	:	Ds. Mondoteko RT 01/RW 04, Kec. Rembang Kab. Rembang, Jawa Tengah
4. HP	:	089506477957
5. E-mail	:	2108046021@student.walisongo.ac.id nelyssaadah1808@gmail.com

B. Riwayat Pendidikan

1. SD Negeri Mondoteko Rembang (2009-2015)
2. MTs Mu'allimin Mu'allimat Rembang (2015-2018)
3. MA Mu'allimin Mu'allimat Rembang (2018-2021)
4. UIN Walisongo Semarang (2021-2025)

Semarang, 18 Juni 2025

Nelis Sa'adah
NIM: 2108046021