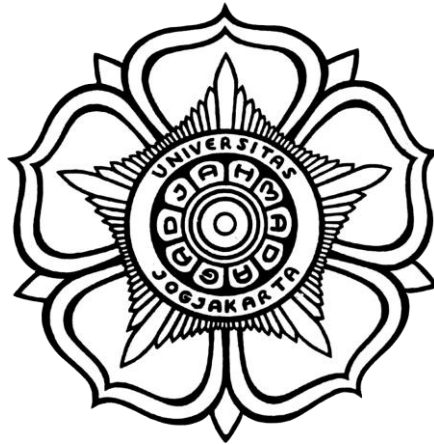


**PREDIKSI FREKUENSI KLAIM ASURANSI BPJS BERDASARKAN TINGKAT
KEPARAHAN GAGAL GINJAL MENGGUNAKAN PENDEKATAN ANALISIS
GENERAL LINEAR MODEL (GLM)**



Yogyakarta, 23 Desember 2023

Disusun oleh : 1. Twenty Yunita Anggie Hutasoit (22/492576/PA/21123)
2. Fanny Indah Oktaviani (22/494675/PA/21281)
3. Kezia Theophila (22/496970/PA/21382)
4. Zumrotul Inayah (22/498498/PA/21524)
5. Cintya Kusumawardhani (22/502898/PA/21591)
6. Febryantama Panji Christiawan (22/503447/PA/21637)

Dosen Pengampu : 1. Danang Teguh Qoyyimi, S.Si., M.Sc., Ph.D.
2. Ibu Noorma Yulia Megawati, S.Si., M.Sc

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS GADJAH MADA
YOGYAKARTA
2023**

KATA PENGANTAR

Puji syukur kami panjatkan ke hadirat Tuhan Yang Maha Esa atas segala rahmat dan hidayah-Nya sehingga kami dapat menyelesaikan laporan ini dengan baik. Laporan ini kami susun sebagai bentuk pertanggungjawaban terhadap tugas yang telah diberikan pada mata kuliah Model Linear Tergeneralisasi yang diampu oleh Bapak Danang Teguh Qoyyimi, S.Si., M.Sc., Ph.D. dan Ibu Noorma Yulia Megawati, S.Si., M.Sc.

Kami mengucapkan terima kasih yang sebesar-besarnya kepada Bapak Danang yang telah memberikan bimbingan, arahan, serta sarannya selama proses pelaksanaan *final project* ini. Kami juga mengucapkan terima kasih kepada Ibu Noorma yang telah memberikan ilmu serta materi dengan baik kepada kami sehingga kami dapat menerapkan ilmu tersebut dalam *final project* ini. Kami juga mengucapkan terima kasih kepada pihak BPJS Kesehatan yang telah memperkenankan kami untuk menggunakan data-data peserta BPJS Kesehatan. Kami juga ingin mengucapkan terima kasih kepada semua pihak yang telah membantu kelancaran pelaksanaan *final project* ini hingga tersusun laporan ini.

Laporan ini membahas tentang pembentukan model prediksi jumlah frekuensi klaim pada pasien gagal ginjal berdasarkan faktor-faktor yang merepresentasikan tingkat keparahan penyakit gagal ginjal dengan data yang diperoleh dari BPJS Kesehatan. Penulis berharap laporan ini dapat menambah wawasan pembaca mengenai pembentukan model dengan metode Analisis Poisson dan Analisis Negatif Binomial dalam kasus nyata. Akan tetapi, penulis menyadari masih banyak kekurangan dalam laporan ini. Oleh karena itu, penulis menyampaikan permohonan maaf serta terbuka untuk kritik dan saran yang membangun demi perbaikan di kemudian hari.

Yogyakarta, 23 Desember 2023

Penyusun

DAFTAR ISI

KATA PENGANTAR.....	i
DAFTAR ISI.....	ii
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah.....	3
1.3. Tujuan Penelitian	3
BAB II LANDASAN TEORI	4
2.1. Generalized Linear Model (GLM).....	4
2.2. Model Regresi Poisson Tergeneralisasi (Generalized Poisson Regression).....	5
2.3. Regresi Binomial Negatif	6
2.4. Pemilihan Model Terbaik	8
BAB III ANALISIS DATA	9
3.1. Deskripsi Data.....	9
3.2. Analisis yang Digunakan	13
3.3. Hasil Analisis	14
3.4. Pembentukan Model Akhir	29
BAB IV PENUTUP	32
4.1. Kesimpulan	32
4.2. Kritik dan Saran	32
LAMPIRAN.....	33
DAFTAR PUSTAKA.....	34

BAB I

PENDAHULUAN

1.1. Latar Belakang

Menurut Kamus Besar Bahasa Indonesia (KBBI), pengertian katastrofe adalah malapetaka besar yang datang secara tiba-tiba. Dengan kata lain, katastrofe berarti bencana yang datang mendadak. Bisa diartikan bahwa katastrofik, yang berasal dari kata “katastrofe” adalah penyakit yang “*high cost, high volume, dan high risk*” yang secara komplikasi dapat membahayakan jiwa (Riza et al., 2023). Penyakit katastrofik merupakan penyakit yang membutuhkan penanganan khusus, menggunakan alat kesehatan yang canggih, dan atau memerlukan penangan dan pelayanan kesehatan seumur hidup. Akibatnya, klaim kesehatan dari penyakit katastrofik menjadi tinggi.

Salah satu penyakit katastrofik adalah gagal ginjal. Disebut penyakit katastrofik karena bersifat kronik yang berarti membutuhkan waktu yang lama baik dalam terjangkitnya maupun dalam penyembuhannya serta membutuhkan waktu yang tidak sedikit bahkan sampai seumur hidup untuk mengendalikannya. Selain itu, penyakit ini juga degeneratif, yang berarti jenis penyakit ini akan semakin sering dijumpai seiring bertambahnya usia (Heniwati & Thabrany, 2017). Penyakit gagal ginjal merupakan penyakit pada organ ginjal dan dapat timbul karena beberapa faktor, misalnya infeksi, tumor, kelainan bawaan, penyakit metabolik, dan lain-lain. Oleh karena itu, penyakit ginjal di dunia menjadi masalah kesehatan serius. Pada umumnya sulit untuk mendeteksi secara dini penyakit ginjal. Hal ini disebabkan karena penyakit ginjal tidak menunjukkan gejala yang khas sehingga sering terlambat untuk terdeteksi. Beberapa gejala yang timbul seperti adanya darah dalam urin, bengkak pada kaki, frekuensi buang air kecil yang meningkat, dan gejala-gejala lainnya.

Berdasarkan hasil penelitian *Global Burden Disease* pada tahun 2010, ditemukan bahwa penyakit gagal ginjal kronis merupakan penyebab kematian peringkat ke-27 pada tahun 1990 dan meningkat menjadi urutan ke-18 pada tahun 2010. Lebih dari 2 juta penduduk di dunia mendapatkan perawatan dengan dialisis atau transplantasi ginjal dan hanya sekitar 10% yang mendapat tindakan tersebut. Selain itu, laporan data *Indonesia Renal Registry* menunjukkan bahwa pasien baru yang menjalani hemodialisis (prosedur perawatan dengan menyaring limbah hasil metabolisme tubuh dari darah dengan menggunakan alat

yang disebut hemodialyzer) meningkat dari 4977 pasien pada tahun 2007 menjadi 21050 pasien pada tahun 2015. Tingginya angka penduduk yang mengidap penyakit ginjal menjadi perhatian utama masyarakat kesehatan di Indonesia karena biaya penanganan dan perawatannya yang tinggi. Hal ini menjadikan pengeluaran kesehatan menjadi bersifat katastrofik (Nugraheni & Hartono, 2017).

Asuransi atau jaminan kesehatan menjadi salah satu cara untuk meringankan beban masyarakat terhadap pembiayaan kesehatan. Hal ini disebabkan karena biaya perawatan dan penanganan medis bagi penderita penyakit katastrofik tidaklah murah. Pengeluaran terus menerus terhadap pengobatan penyakit katastrofik tidak hanya menjadi masalah bagi masyarakat yang kurang berkecukupan, tetapi masyarakat yang mampu pun bisa merasakan kesulitan dalam pembiayaan kesehatan. Alasannya, karena penyakit katastrofik membutuhkan penanganan yang terus menerus dan dalam jangka waktu lama, bahkan tak jarang membutuhkan perawatan seumur hidupnya (Nugraheni & Hartono, 2017). Salah satu jaminan kesehatan di Indonesia adalah BPJS Kesehatan yang memiliki tujuan untuk mengurangi risiko masyarakat dalam menanggung pembiayaan kesehatan secara pribadi atau dari kantong sendiri. Jaminan kesehatan ini menjadi salah satu upaya untuk mengurangi risiko pribadi dan mengubahnya menjadi risiko kelompok (*sharing risiko*) sehingga biaya yang dikeluarkan dalam asuransi kesehatan merupakan hasil gotong royong dari semua masyarakat pengguna BPJS Kesehatan. Oleh karena hal tersebut, BPJS Kesehatan memiliki tujuan agar seluruh pesertanya dapat memperoleh manfaat pemeliharaan kesehatan dan perlindungan dalam memenuhi kebutuhan dasar kesehatan.

Berdasarkan data dari (BPJS Kesehatan, 2023) pada tahun 2022 mencatat bahwa klaim yang harus dibayar BPJS Kesehatan untuk sakit gagal ginjal sebesar Rp2,16 triliun dengan 1,32 juta kasus. Menurut Peraturan Badan Penyelenggara Jaminan Sosial Kesehatan Nomor 3 Tahun 2017 Tentang Pengelolaan Administrasi Klaim Fasilitas Kesehatan Dalam Penyelenggaraan Jaminan Kesehatan Nasional), fasilitas kesehatan mengajukan klaim setiap bulan secara reguler paling lambat tanggal 10 bulan berikutnya. Akan tetapi, sulit untuk memperkirakan biaya kesehatan yang akan diklaim oleh masyarakat, salah satunya bagi pengidap gagal ginjal. Apakah semakin parah gagal ginjalnya maka klaim BPJS Kesehatan akan semakin banyak? Apakah ada hal-hal lain yang mempengaruhi frekuensi klaim BPJS

terhadap penanganan penyakit gagal ginjal? Untuk menjawab pertanyaan tersebut, maka penelitian ini dilakukan.

1.2. Rumusan Masalah

1. Bagaimana tingkat keparahan gagal ginjal dapat mempengaruhi frekuensi klaim BPJS?
2. Bagaimana hubungan variabel-variabel lain dari pengidap gagal ginjal terhadap frekuensi klaim BPJS?
3. Bagaimana pengembangan model prediktif untuk memproyeksikan jumlah klaim BPJS berdasarkan tingkat keparahan tersebut?

1.3. Tujuan Penelitian

1. Mengetahui hubungan tingkat keparahan gagal ginjal terhadap frekuensi klaim BPJS.
2. Menjelaskan hubungan antara variabel-variabel yang kompleks terhadap frekuensi gagal ginjal.
3. Memprediksi jumlah klaim BPJS berdasarkan tingkat keparahan penyakit gagal ginjal.

BAB II

LANDASAN TEORI

2.1. Generalized Linear Model (GLM)

Model Linier Tergeneralisasi (MLT) atau Generalized Linear Model (GLM) merupakan sebuah metode statistika yang digunakan untuk menganalisis hubungan antara variabel dependen dengan satu atau lebih variabel independen. GLM dapat digunakan untuk berbagai jenis data, termasuk data yang tidak mengikuti distribusi normal. Metode ini sangat berguna dalam menangani berbagai jenis permasalahan di berbagai bidang, mulai dari ilmu sosial, kesehatan, hingga industri.

GLM dikembangkan sebagai perluasan dari model linier klasik. Model ini memungkinkan fleksibilitas lebih besar dengan tidak mensyaratkan distribusi normal pada variabel dependen. Secara umum, GLM terdiri dari tiga komponen utama: komponen acak, komponen sistematis, dan fungsi penghubung (link function).

- 1) Komponen acak, yaitu variabel respon Y berdistribusi keluarga eksponensial. Suatu distribusi dapat dikategorikan sebagai distribusi keluarga eksponensial apabila distribusi tersebut memiliki fungsi kepadatan probabilitas atau probability density function (pdf) yang dapat dinyatakan dalam bentuk umum

$$f(y_i|\theta_i, \phi) = \exp \left\{ \frac{y_i \cdot \theta_i - b(\theta_i)}{a_i(\phi)} + c(y_i, \phi) \right\}$$

- 2) Komponen sistematis, yakni kombinasi linier antara variabel penjelas (X_1, X_2, \dots, X_k) dengan variabel prediktor, $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$
- 3) *Link function*, η or $g(\mu)$ yang menghubungkan antara komponen acak dengan komponen sistematis. Link function menjelaskan bagaimana nilai harapan dari respons berkaitan dengan prediktor linier dari variabel-variabel penjelas.

Fungsi penghubung untuk komponen acak dengan sebaran keluarga eksponensial dapat dilihat pada tabel.

Sebaran	Penghubung	Link function	Link function inverse	Fungsi ragam
Normal	Identitas	identitas	identitas	1
Binomial/NB	Logit	$\ln[(\mu/1 - \mu)]$	$e^\eta(1 + e^\eta)$	$\mu(1 - \mu)/n$
Poisson	Logit	$\ln \mu$	e^η	μ
Gamma	Invers	$1/\mu$	$1/1/\eta$	μ^2

Dalam konteks prediksi klaim, variabel respon yang diharapkan adalah variabel respon yang bersifat *count* atau cacah, sehingga ada dua distribusi yang umum digunakan, yakni Poisson dan Negative Binomial.

2.2. Model Regresi Poisson Tergeneralisasi (Generalized Poisson Regression)

Regresi Poisson merupakan salah satu bentuk regresi nonlinier yang umum digunakan untuk menggambarkan keterkaitan antara variabel respons dalam bentuk data diskrit dan variabel prediktor yang dapat berupa data diskrit atau kontinu. *Probability Mass Function* dari distribusi Poisson yaitu:

$$P(y; \mu) = \frac{e^{-\mu} \mu^y}{y!}; \quad y = 0, 1, \dots \text{ and } \mu > 0$$

di mana μ adalah jumlah rata-rata kejadian dalam suatu interval tertentu. Nilai harapan dan varians dari distribusi Poisson sama, yaitu

$$E[y] = \text{Var}[y] = \mu$$

Maka model regresi Poisson dapat ditulis

$$y_i = \mu_i + \varepsilon_i = \exp(x_i' \beta) + \varepsilon_i; \quad i = 1, 2, 3, \dots, n$$

Model regresi poisson tergeneralisasi atau Generalized Poisson Regression (GPR) adalah perluasan dari distribusi poisson yang berguna untuk memodelkan data hitungan yang mengalami overdispersion atau underdispersion. Didefinisikan respon variabel Y yang memiliki distribusi GPR memiliki distribusi peluang

$$P(y; \mu, \phi) = \left(\frac{\mu}{1 + \phi\mu} \right)^y \frac{(1 + \phi\mu)^{y-1}}{y!} \exp \left[\frac{-\mu(1 + \phi\mu)}{1 + \phi\mu} \right]$$

dengan $y = 0, 1, 2, \dots$, dst., nilai mean μ , dan variasi $\mu(1 + \phi\mu)^2$.

Kondisi ekuidispersi terjadi ketika nilai $\phi = 0$, jadi GPR kembali ke dalam bentuk distribusi Poisson:

$$P(y; \mu, \phi = 0) = \left(\frac{\mu}{1 + 0\mu} \right)^y \frac{(1 + 0y)^{y-1}}{y!} \exp \left[\frac{-\mu(1 + 0y)}{1 + 0\mu} \right] = \frac{e^{-\mu} \mu^y}{y!}$$

Model regresi Poisson Tergeneralisasi dapat dituliskan sebagai berikut:

$$\log(\mu_i) = \mathbf{x}_i' \boldsymbol{\beta} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}$$

Dengan x_i adalah vektor berdimensi k-1 dan β adalah vektor parameter regresi berdimensi k.

Regresi Poisson Tergeneralisasi dapat digunakan untuk mengatasi overdispersion atau underdispersion. Kondisi overdispersion terjadi pada regresi Poisson dengan $\text{Var}(y) < E(y)$ jika ($\phi > 0$), sementara kondisi underdispersion terjadi jika ($\phi < 0$). Asumsi parameter dispersi dan koefisien regresi Poisson Generalized dilakukan dengan menggunakan metode asumsi kemungkinan maksimum. Fungsi probabilitas diperoleh dengan mengalikan fungsi probabilitas Poisson Generalized. Fungsi logaritma natural probabilitas dari distribusi Poisson Generalized adalah sebagai berikut:

$$l(y|\beta, \phi) = \sum_{i=1}^n y_i \ln \left(\frac{\mu_i}{1 + \phi \mu_i} \right) + [y_i - 1] \ln(1 + \phi y_i) - \frac{\mu_i(1 + \phi y_i)}{1 + \phi \mu_i} - \ln(y!)$$

Fungsi ini dapat dimaksimalkan dengan mencari turunan terhadap setiap parameter dan menyamakannya dengan nol.

2.3. Regresi Binomial Negatif

Regresi binomial negatif merupakan generalisasi dari regresi Poisson dengan melemahkan asumsi bahwa mean sama dengan varians. Distribusi binomial negatif dapat dijelaskan sebagai pendekatan yang ingin melihat besaran probabilitas dari r sukses setelah beberapa kali kejadian gagal. Diketahui r adalah banyaknya kejadian berhasil dan variabel acak Y adalah banyaknya kejadian gagal sebelum kejadian berhasil ke- r , bentuk distribusi binomial negatif dengan peluang berhasil p , dan peluang gagal $(1 - p)$ adalah

$$P(Y = y|r, p) = \binom{y-1}{r-1} p^r (1-p)^{y-r}; \quad Y = r, r+1, r+2, \dots$$

Dengan nilai harapan dan variansi

$$E[y] = r \frac{(1-p)}{p} \text{ dan } \text{Var}[Y] = r \frac{(1-p)}{p^2}$$

Model Regresi Binomial Negatif adalah salah satu pendekatan untuk mengatasi masalah overdispersi berdasarkan model campuran Poisson-Gamma. Dengan mengasumsikan adanya variabel Gamma-spreading dengan mean 1 dan varians Φ dalam rata-rata distribusi Poisson, misalnya m sebagai sumber varians yang tidak teramati. Maka, nilai rata-rata dari distribusi campuran Poisson-Gamma adalah:

$$E(y_i) = \tilde{\mu}_i = \exp(x_i^T \beta + m_i) = \exp(x_i^T \beta) \exp(m_i) = \mu_i \delta_i$$

dengan $\mu_i = \exp(x_i^T \beta)$ adalah mean dari model Poisson. Fungsi probabilitas dari gabungan distribusi Poisson-Gamma dapat dituliskan:

$$f(y_i | x_i, \beta, \delta_i) = \frac{e^{-(\mu_i \delta_i)} (\mu_i \delta_i)^{y_i}}{y_i!}$$

Fungsi probabilitas Gamma:

$$g(\delta_i) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \delta_i^{\alpha-1} e^{-\delta_i/\beta}$$

Dengan nilai harapan $E(\delta_i) = \alpha\beta$ dan asumsi $E(\delta_i) = 1$ maka $\alpha = 1/\beta$. Jika parameter $\alpha = 1/\Phi$, maka fungsi probabilitas gama menjadi:

$$g(\delta_i) = \frac{1}{\Phi^{\Phi-1} \Gamma(\Phi^{-1})} \delta_i^{\Phi^{-1}-1} e^{-\delta_i/\Phi}$$

Fungsi gabungan Poisson-Gamma bisa didapatkan dari mengintralkan variabel δ_i ke dalam fungsi probabilitas Poisson:

$$f(y_i | x_i, \beta, \Phi) = \int_0^\infty f(y_i | \Phi, \delta_i) g(\delta_i) d\delta_i = \frac{\Gamma(y_i + \Phi^{-1})}{\Gamma(\Phi^{-1}) y_i!} \left(\frac{\Phi \mu_i}{1 + \Phi \mu_i} \right)^{y_i} \left(\frac{1}{1 + \Phi \mu_i} \right)^{\Phi^{-1}}$$

dengan

$$\Phi > 0, E[y_i] = \mu_i \text{ dan } \text{Var}[y_i] = \mu_i + \Phi \mu_i^2$$

Asumsi parameter dispersi dan koefisien regresi dalam Regresi Binomial Negatif diestimasi menggunakan metode asumsi likelihood maksimum. Fungsi probabilitas dari distribusi binomial negatif sebagai berikut:

$$L(\beta, \Phi | y, x) = \prod_{i=1}^n \left\{ \frac{\Gamma(y_i + \Phi^{-1})}{\Gamma(\Phi^{-1}) y_i!} \left(\frac{\Phi \mu_i}{1 + \Phi \mu_i} \right)^{y_i} \left(\frac{1}{1 + \Phi \mu_i} \right)^{\Phi^{-1}} \right\}$$

Fungsi probabilitas logaritma naturalnya:

$$\ln L(\beta, \Phi | y, x) = \sum_{i=1}^n \left\{ \ln \left[\frac{\Gamma(y_i + \Phi^{-1})}{\Gamma(\Phi^{-1})\Gamma(y_i + 1)} \right] - (y_i + \Phi^{-1}) \ln(1 + \Phi\mu_i) + y_i \ln(\Phi\mu_i) \right\}$$

1.1. Pemilihan Model Terbaik

1. Pemenuhan asumsi ekuidispersi

Asumsi ekuidispersi merupakan asumsi bahwa nilai mean data sama dengan nilai variansinya. Syarat ekuidispersi diuji dengan menghitung hasil bagi *residual deviance* oleh derajat bebasnya. Dengan asumsi nilai mean dan variansi sama, maka model akan semakin baik jika hasil baginya semakin mendekati satu. Sedangkan jika hasil baginya lebih dari 1, maka terjadi kasus overdispersi sehingga digunakan analisis alternatif, seperti penggunaan regresi binomial negatif.

2. Aikake's Information Criteriation (AIC)

Pemilihan model terbaik antara Regresi Poisson, Regresi Binomial Negatif, dan Regresi Poisson Tergeneralisasi dapat dilakukan dengan menggunakan nilai Akaike's Information Criteriation (AIC). AIC diformulasikan sebagai berikut:

$$AIC = -2(\mathcal{L} - k)$$

di mana \mathcal{L} adalah log-likelihood model, dan k adalah jumlah parameter dalam model.

Model terbaik ditentukan oleh model yang memiliki nilai AIC terkecil.

BAB III

ANALISIS DATA

3.1. Deskripsi Data

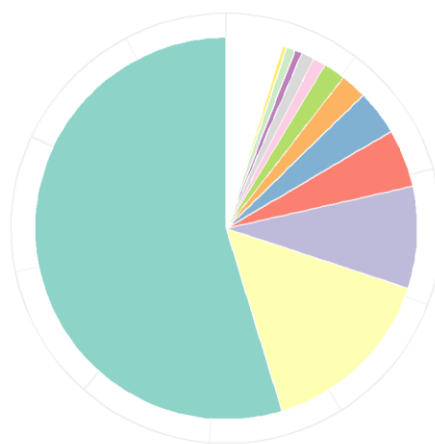
Data yang digunakan untuk analisis Prediksi Frekuensi Klaim BPJS Berdasarkan Tingkat Keparahan Gagal Ginjal menggunakan Pendekatan Analisis General Linear Model (GLM) adalah data BPJS tahun 2015-2020. Peserta yang diikuti dalam analisis adalah gabungan dari peserta FKTP Kapitasi yang dirujuk ke FKRTL, peserta FKRTL yang tidak berasal dari rujukan FKTP, dan peserta FKTP Nonkapitasi yang memiliki diagnosis akhir gagal ginjal dengan nomor kode N17, N18, dan N19.

Variabel yang digunakan dalam analisis Prediksi Frekuensi Klaim BPJS Berdasarkan Tingkat Keparahan Gagal Ginjal menggunakan Pendekatan Analisis General Linear Model (GLM) dibagi menjadi dua, yaitu variabel respons dan variabel prediktor.

a. Variabel Respons

Variabel respons adalah variabel yang nilainya ingin kita prediksi atau jelaskan dalam suatu penelitian atau analisis. Dalam analisis ini dipilih variabel respons, yaitu frekuensi klaim BPJS.

Persebaran Frekuensi Klaim



Frekuensi

1	18	43	88	162
2	19	45	90	178
3	20	47	91	185
4	21	49	93	192
5	22	55	96	194
6	23	57	99	200
7	26	58	107	204
8	28	60	108	205
9	30	62	110	206
10	31	66	115	207
11	32	68	123	208
12	35	70	124	209
13	37	71	138	211
14	39	73	140	
15	40	76	142	
16	41	77	149	
17	42	83	154	

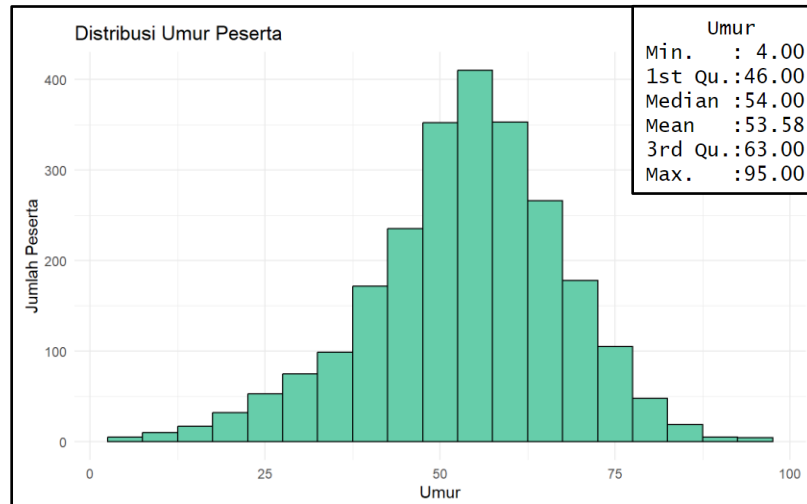
b. Variabel Prediktor

Variabel prediktor adalah variabel yang digunakan untuk memprediksi atau menjelaskan variasi dalam variabel respon. Variabel prediktor yang digunakan dalam

analisis ini adalah umur, jenis kelamin, riwayat penyakit katastropik lain selain gagal ginjal, riwayat keturunan, frekuensi rawat inap.

1. Variabel Umur

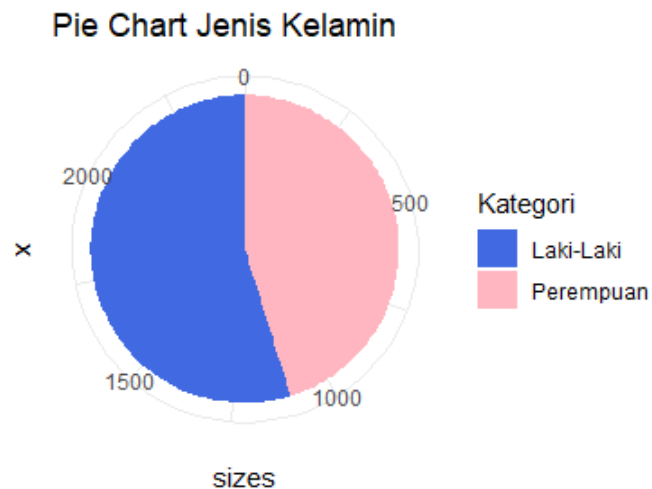
Variabel umur yang digunakan adalah umur minimum peserta yaitu umur saat kunjungan pertama. Variabel umur didapatkan dari tanggal kunjungan pertama dikurangi dengan tanggal lahir peserta.



Distribusi umur peserta yang diikuti berada dalam interval 4 - 95 tahun dengan rata-rata umur 53,5 tahun.

2. Variabel Jenis Kelamin

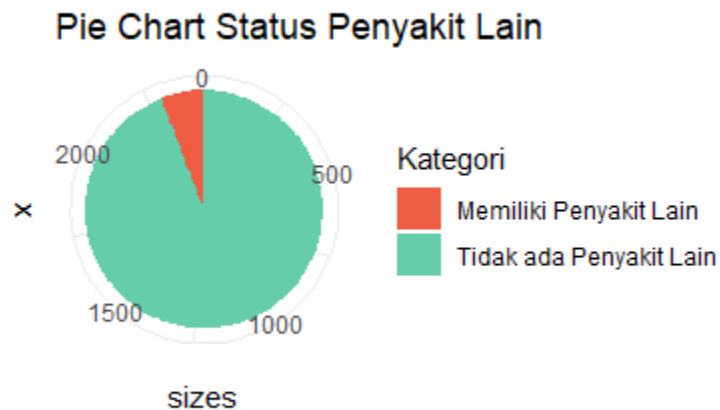
Variabel jenis kelamin didapatkan dari data kepesertaan kolom PSTV05 (Jenis Kelamin)



Distribusi jenis kelamin peserta yang diikutkan adalah 54.8% laki-laki atau sebanyak 1336 peserta laki-laki.

3. Variabel Riwayat Penyakit Katastropik Lain selain Gagal Ginjal

Variabel riwayat penyakit katastropik lain selain gagal ginjal didapatkan dari memfilter data awal dengan kode-kode penyakit katastropik selain gagal ginjal kemudian dilakukan inner join dengan nomor peserta yang memiliki penyakit gagal ginjal.

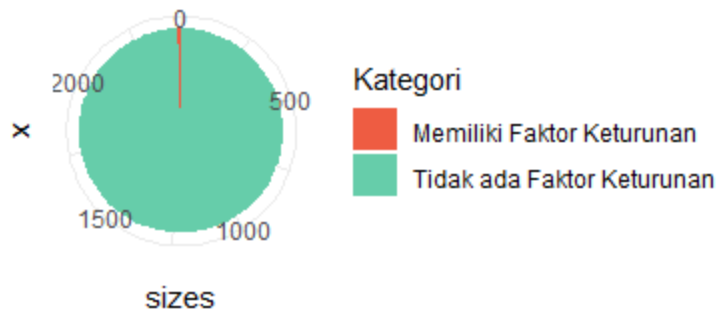


Didapatkan peserta yang memiliki penyakit gagal ginjal namun juga memiliki penyakit katastropik lain sebanyak 143 atau 5,8%.

4. Variabel Riwayat Keturunan

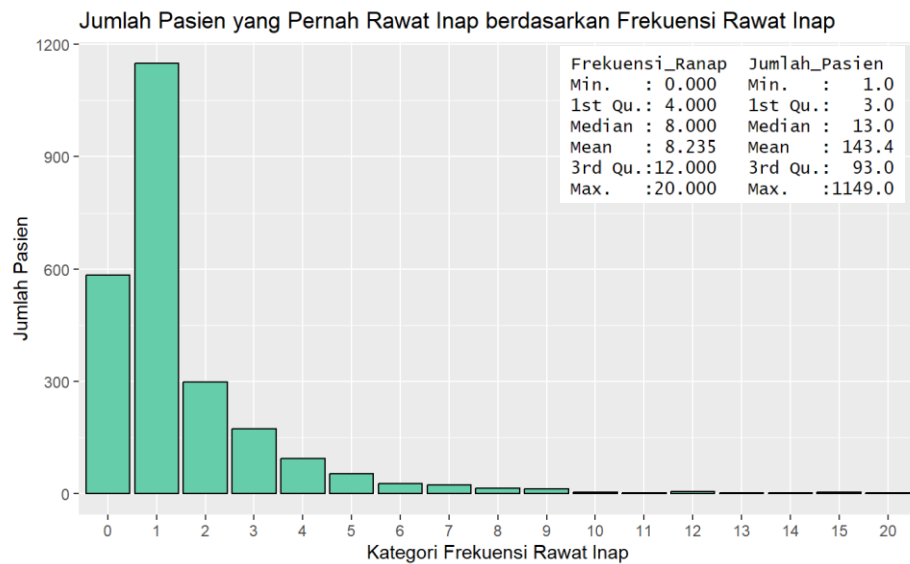
Variabel riwayat keturunan didapatkan dengan melakukan join distinct peserta dengan nomor peserta berbeda tetapi memiliki nomor keluarga yang sama.

Pie Chart Faktor Keturunan



Didapatkan peserta yang memiliki penyakit gagal ginjal dan memiliki faktor keturunan atau terdapat anggota keluarga lain yang juga memiliki gagal ginjal sebanyak 12 atau 0,5%.

5. Variabel Frekuensi Rawat Inap



Frekuensi rawat inap peserta yang diikutkan adalah antara 0-20 kali dengan frekuensi terbanyak adalah 1 kali rawat inap.

3.2. Analisis yang Digunakan

Analisis yang digunakan untuk melakukan prediksi frekuensi klaim BPJS berdasarkan tingkat keparahan gagal ginjal adalah analisis Poisson dan analisis Negatif Binomial kemudian dicari model terbaiknya.

Analisis Poisson adalah metode statistik yang digunakan untuk memodelkan frekuensi peristiwa jarang atau tidak terduga dalam suatu periode waktu tertentu. Karakteristik utama dari distribusi Poisson adalah cocok digunakan untuk memodelkan kejadian yang bersifat acak dan langka, di mana setiap kejadian terjadi secara independen dengan tingkat kejadian yang konstan dalam interval waktu atau ruang tertentu. Frekuensi klaim asuransi sering kali sesuai dengan pola distribusi Poisson di mana kejadian klaim cenderung terjadi secara jarang namun bisa menjadi lebih sering dalam interval waktu yang spesifik. Distribusi Poisson mengasumsikan bahwa varians sama dengan mean. Namun, dalam beberapa situasi, varians mungkin jauh lebih besar daripada mean atau sering disebut dengan overdispersi. Analisis Regresi Poisson tidak dapat digunakan untuk menangani data yang overdispersi sehingga untuk menanganinya dapat dilakukan Analisis Negatif Binomial.

Analisis Negatif Binomial adalah metode statistik yang digunakan untuk memodelkan data yang menunjukkan pola distribusi overdispersi, yaitu variasi yang lebih besar dari yang dapat dijelaskan oleh distribusi Poisson. Metode ini umumnya digunakan ketika data memiliki variasi yang lebih besar daripada yang diharapkan berdasarkan asumsi distribusi Poisson.

3.3. Hasil Analisis

a. Analisis *Generalized Linear Model* (GLM) Poisson

Model

```
call:
glm(formula = Frekuensi ~ Umur + sex + fams + ill + ranap + Umur *
  sex + Umur * fams + Umur * ill + Umur * ranap + sex * fams +
  sex * ill + sex * ranap + fams * ill + fams * ranap + ill *
  ranap, family = "poisson", data = dataprediksi)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-5.477  -2.024  -1.797  -1.176   35.581

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.8800284  0.0514486  36.542 < 2e-16 ***
Umur         -0.0080132  0.0009617  -8.332 < 2e-16 ***
sex2         -0.4000302  0.0706677  -5.661 1.51e-08 ***
fams1        -1.5765753  1.7979446  -0.877  0.381
ill1         1.2067895  0.1355355   8.904 < 2e-16 ***
ranap         0.2912072  0.0098725  29.497 < 2e-16 ***
Umur:sex2     0.0063788  0.0012603   5.061 4.17e-07 ***
Umur:fams1    0.0026507  0.0330290   0.080  0.936
Umur:ill1    -0.0199303  0.0025367  -7.857 3.94e-15 ***
Umur:ranap   -0.0031091  0.0002134 -14.569 < 2e-16 ***
sex2:fams1   -0.3247274  0.5411549  -0.600  0.548
sex2:ill1     0.3694440  0.0671261   5.504 3.72e-08 ***
sex2:ranap    0.0042955  0.0064497   0.666  0.505
fams1:ill1   -2.0844858  1.7139389  -1.216  0.224
fams1:ranap   0.4264319  0.2615411   1.630  0.103
ill1:ranap   -0.0565820  0.0098959  -5.718 1.08e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 41421  on 2437  degrees of freedom
Residual deviance: 38625  on 2422  degrees of freedom
AIC: 45058
```

● Goodness of Fit Test

Diketahui bahwa suatu model dapat dikatakan layak jika model tersebut memiliki nilai $\frac{\text{Residual Deviance}}{\text{Degree of Freedom Residual}} < 1$ atau mendekati 1. Berdasarkan keluaran dari hasil komputasi menggunakan software Rstudio di atas, diketahui nilai Residual Deviance sebesar 38625 dengan degrees of freedomnya sebesar 2422 sehingga diperoleh nilai $\frac{\text{Residual Deviance}}{\text{Degree of Freedom Residual}} = 15.95 > 1$, sehingga model tidak dapat dikatakan layak untuk digunakan. Dari besar nilai yang didapatkan, diduga data yang digunakan pada model GLM Poisson tersebut merupakan data yang overdispersi.

- **Overdispersion Test**

```
Overdispersion test  
  
data: model_pois  
z = 5.5618, p-value = 1.335e-08  
alternative hypothesis: true dispersion is greater than 1  
sample estimates:  
dispersion  
73.01547
```

- a. **Hipotesis**

H_0 : Tidak ada overdispersi, model Poisson cukup untuk menjelaskan variasi dalam data.

H_1 : Ada overdispersi, model Poisson tidak cukup untuk menjelaskan variasi dalam data.

- b. **Tingkat Signifikansi**

$\alpha = 0.05$

- c. **Statistik Uji**

P-Value = 1.335E-08

- d. **Daerah Kritik**

H_0 ditolak apabila nilai P-Value $< \alpha$

- e. **Kesimpulan**

Berdasarkan hasil keluaran Overdispersion Test menggunakan software Rstudio, diketahui nilai P-Value = 1.335E-08 $< 0.05 = \alpha$, sehingga H_0 ditolak. Dengan demikian, dapat diambil kesimpulan bahwa data yang digunakan pada model Poisson tersebut terdapat overdispersi sehingga model Poisson tidak cukup untuk menjelaskan variasi dalam data. Oleh karena itu, untuk menganalisis data tersebut, dilakukan analisis *Generalized Linear Model* (GLM) Negative Binomial.

b. Analisis *Generalized Linear Model* (GLM) Negative Binomial

Model Pertama

```
Call:
glm.nb(formula = Frekuensi ~ Umur + sex + fams + i11 + ranap +
  Umur * sex + Umur * fams + Umur * i11 + Umur * ranap + sex *
  fams + sex * i11 + sex * ranap + fams * i11 + fams * ranap +
  i11 * ranap, data = dataprediksi, init.theta = 0.6396696533,
  link = log)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.5629  -0.8431  -0.7541  -0.4190   7.8009

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.039510   0.172210  11.843  < 2e-16 ***
Umur        -0.011806   0.003107  -3.800  0.000145 ***
sex2        -0.620006   0.225733  -2.747  0.006021 **
fams1       -1.520647   3.143139  -0.484  0.628529
i111        2.200226   0.517558   4.251  2.13e-05 ***
ranap        0.282301   0.059284   4.762  1.92e-06 ***
Umur:sex2    0.010193   0.003937   2.589  0.009628 **
Umur:fams1   0.003213   0.056768   0.057  0.954862
Umur:i111   -0.039144   0.009092  -4.305  1.67e-05 ***
Umur:ranap  -0.002425   0.001136  -2.134  0.032831 *
sex2:fams1  -0.287366   1.002101  -0.287  0.774294
sex2:i111    0.477271   0.232918   2.049  0.040453 *
sex2:ranap   0.006587   0.029809   0.221  0.825105
fams1:i111  -2.303476   3.368472  -0.684  0.494080
fams1:ranap  0.351556   0.542940   0.648  0.517305
i111:ranap  -0.083256   0.051395  -1.620  0.105249
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(0.6397) family taken to be 1)

Null deviance: 2787.3 on 2437 degrees of freedom
Residual deviance: 2544.2 on 2422 degrees of freedom
AIC: 12989

Number of Fisher Scoring iterations: 1

            Theta: 0.6397
            Std. Err.: 0.0177

2 x log-likelihood: -12955.4840
```

• Uji Goodness of Fit

Diketahui bahwa suatu model dapat dikatakan layak jika model tersebut memiliki nilai $\frac{Residual\ Deviance}{Degree\ of\ Freedom\ Residual} < 1$ atau mendekati 1. Berdasarkan keluaran dari hasil komputasi menggunakan software Rstudio di atas, diketahui nilai Residual Deviance sebesar 2544.2 dengan degrees of freedomnya sebesar 2422 sehingga diperoleh nilai $\frac{Residual\ Deviance}{Degree\ of\ Freedom\ Residual} = 1.05 \approx 1$, sehingga model pertama GLM Negative Binomial dapat dikatakan layak untuk digunakan.

• Uji Parsial untuk Konstanta

a. Hipotesis

H_0 : Konstanta tidak mempengaruhi model pertama secara signifikan

H_1 : Konstanta mempengaruhi model pertama secara signifikan

b. Tingkat Signifikansi

$\alpha = 0.05$

c. Statistik Uji

$P\text{-Value} < 2E-16$

d. Daerah Kritik

H_0 ditolak apabila nilai $P\text{-Value} < \alpha$

e. Kesimpulan

Berdasarkan hasil keluaran Analisis Negative Binomial menggunakan software Rstudio, diketahui nilai $P\text{-Value}$ dari konstanta model sebesar $2E-16 < 0.05 = \alpha$, sehingga H_0 ditolak. Dengan demikian, dapat diambil kesimpulan bahwa konstanta mempengaruhi model secara signifikan.

● **Uji Parsial untuk Koefisien**

a. Hipotesis

H_0 : Koefisien variabel ke- i tidak mempengaruhi model secara signifikan, dengan $i = 1, 2, 3, \dots, 15$.

H_1 : Koefisien variabel ke- i mempengaruhi model secara signifikan, dengan $i = 1, 2, 3, \dots, 15$.

b. Tingkat Signifikansi

$\alpha = 0.05$

c. Daerah Kritik

H_0 ditolak apabila nilai $P\text{-Value} < \alpha$

d. Statistik Uji dan Kesimpulan

No	Variabel	P-Value	Kesimpulan
1.	Umur	0.000145	Berdasarkan hasil keluaran, diketahui nilai $P\text{-Value}$ dari konstanta model sebesar $0.000145 < 0.05 = \alpha$, sehingga H_0 ditolak. Dengan demikian, dapat diambil kesimpulan bahwa koefisien variabel Umur mempengaruhi model secara signifikan.

2.	sex2	0.006021	Berdasarkan hasil keluaran, diketahui nilai P-Value dari konstanta model sebesar $0.006021 < 0.05 = \alpha$, sehingga H_0 ditolak. Dengan demikian, dapat diambil kesimpulan bahwa koefisien variabel Sex2 mempengaruhi model secara signifikan.
3.	fams1	0.628529	Berdasarkan hasil keluaran, diketahui nilai P-Value dari konstanta model sebesar $0.628529 > 0.05 = \alpha$, sehingga H_0 tidak ditolak. Dengan demikian, dapat diambil kesimpulan bahwa koefisien variabel fams1 tidak mempengaruhi model secara signifikan.
4.	ill1	2.13E-05	Berdasarkan hasil keluaran, diketahui nilai P-Value dari konstanta model sebesar $2.13E-05 < 0.05 = \alpha$, sehingga H_0 ditolak. Dengan demikian, dapat diambil kesimpulan bahwa koefisien variabel ill1 mempengaruhi model secara signifikan.
5.	ranap	1.92E-06	Berdasarkan hasil keluaran, diketahui nilai P-Value dari konstanta model sebesar $1.92E-06 < 0.05 = \alpha$, sehingga H_0 ditolak. Dengan demikian, dapat diambil kesimpulan bahwa koefisien variabel ranap mempengaruhi model secara signifikan.

6.	Umur:sex2	0.009628	Berdasarkan hasil keluaran, diketahui nilai P-Value dari konstanta model sebesar $= 0.009628 < 0.05 = \alpha$, sehingga H_0 ditolak. Dengan demikian, dapat diambil kesimpulan bahwa koefisien variabel Umur:sex2 mempengaruhi model secara signifikan.
7.	Umur:fams1	0.954862	Berdasarkan hasil keluaran, diketahui nilai P-Value dari konstanta model sebesar $= 0.954862 > 0.05 = \alpha$, sehingga H_0 tidak ditolak. Dengan demikian, dapat diambil kesimpulan bahwa koefisien variabel Umur:fams1 tidak mempengaruhi model secara signifikan.
8.	Umur:ill1	1.67E-05	Berdasarkan hasil keluaran, diketahui nilai P-Value dari konstanta model sebesar $= 1.67E-05 < 0.05 = \alpha$, sehingga H_0 ditolak. Dengan demikian, dapat diambil kesimpulan bahwa koefisien variabel Umur:ill1 mempengaruhi model secara signifikan.
9.	Umur:ranap	0.032831	Berdasarkan hasil keluaran, diketahui nilai P-Value dari konstanta model sebesar $= 0.032831 < 0.05 = \alpha$, sehingga H_0 ditolak. Dengan demikian, dapat diambil kesimpulan bahwa koefisien variabel Umur:ranap mempengaruhi model secara signifikan.

10.	sex2:fams1	0.774294	Berdasarkan hasil keluaran, diketahui nilai P-Value dari konstanta model sebesar $= 0.774294 > 0.05 = \alpha$, sehingga H_0 tidak ditolak. Dengan demikian, dapat diambil kesimpulan bahwa koefisien variabel sex2:fams1 tidak mempengaruhi model secara signifikan.
11.	sex2:ill1	0.040453	Berdasarkan hasil keluaran, diketahui nilai P-Value dari konstanta model sebesar $= 0.040453 < 0.05 = \alpha$, sehingga H_0 ditolak. Dengan demikian, dapat diambil kesimpulan bahwa koefisien variabel sex2:ill1 mempengaruhi model secara signifikan.
12.	sex2:ranap	0.825105	Berdasarkan hasil keluaran, diketahui nilai P-Value dari konstanta model sebesar $= 0.825105 > 0.05 = \alpha$, sehingga H_0 tidak ditolak. Dengan demikian, dapat diambil kesimpulan bahwa koefisien variabel sex2:ranap tidak mempengaruhi model secara signifikan.
13.	fams1:ill1	0.494080	Berdasarkan hasil keluaran, diketahui nilai P-Value dari konstanta model sebesar $= 0.494080 > 0.05 = \alpha$, sehingga H_0 tidak ditolak. Dengan demikian, dapat diambil kesimpulan bahwa koefisien variabel fams1:ill1 mempengaruhi model secara signifikan.

14.	fams1:ranap	0.517305	Berdasarkan hasil keluaran, diketahui nilai P-Value dari konstanta model sebesar $= 0.517305 > 0.05 = \alpha$, sehingga H_0 tidak ditolak. Dengan demikian, dapat diambil kesimpulan bahwa koefisien variabel fams1:ranap tidak mempengaruhi model secara signifikan.
15.	ill1:ranap	0.105249	Berdasarkan hasil keluaran, diketahui nilai P-Value dari konstanta model sebesar $= 0.105249 > 0.05 = \alpha$, sehingga H_0 tidak ditolak. Dengan demikian, dapat diambil kesimpulan bahwa koefisien variabel ill1:ranap tidak mempengaruhi model secara signifikan.

Interpretasi :

Uji Goodness of Fit dilakukan untuk mengetahui apakah model sudah fit (layak) untuk digunakan, sedangkan Uji Parsial, baik untuk Konstanta maupun Koefisien, dilakukan untuk mengetahui apakah nilai konstanta dan koefisien signifikan terhadap model GLM yang didapatkan. Berdasarkan keluaran dari Analisis GLM Negative binomial di atas, diketahui bahwa, melalui Uji Goodness of Fit, model dinyatakan layak untuk digunakan. Kemudian, melalui Uji Parsial, diketahui bahwa konstanta dan 8 dari 15 koefisien variabel prediktor dinyatakan mempengaruhi model secara signifikan. Hal ini berarti masih terdapat variabel prediktor yang dinyatakan tidak mempengaruhi model yang dimiliki, sehingga dilakukan pengeluaran variabel prediktor dengan nilai P-Value terbesar dari mode. Hal ini dilakukan berulang-ulang hingga mendapatkan model GLM yang layak digunakan dengan konstanta dan seluruh koefisien variabel prediktornya signifikan mempengaruhi model. Pada kasus ini, telah dilakukan pengulangan sebanyak enam kali hingga mendapatkan model yang layak digunakan dan konstanta serta seluruh koefisien variabelnya signifikan terhadap model pada model ketujuh.

Model Kedua

```
Call:
glm.nb(formula = Frekuensi ~ Umur + sex + fams + ill + ranap +
  Umur * sex + Umur * ill + Umur * ranap + sex * fams + sex *
  ill + sex * ranap + fams * ill + fams * ranap + ill * ranap,
  data = dataprediksi, init.theta = 0.6396692069, link = log)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.5629  -0.8431  -0.7542  -0.4190   7.8004

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.039375   0.172200  11.843 < 2e-16 ***
Umur        -0.011804   0.003107  -3.799 0.000145 ***
sex2        -0.620411   0.225653  -2.749 0.005970 **
fams1       -1.352303   0.949042  -1.425 0.154182
ill1        2.200451   0.517544   4.252 2.12e-05 ***
ranap        0.282341   0.059281   4.763 1.91e-06 ***
Umur:sex2    0.010201   0.003936   2.592 0.009545 **
Umur:ill1   -0.039148   0.009091  -4.306 1.66e-05 ***
Umur:ranap  -0.002426   0.001136  -2.135 0.032776 *
sex2:fams1  -0.265749   0.937778  -0.283 0.776885
sex2:ill1    0.477267   0.232919   2.049 0.040455 *
sex2:ranap   0.006590   0.029809   0.221 0.825036
fams1:ill1  -2.433529   2.479379  -0.982 0.326343
fams1:ranap  0.358649   0.530279   0.676 0.498825
ill1:ranap  -0.083261   0.051395  -1.620 0.105229
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(0.6397) family taken to be 1)

    Null deviance: 2787.3  on 2437  degrees of freedom
Residual deviance: 2544.2  on 2423  degrees of freedom
AIC: 12987

Number of Fisher Scoring iterations: 1

            Theta:  0.6397
            Std. Err.:  0.0177

2 x log-likelihood:  -12955.4870
```

Model Ketiga

```
Call:
glm.nb(formula = Frekuensi ~ Umur + sex + fams + i11 + ranap +
        Umur * sex + Umur * i11 + Umur * ranap + sex * fams + sex *
        i11 + fams * i11 + fams * ranap + i11 * ranap, data = dataprediksi,
        init.theta = 0.6396471231, link = log)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.5624  -0.8428  -0.7555  -0.4168   7.7857

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.036037   0.170747  11.924 < 2e-16 ***
Umur        -0.011826   0.003107  -3.806 0.000141 ***
sex2        -0.603860   0.216272  -2.792 0.005236 **
fams1       -1.350597   0.949227  -1.423 0.154783
i111        2.201099   0.517290   4.255 2.09e-05 ***
ranap        0.284235   0.058625   4.848 1.24e-06 ***
Umur:sex2    0.010081   0.003920   2.572 0.010109 *
Umur:i111   -0.039128   0.009089  -4.305 1.67e-05 ***
Umur:ranap  -0.002407   0.001135  -2.121 0.033916 *
sex2:fams1  -0.268164   0.937854  -0.286 0.774929
sex2:i111    0.477923   0.232854   2.052 0.040125 *
fams1:i111  -2.433943   2.479064  -0.982 0.326199
fams1:ranap  0.357692   0.530252   0.675 0.499949
i111:ranap  -0.084350   0.051360  -1.642 0.100525
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(0.6396) family taken to be 1)

    Null deviance: 2787.2  on 2437  degrees of freedom
Residual deviance: 2544.2  on 2424  degrees of freedom
AIC: 12986

Number of Fisher Scoring iterations: 1

            Theta: 0.6396
            Std. Err.: 0.0177

2 x log-likelihood: -12955.5320
```

Model Keempat

```
Call:
glm.nb(formula = Frekuensi ~ Umur + sex + fams + i11 + ranap +
        Umur * sex + Umur * i11 + Umur * ranap + sex * i11 + fams *
        i11 + fams * ranap + i11 * ranap, data = dataprediksi, init.theta = 0.6396320603,
        link = log)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.5623  -0.8424  -0.7551  -0.4166   7.7889

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.036431   0.170739  11.927 < 2e-16 ***
Umur        -0.011825   0.003107  -3.806 0.000141 ***
sex2        -0.603329   0.216267  -2.790 0.005275 **
fams1       -1.527045   0.733028  -2.083 0.037233 *
i111        2.200257   0.517287   4.253 2.10e-05 ***
ranap        0.284178   0.058624   4.847 1.25e-06 ***
Umur:sex2    0.010054   0.003919   2.566 0.010298 *
Umur:i111   -0.039121   0.009089  -4.304 1.67e-05 ***
Umur:ranap  -0.002406   0.001135  -2.120 0.033971 *
sex2:i111    0.478868   0.232831   2.057 0.039713 *
fams1:i111  -2.457458   2.477609  -0.992 0.321263
fams1:ranap  0.397758   0.517474   0.769 0.442099
i111:ranap  -0.084332   0.051361  -1.642 0.100599
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(0.6396) family taken to be 1)

    Null deviance: 2787.1  on 2437  degrees of freedom
Residual deviance: 2544.2  on 2425  degrees of freedom
AIC: 12984

Number of Fisher Scoring iterations: 1

            Theta: 0.6396
            Std. Err.: 0.0177

2 x log-likelihood: -12955.6110
```

Model Kelima

```
Call:
glm.nb(formula = Frekuensi ~ Umur + sex + fams + i11 + ranap +
  Umur * sex + Umur * i11 + Umur * ranap + sex * i11 + fams *
  i11 + i11 * ranap, data = dataprediksi, init.theta = 0.6394931428,
  link = log)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.5620  -0.8419  -0.7538  -0.4167   7.7955

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.035474   0.170754  11.920 < 2e-16 ***
Umur         -0.011814   0.003107  -3.802 0.000143 ***
sex2         -0.601044   0.216279  -2.779 0.005452 **
fams1        -1.044399   0.448730  -2.327 0.019941 *
i111         2.201711   0.517337   4.256 2.08e-05 ***
ranap         0.283502   0.058623   4.836 1.32e-06 ***
Umur:sex2     0.009995   0.003919   2.551 0.010755 *
Umur:i111    -0.039141   0.009089  -4.306 1.66e-05 ***
Umur:ranap   -0.002383   0.001134  -2.101 0.035671 *
sex2:i111    0.479788   0.232851   2.060 0.039352 *
fams1:i111   -0.948271   1.435725  -0.660 0.508944
i111:ranap   -0.084846   0.051363  -1.652 0.098561 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(0.6395) family taken to be 1)

Null deviance: 2786.6  on 2437  degrees of freedom
Residual deviance: 2544.4  on 2426  degrees of freedom
AIC: 12982

Number of Fisher Scoring iterations: 1

            Theta: 0.6395
            Std. Err.: 0.0177

2 x log-likelihood: -12956.3470
```

Model Keenam

```
Call:
glm.nb(formula = Frekuensi ~ Umur + sex + fams + i11 + ranap +
  Umur * sex + Umur * i11 + Umur * ranap + sex * i11 + i11 *
  ranap, data = dataprediksi, init.theta = 0.6394179858, link = log)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.5471  -0.8420  -0.7528  -0.4168   7.7994

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.038393   0.170713  11.940 < 2e-16 ***
Umur         -0.011866   0.003106  -3.820 0.000134 ***
sex2         -0.597847   0.216177  -2.766 0.005683 **
fams1        -1.111879   0.429077  -2.591 0.009561 **
i111         2.136679   0.509143   4.197 2.71e-05 ***
ranap         0.280759   0.058432   4.805 1.55e-06 ***
Umur:sex2     0.009939   0.003917   2.537 0.011166 *
Umur:i111    -0.038016   0.008951  -4.247 2.17e-05 ***
Umur:ranap   -0.002330   0.001131  -2.061 0.039315 *
sex2:i111    0.486332   0.232364   2.093 0.036351 *
i111:ranap   -0.087041   0.051119  -1.703 0.088623 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(0.6394) family taken to be 1)

Null deviance: 2786.3  on 2437  degrees of freedom
Residual deviance: 2544.5  on 2427  degrees of freedom
AIC: 12981

Number of Fisher Scoring iterations: 1

            Theta: 0.6394
            Std. Err.: 0.0177

2 x log-likelihood: -12956.7060
```

Model Ketujuh (Model Akhir)

```
Call:
glm.nb(formula = Frekuensi ~ Umur + sex + fams + i11 + ranap +
  Umur * sex + Umur * i11 + Umur * ranap + sex * i11, data = dataprediksi,
  init.theta = 0.6388014301, link = log)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.5246  -0.8431  -0.7558  -0.4212   7.7599

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.048981   0.170520  12.016 < 2e-16 ***
Umur         -0.011847   0.003107   -3.813 0.000137 ***
sex2         -0.589236   0.216213   -2.725 0.006425 **
fams1        -1.124977   0.429239   -2.621 0.008771 **
i11          1.963194   0.491729   3.992 6.54e-05 ***
ranap         0.275284   0.058403   4.713 2.43e-06 ***
Umur:sex2     0.009772   0.003918    2.494 0.012623 *
Umur:i11     -0.037010   0.008894   -4.161 3.16e-05 ***
Umur:ranap   -0.002367   0.001131   -2.093 0.036374 *
sex2:i11      0.470138   0.231970    2.027 0.042691 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(0.6388) family taken to be 1)

Null deviance: 2783.9  on 2437  degrees of freedom
Residual deviance: 2545.1  on 2428  degrees of freedom
AIC: 12981

Number of Fisher Scoring iterations: 1

              Theta:  0.6388
            Std. Err.:  0.0176

2 x log-likelihood:  -12959.4050
```

- **Uji Goodness of Fit**

Diketahui bahwa suatu model dapat dikatakan layak jika model tersebut memiliki nilai $\frac{\text{Residual Deviance}}{\text{Degree of Freedom Residual}} < 1$ atau mendekati 1. Berdasarkan keluaran dari hasil komputasi menggunakan software Rstudio di atas, diketahui nilai Residual Deviance sebesar 2545.1 dengan degrees of freedomnya sebesar 2428 sehingga diperoleh nilai $\frac{\text{Residual Deviance}}{\text{Degree of Freedom Residual}} = 1.05 \approx 1$, sehingga model pertama GLM Negative Binomial dapat dikatakan layak untuk digunakan.

- **Uji Parsial untuk Konstanta**

- a. **Hipotesis**

H_0 : Konstanta tidak mempengaruhi model ketujuh (akhir) secara signifikan

H_1 : Konstanta mempengaruhi model ketujuh (akhir) secara signifikan

b. Tingkat Signifikansi

$$\alpha = 0.05$$

c. Statistik Uji

$$P\text{-Value} < 2E-16$$

d. Daerah Kritik

$$H_0 \text{ ditolak apabila nilai } P\text{-Value} < \alpha$$

e. Kesimpulan

Berdasarkan hasil keluaran Analisis Negative Binomial menggunakan software Rstudio, diketahui nilai P-Value dari konstanta model sebesar $2E-16 < 0.05 = \alpha$, sehingga H_0 ditolak. Dengan demikian, dapat diambil kesimpulan bahwa konstanta mempengaruhi model ketujuh (akhir) secara signifikan.

• **Uji Parsial untuk Koefisien**

a. Hipotesis

H_0 : Koefisien variabel ke-i tidak mempengaruhi model ketujuh (akhir) secara signifikan, dengan $i = 1, 2, 3, \dots, 15$.

H_1 : Koefisien variabel ke-i mempengaruhi model ketujuh (akhir) secara signifikan, dengan $i = 1, 2, 3, \dots, 15$.

b. Tingkat Signifikansi

$$\alpha = 0.05$$

c. Daerah Kritik

$$H_0 \text{ ditolak apabila nilai } P\text{-Value} < \alpha$$

d. Statistik Uji dan Kesimpulan

No	Variabel	P-Value	Kesimpulan
1.	Umur	0.000137	Berdasarkan hasil keluaran, diketahui nilai P-Value dari konstanta model sebesar $0.000137 < 0.05 = \alpha$, sehingga H_0 ditolak. Dengan demikian, dapat diambil kesimpulan bahwa koefisien

			variabel Umur mempengaruhi model ketujuh (akhir) secara signifikan.
2.	sex2	0.006425	Berdasarkan hasil keluaran, diketahui nilai P-Value dari konstanta model sebesar $0.006425 < 0.05 = \alpha$, sehingga H_0 ditolak. Dengan demikian, dapat diambil kesimpulan bahwa koefisien variabel Sex2 mempengaruhi model ketujuh (akhir) secara signifikan.
3.	fams1	0.008771	Berdasarkan hasil keluaran, diketahui nilai P-Value dari konstanta model sebesar $0.008771 < 0.05 = \alpha$, sehingga H_0 ditolak. Dengan demikian, dapat diambil kesimpulan bahwa koefisien variabel fams1 mempengaruhi model ketujuh (akhir) secara signifikan.
4.	ill1	6.54E-05	Berdasarkan hasil keluaran, diketahui nilai P-Value dari konstanta model sebesar $6.54E-05 < 0.05 = \alpha$, sehingga H_0 ditolak. Dengan demikian, dapat diambil kesimpulan bahwa koefisien variabel ill1 mempengaruhi model ketujuh (akhir) secara signifikan.
5.	ranap	2.43E-06	Berdasarkan hasil keluaran, diketahui nilai P-Value dari konstanta model sebesar $2.43E-06 < 0.05 = \alpha$, sehingga H_0 ditolak. Dengan demikian, dapat diambil kesimpulan bahwa koefisien

			variabel ranap mempengaruhi model ketujuh (akhir) secara signifikan.
6.	Umur:sex2	0.012623	Berdasarkan hasil keluaran, diketahui nilai P-Value dari konstanta model sebesar $0.012623 < 0.05 = \alpha$, sehingga H_0 ditolak. Dengan demikian, dapat diambil kesimpulan bahwa koefisien variabel Umur:sex2 mempengaruhi model ketujuh (akhir) secara signifikan.
8.	Umur:ill1	3.16E-05	Berdasarkan hasil keluaran, diketahui nilai P-Value dari konstanta model sebesar $3.16E-05 < 0.05 = \alpha$, sehingga H_0 ditolak. Dengan demikian, dapat diambil kesimpulan bahwa koefisien variabel Umur:ill1 mempengaruhi model ketujuh (akhir) secara signifikan.
9.	Umur:ranap	0.036374	Berdasarkan hasil keluaran, diketahui nilai P-Value dari konstanta model sebesar $0.036374 < 0.05 = \alpha$, sehingga H_0 ditolak. Dengan demikian, dapat diambil kesimpulan bahwa koefisien variabel Umur:ranap mempengaruhi model ketujuh (akhir) secara signifikan.
11.	sex2:ill1	0.042691	Berdasarkan hasil keluaran, diketahui nilai P-Value dari konstanta model sebesar $0.042691 < 0.05 = \alpha$, sehingga H_0 ditolak. Dengan demikian, dapat diambil kesimpulan bahwa koefisien

			variabel sex2:ill1 mempengaruhi model ketujuh (akhir) secara signifikan.
--	--	--	--

Interpretasi :

Uji Goodness of Fit dilakukan untuk mengetahui apakah model sudah fit (layak) untuk digunakan, sedangkan Uji Parsial, baik untuk Konstanta maupun Koefisien, dilakukan untuk mengetahui apakah nilai konstanta dan koefisien signifikan terhadap model GLM yang didapatkan. Berdasarkan keluaran dari Analisis GLM Negative binomial di atas, diketahui bahwa, melalui Uji Goodness of Fit, model dinyatakan layak untuk digunakan. Kemudian, melalui Uji Parsial, diketahui bahwa konstanta dan 8 dari 15 koefisien variabel prediktor dinyatakan mempengaruhi model secara signifikan. Hal ini berarti masih terdapat variabel prediktor yang dinyatakan tidak mempengaruhi model yang dimiliki, sehingga dilakukan pengeluaran variabel prediktor dengan nilai P-Value terbesar dari mode. Hal ini dilakukan berulang-ulang hingga mendapatkan model GLM yang layak digunakan dengan konstanta dan seluruh koefisien variabel prediktornya signifikan mempengaruhi model. Pada kasus ini, telah dilakukan pengulangan sebanyak enam kali hingga mendapatkan model yang layak digunakan dan konstanta serta seluruh koefisien variabelnya signifikan terhadap model pada model ketujuh.

3.4. Pembentukan Model Akhir

Berdasarkan model ketujuh analisis negatif binomial, diperoleh model akhir yang layak digunakan, sebagai berikut:

$$\begin{aligned} \log(\mu) = & 2.048981 - 0.011847 * \text{Umur} - 0.589236 * \text{sex2} - 1.124977 * \text{fams1} \\ & + 1.963194 * \text{ill1} + 0.275284 * \text{ranap} + 0.009772 * (\text{Umur} \times \text{sex2}) \\ & - 0.037010 * (\text{Umur} \times \text{ill1}) - 0.002367 * (\text{Umur} \times \text{ranap}) \\ & + 0.470138 * (\text{sex2} \times \text{ill1}) \end{aligned}$$

dengan

- μ : nilai harapan (mean) frekuensi klaim asuransi BPJS
- Variabel independen : Umur, sex, fams, ill, dan ranap

- Koefisien masing-masing variabel independen yang bernilai positif menunjukkan bahwa dengan peningkatan nilai variabel independen, log dari nilai harapan meningkat.
- Koefisien masing-masing variabel independen yang bernilai negatif menunjukkan bahwa dengan peningkatan nilai variabel independen, log dari nilai harapan menurun.

Intepretasi :

1. **Koefisien variabel Umur adalah -0.011847.** Artinya, setiap peningkatan satu unit dalam Umur akan menurunkan log dari nilai harapan (μ) frekuensi klaim sebesar 0.011847 dengan mengontrol variabel lain dalam model.
2. **Koefisien untuk Jenis kelamin perempuan (sex2) adalah -0.589236.** Artinya, peserta laki-laki memiliki nilai harapan (μ) frekuensi klaim $e^{0.589236}$ lebih besar daripada peserta perempuan dengan mengontrol variabel lain dalam model.
3. **Koefisien untuk riwayat keturunan (fams1) adalah -1.124977.** Artinya, peserta yang tidak memiliki riwayat keturunan penyakit gagal ginjal memiliki nilai nilai harapan (μ) frekuensi klaim $e^{1.124977}$ lebih besar daripada peserta yang memiliki riwayat keturunan dengan mengontrol variabel lain dalam model.
4. **Koefisien untuk riwayat penyakit katastropik selain gagal ginjal (ill1) adalah 1.963194.** Artinya, peserta yang memiliki riwayat penyakit katastropik lain selain gagal ginjal memiliki nilai harapan (μ) frekuensi klaim $e^{1.963194}$ lebih besar daripada peserta yang tidak memiliki riwayat dengan mengontrol variabel lain dalam model.
5. **Koefisien variabel ranap adalah 0.275284.** Artinya, setiap peningkatan satu unit dalam ranap akan menaikkan log dari nilai harapan (μ) frekuensi klaim sebesar 0.275284 dengan mengontrol variabel lain dalam model.
6. **Koefisien untuk interaksi Umur \times sex2 adalah 0.009772.** Artinya, interaksi antara Umur dan sex2 akan mempengaruhi peningkatan nilai log dari nilai harapan (μ) frekuensi klaim sebesar 0.009772 dengan mengontrol variabel lain dalam model.
7. **Koefisien untuk interaksi Umur \times ill1 adalah -0.037010.** Artinya, interaksi antara Umur dan ill1 akan mempengaruhi penurunan nilai log dari nilai harapan (μ) frekuensi klaim sebesar 0.037010 dengan mengontrol variabel lain dalam model.

8. **Koefisien untuk interaksi Umur×ranap adalah -0.002367.** Artinya, interaksi antara Umur dan ranap akan mempengaruhi penurunan nilai log dari nilai harapan (μ) frekuensi klaim sebesar 0.002367 dengan mengontrol variabel lain dalam model.
9. **Koefisien untuk interaksi sex2×ill1 adalah 0.470138.** Artinya, interaksi antara sex2 dan ill1 akan mempengaruhi peningkatan nilai log dari nilai harapan (μ) frekuensi klaim sebesar 0.470138 dengan mengontrol variabel lain dalam model.

BAB IV

PENUTUP

4.1. Kesimpulan

Gagal ginjal telah menjadi perhatian serius dalam sektor kesehatan karena tidak hanya membutuhkan pengobatan yang lama dan mahal, tetapi juga sulit untuk dideteksi pada tahap awal. Seiring bertambahnya jumlah individu yang terkena penyakit ini, terjadi peningkatan yang signifikan dalam klaim kesehatan terkait perawatan dialisis dan transplantasi ginjal. Data dari Global Burden Disease dan Indonesia Renal Registry menunjukkan pertumbuhan yang tinggi dari pasien yang membutuhkan perawatan khusus ini. Penanganan gagal ginjal bukanlah hanya masalah individual, tetapi juga memberikan beban finansial yang signifikan bagi masyarakat dan sistem kesehatan secara menyeluruh. Salah satu cara untuk mengurangi beban finansial yang terkait dengan penyakit katastrofik seperti gagal ginjal adalah melalui perlindungan kesehatan seperti yang ditawarkan oleh BPJS Kesehatan di Indonesia.

BPJS Kesehatan berperan dalam mengurangi risiko finansial pada individu dan mengalihkannya menjadi risiko bersama (*sharing risk*), sehingga membantu dalam pembiayaan perawatan yang kontinu dan seringkali seumur hidup bagi penderita gagal ginjal. Meski sudah ada sistem jaminan kesehatan, klaim yang harus dipenuhi oleh BPJS Kesehatan untuk kasus gagal ginjal tetap tinggi, mencapai jumlah yang besar pada tahun 2022, sebesar triliunan rupiah dengan jutaan kasus. Penelitian ini bertujuan untuk mengeksplorasi faktor-faktor yang memengaruhi frekuensi klaim BPJS terhadap penanganan penyakit gagal ginjal.

Dengan melihat hasil dari analisis regresi *poisson* dan regresi negatif binomial pada model di atas akhirnya mendapat kesimpulan bahwa regresi negatif binomial lah yang menjadi model yang paling layak untuk digunakan dalam menentukan berapa jumlah klaim asuransi yang diperlukan dengan beberapa variabel yang signifikan terhadap model.

4.2. Kritik dan Saran

Mungkin sebaiknya kepada pihak BPJS bisa menggunakan contoh model yang telah kita uji ini untuk menentukan jumlah klaim asuransi. Selain itu, juga untuk bisa menggunakan beberapa percobaan model untuk menentukan metode terbaik dan tidak hanya terpaku kepada 1 model saja.

LAMPIRAN

- Syntax R, Data yang digunakan, PPT Presentasi, dan Video Presentasi
<https://drive.google.com/drive/folders/1Zd7SGZenjb12bZF9y0hN3-geKcOPikfF?usp=sharing>

DAFTAR PUSTAKA

- BPJS Kesehatan. (2023). Laporan Pengelolaan Program Tahun 2022 & Laporan Keuangan.
- Heniwati & Thabrany, H. (2017). Perbandingan Klaim Penyakit Katastropik Peserta Jaminan.
- KBBI. (n.d.). Katastrofe. In KBBI. Retrieved November 27, 2023, from <https://kbbi.web.id/katastrofe>
- Kesehatan Nasional di Provinsi DKI Jakarta dan Nusa Tenggara Timur Tahun 2014. *Jurnal Ekonomi Kesehatan Indonesia*, 1(2). <https://doi.org/10.7454/eki.v1i2.1771>
- Nugraheni, W. P., & Hartono, R. K. (2017). Determinan Pengeluaran Kesehatan Katastropik Rumah Tangga Indonesia Pada Tahun Pertama Implementasi Program JKN. *Buletin Penelitian Kesehatan*, 45(1). <https://doi.org/10.22435/bpk.v45i1.6069.27-36>
- Peraturan Badan Penyelenggara Jaminan Sosial Kesehatan Nomor 3 Tahun 2017 Tentang Pengelolaan Administrasi Klaim Fasilitas Kesehatan dalam Penyelenggaraan Jaminan Kesehatan Nasional, Pub. L. No. 3 (2017). <https://jdih.djsn.go.id/assets/file/03-bpjs-2017.pdf>
- Perkumpulan Nefrologi Indonesia (PERNEFRI). (2016). Indonesia Renal Registry. <https://www.indonesianrenalregistry.org/data/>
- Riza, Y., Budiarto, W., Haksama, S., Kuntoro, Yudhastuti, R., Wibowo, A., & Notobroto, H. B. (2023). Health services for catastrophic patients through the national health insurance program: literature review. *Journal of Public Health in Africa*, 14(S2). <https://doi.org/10.4081/jphia.2023.2559>
- Tahun 2022 (Auditan). <https://web.bpjs-kesehatan.go.id/uploads/information/27072023114426-4f131b52-f725-4d9e-bc41-3ab9d344ee08.pdf>
- University of Washington. Institute for Health Metrics and Evaluation. (2010). *The Global Burden of Disease : Generating Evidence, Guiding Policy*.
- Haris, M. A., & Arum, P. R. (2022). Negative Binomial Regression and Generalized Poisson Regression Models On The Number of Traffic Accidents in Central Java. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, 16(2), 471–482. <https://doi.org/10.30598/barekengvol16iss2pp471-482>

- Santi, V. M., Wiyono, A., & Sudarwanto. (2021). Pemodelan Jumlah Kasus Malaria di Indonesia Menggunakan Generalized Linear Model. *Jurnal Statistika Dan Aplikasinya*, 5(1), 112–120. <https://doi.org/10.21009/jsa.05111>
- Sastri, R., & Setiadi, Y. (2018). Generalized Linear Mixed Model Untuk Data Kematian Bayi Di Indonesia, 1–17. https://stis.ac.id/sipadu/pegawai/upload_jurnal/file_1549952003.pdf