

BAB I PENDAHULUAN

1.1 Latar Belakang

Analisis regresi merupakan salah satu metode statistika yang dapat digunakan untuk memodelkan hubungan antara variabel prediktor dan respon. Tipe data variabel respon dapat berupa kontinu atau diskrit. Salah satu jenis data diskrit adalah data *count* (jumlah). Data *count* adalah data yang menggambarkan banyaknya kejadian yang terjadi dalam jangka waktu tertentu (Hidayati et al., 2019). Analisis regresi yang digunakan untuk memodelkan variabel respon bersifat diskrit adalah regresi Poisson.

Regresi Poisson adalah model regresi yang digunakan untuk menganalisis variabel respon bertipe diskrit dan integer tidak negatif atau data *count* (Rahayu, 2020). Terdapat asumsi yang harus dipenuhi dalam regresi Poisson, yaitu ekuidispersi dimana nilai varians dan rata-rata dari variabel respon (Y) harus memiliki nilai yang sama. Namun pada kenyataannya, sering terjadi pelanggaran asumsi ini dimana nilai variansnya lebih besar dari nilai rata-rata yang dinamakan dengan overdispersi atau nilai variansnya lebih kecil dari nilai rata-rata yang dinamakan dengan underdispersi. Menggunakan analisis regresi Poisson saat terjadi overdispersi pada data menjadi kurang akurat karena berdampak pada nilai *standard error* menjadi *underestimate* atau lebih kecil dari nilai sesungguhnya sehingga dapat menghasilkan kesimpulan yang kurang tepat (Prahutama & Sudarno, 2018). Untuk mengatasi masalah ini, model regresi yang dapat digunakan dua diantaranya adalah dengan menggunakan pemodelan *Generalized Poisson Regression* (GPR) dan Binomial Negatif (Saputro et al., 2021). Model GPR merupakan perluasan dari regresi Poisson yang mampu menangani overdispersi dengan lebih baik. GPR dengan regresi Poisson merupakan suatu model *generalized linear model* (GLM). Perbedaan dengan regresi Poisson adalah GPR mengasumsikan komponen randomnya berdistribusi *generalized Poisson*. Dengan kata lain, model ini dapat digunakan untuk data diskrit yang berdistribusi Poisson tanpa adanya asumsi ekuidispersi (Chaniago & Sri, 2022). Penelitian mengenai *Generalized Poisson Regression* pernah dilakukan Famoye et al. (2004), Grover et al. (2015), Savitri et al. (2022), dan Haris et al. (2022) menemukan bahwa pemodelan *Generalized Poisson Regression* (GPR) merupakan pemodelan yang tepat untuk mengatasi kasus overdispersi.

Beberapa penelitian tersebut menggunakan pendekatan regresi parametrik, yaitu pendekatan yang digunakan untuk bentuk kurva regresi diketahui. Namun, pada kenyataannya seringkali hubungan antara variabel prediktor dan variabel respon tidak mengikuti pola tertentu tetapi cenderung tidak beraturan sehingga pendekatan parametrik kurang tepat digunakan. Pada kondisi ini pendekatan nonparametrik menjadi alternatif pilihan. Regresi nonparametrik merupakan pendekatan yang tidak mengasumsikan bentuk kurva regresi seperti pada regresi parametrik. Pendekatan regresi nonparametrik dapat memodelkan fungsi secara fleksibel sesuai dengan data yang ada. Regresi nonparametrik menyediakan opsi

serbaguna untuk pemodelan data, khususnya dalam keberadaan pola nonlinier. Regresi nonparametrik memiliki fleksibilitas yang tinggi dalam mengestimasi kurva regresi (Eubank, 1999).

Salah satu pendekatan dalam regresi nonparametrik adalah model *Multivariate Adaptive Regression Splines* (MARS) yang diperkenalkan oleh Friedman pada tahun 1991. MARS dapat mengidentifikasi dan memodelkan hubungan nonlinier yang kompleks antara variabel respon dan variabel prediktor tanpa memerlukan asumsi tertentu. Model ini dapat digunakan pada data berdimensi tinggi, khususnya ketika jumlah variabel prediktor berada dalam rentang 3 hingga 20 (Friedman, 1991). MARS merupakan pengembangan dari pendekatan *Recursive Partitioning Regression* (RPR) yang dikombinasikan dengan metode spline, sehingga menghasilkan model regresi yang kontinu pada titik-titik *knot*, di mana setiap *knot* tersambung secara mulus dengan fungsi basisnya (Otok et al., 2006). Salah satu keunggulan MARS lainnya adalah kemampuannya dalam menentukan letak *knot* secara otomatis berdasarkan nilai *Generalized Cross Validation* (GCV) minimum, yang digunakan untuk memperoleh model terbaik dengan kompleksitas yang optimal (Adiguzel & Mehmet, 2023).

Penelitian tentang MARS pernah dilakukan Dukalang (2017) untuk memodelkan jumlah kematian bayi dengan menggunakan pendekatan MARS, Meilisa et al. (2023) memodelkan MARS untuk menentukan faktor-faktor yang menyebabkan demam berdarah di Rumah Sakit Islam Jemursari Surabaya, dan Permatasi et al. (2021) untuk menentukan faktor-faktor yang mempengaruhi Angka Kematian Bayi (AKB) di Nusa Tenggara Timur dengan menggunakan pendekatan MARS.

Kematian ibu merupakan indikator penting untuk menilai derajat kesehatan masyarakat dan kualitas layanan kesehatan suatu negara. Menurut Organisasi Kesehatan Dunia (WHO), kematian ibu adalah kematian seorang wanita selama kehamilan atau dalam 42 hari setelah kehamilan berakhir, yang disebabkan oleh kondisi yang terkait atau diperberat oleh kehamilan atau penanganannya, namun bukan akibat kecelakaan atau cedera. Di Indonesia, kematian ibu masih menjadi masalah utama dalam kesehatan ibu dan anak. Meskipun sebagian besar persalinan ditolong oleh tenaga kesehatan terlatih, Angka Kematian Ibu (AKI) masih tergolong tinggi (UNFPA, 2022). Penelitian oleh Indriyani & Puspitasari (2016), Ayu et al. (2022), Perdana et al. (2023), Jones et al. (2024), dan Sari et al. (2023) menunjukkan bahwa beberapa faktor yang memengaruhi kematian ibu antara lain adalah persentase ibu hamil yang mengikuti program K1 dan K4, persentase persalinan di fasilitas kesehatan, pemberian imunisasi Td3/Td4, serta pemberian vitamin A pada masa nifas. Di Provinsi Sulawesi Selatan, angka kematian ibu pada tahun 2021 tercatat sebanyak 198 kasus, meningkat dibandingkan tahun-tahun sebelumnya, yaitu 139 kasus pada 2018, 144 kasus pada 2019, dan 133 kasus pada 2020. Tingginya angka kematian ibu menunjukkan perlunya perhatian serius untuk mengidentifikasi dan mengatasi faktor-faktor penyebabnya.

Berdasarkan uraian diatas, penulis menggunakan estimasi *Generalized Poisson Regression* pada metode MARS, sehingga menjadi metode kombinasi

Multivariate Adaptive Generalized Poisson Regression Splines (MAGPRS) yang diterapkan pada kasus Angka Kematian Ibu di Provinsi Sulawesi Selatan. Hal ini dilakukan karena pola data tidak linier, variabel respon merupakan variabel *count* (jumlah), dan diduga berdistribusi Poisson. *Generalized Poisson Regression* digunakan karena diduga asumsi distribusi Poisson tidak terpenuhi, yaitu nilai rata-rata tidak sama dengan varians (Prastika et al., 2021).

Kajian mengenai metode MAGPRS sebelumnya telah dilakukan oleh Otok et al. (2019) yang menerapkannya pada kasus jumlah penderita ISPA pada bayi di Surabaya tahun 2017, serta oleh Kartini et al. (2022) untuk memodelkan dan memprediksi produksi padi di Kabupaten Bojonegoro. Namun, pada penelitian-penelitian tersebut, proses estimasi parameter model dilakukan menggunakan pendekatan *Weighted Least Squares* (WLS). Berbeda dengan penelitian-penelitian terdahulu, penelitian ini menggunakan pendekatan estimasi parameter *Maximum Likelihood* Terboboti (*Weighted Maximum Likelihood*). Pendekatan ini dipilih karena lebih sesuai untuk memaksimalkan fungsi likelihood pada model *Generalized Linear Model* (GLM) dengan distribusi *Generalized Poisson*, sehingga diharapkan menghasilkan estimasi parameter yang lebih efisien dan akurat, khususnya dalam konteks data *count* dengan overdispersi.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, rumusan masalah yang dikaji pada penelitian ini, yaitu:

1. Bagaimana estimasi parameter model *Multivariate Adaptive Generalized Poisson Regression Spline* dengan *maximum likelihood* terboboti?
2. Faktor-faktor apa saja yang berpengaruh terhadap kematian ibu di Provinsi Sulawesi Selatan dengan menggunakan metode *Multivariate Adaptive Generalized Poisson Regression Spline*?

1.3 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Penelitian ini hanya menggunakan data kematian ibu yang dikumpulkan dari Dinkes Sulawesi Selatan tahun 2021.
2. Kombinasi *Basis Function* (BF) yang digunakan adalah 10, 15 dan 20.
3. *Minimum Observation* (MO) yang digunakan adalah 0, 1, 2 dan 3.
4. Jumlah *Maximum Interaction* (MI) yang digunakan yaitu 1, 2, dan 3.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah, tujuan dari penelitian ini adalah:

1. Mengestimasi parameter model *Multivariate Adaptive Generalized Poisson Regression Spline* dengan *maximum likelihood* terboboti.
2. Menentukan faktor-faktor yang berpengaruh terhadap kematian ibu di Provinsi Sulawesi Selatan dengan menggunakan metode *Multivariate Adaptive Generalized Poisson Regression Spline*.

1.5 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat sebagai berikut:

1. Menambah wawasan dan pengetahuan tentang estimasi parameter model *Multivariate Adaptive Generalized Poisson Regression Spline* pada data yang mengandung overdispersi.
2. Dapat memberikan wawasan mengenai penerapan model *Multivariate Adaptive Generalized Poisson Regression Spline* pada data kasus kematian ibu, yang diharapkan dapat berkontribusi dalam upaya mengurangi angka kematian ibu.

1.6 Teori

Sub-bab teori ini membahas mengenai beberapa teori yang digunakan untuk membahas pokok pembahasan terkait metode MAGPRS.

1.6.1 Distribusi Poisson

Distribusi Poisson merupakan distribusi dimana kejadiannya tergantung pada interval waktu tertentu atau disuatu daerah tertentu dengan hasil pengamatan berupa variabel diskrit. Jika variabel acak diskrit Y mengikuti distribusi Poisson dengan parameter μ , maka fungsi probabilitasnya diberikan oleh (Myers, 2010):

$$P(Y = y; \mu) = \frac{e^{-\mu} \mu^y}{y!}; y = 0, 1, 2, \dots; \mu > 0 \quad (1)$$

dengan

y : jumlah kejadian dalam suatu interval waktu atau wilayah tertentu.

μ : rata-rata jumlah kejadian dalam interval waktu atau wilayah tertentu.

Dalam distribusi Poisson terdapat keadaan dimana nilai rata-rata dan variansinya bernilai sama, yaitu $E[Y] = Var[Y] = \mu$ yang disebut dengan asumsi ekuidispersi. Asumsi tersebut dapat dibuktikan dengan menggunakan fungsi pembangkit momen, yaitu:

1. Fungsi pembangkit momen untuk distribusi Poisson

$$\begin{aligned} M_Y(t) &= E(e^{tY}) \\ &= \sum_{y=0}^{\infty} e^{ty} P(Y = y) \\ &= \sum_{y=0}^{\infty} e^{ty} \frac{e^{-\mu} \mu^y}{y!} \\ &= e^{-\mu} \sum_{y=0}^{\infty} \frac{(e^t \mu)^y}{y!} \\ &= e^{-\mu} e^{\mu e^t} \\ M_Y(t) &= e^{\mu(e^t - 1)} \end{aligned} \quad (2)$$

2. Rata-rata

$$\begin{aligned} M'_Y(t) &= \frac{d}{dt} e^{\mu(e^t - 1)} \\ &= e^{\mu(e^t - 1)} \mu e^t \\ E[Y] &= M'_Y(t = 0) \\ &= e^{\mu(e^0 - 1)} \mu e^0 \\ &= e^{\mu(1 - 1)} \mu \times 1 \end{aligned}$$

$$E[Y] = \mu \quad (3)$$

3. Varians

$$\begin{aligned} M_Y''(t) &= \frac{d^2}{dt^2} e^{\mu(e^t-1)} \\ &= e^{\mu(e^t-1)}(\mu e^t)(\mu e^t) + e^{\mu(e^t-1)}(\mu e^t) \\ &= e^{\mu(e^t-1)} \times \mu e^t(\mu e^t + 1) \\ E[Y^2] &= M_Y''(t=0) \\ &= e^{\mu(e^0-1)} \times \mu e^0(\mu e^0 + 1) \\ &= 1 \times \mu \times 1(\mu + 1) \\ &= \mu^2 + \mu \end{aligned}$$

Sehingga:

$$\begin{aligned} Var[Y] &= E[Y^2] - (E[Y])^2 \\ &= (\mu^2 + \mu) - \mu^2 \\ Var[Y] &= \mu \end{aligned} \quad (4)$$

Untuk mengetahui apakah data yang diamati berdistribusi Poisson atau tidak, dapat dilakukan pengujian dengan Uji *Kolmogorov-Smirnov*. Dimana hipotesis pengujiannya sebagai berikut:

H_0 : Data Y berdistribusi Poisson

H_1 : Data Y tidak berdistribusi Poisson

Statistik uji yang digunakan, yaitu:

$$D_{hitung} = \max|F_0(y) - S_n(y)| \quad (5)$$

dengan

- D : nilai standar deviasi absolut maksimum antara $F_0(y)$ dan $S_n(y)$
- $F_0(y)$: fungsi distribusi frekuensi kumulatif relatif dari distribusi teoritis dibawah H_0
- $S_n(y)$: distribusi frekuensi kumulatif pengamatan sebanyak sampel

Dengan kriteria uji, yaitu tolak H_0 jika nilai $D_{hitung} > D_{tabel}$ (tabel *Kolmogorov-Smirnov*) atau nilai signifikansi $p < \alpha$ (tingkat signifikansi yang digunakan $\alpha = 0.05$) (Safitri et al., 2011).

1.6.2 Regresi Poisson

Regresi Poisson adalah salah satu bentuk model analisis regresi yang digunakan untuk memodelkan data diskrit (*count data*). Model ini merupakan bagian dari *Generalized Linear Model* (GLM) karena peubah respon memiliki sebaran dalam keluarga eksponensial, yaitu distribusi Poisson. Dalam model ini, terdapat fungsi penghubung (*link function*) $g(\cdot)$ yang menghubungkan rata-rata variabel respon dengan variabel prediktor, yaitu (Sinaga & Sinulingga, 2021):

$$g(\mu_i) = \eta_i = \beta_0 + \sum_{j=1}^p \beta_j x_{ij} = \mathbf{x}_i^T \boldsymbol{\beta} \quad (6)$$

dengan $i = 1, 2, \dots, n$ dan $j = 1, 2, \dots, p$

Pada model regresi Poisson, biasanya fungsi penghubung yang digunakan untuk adalah $\log$, karena rata-rata dari variabel responnya akan berbentuk fungsi eksponensial dan menjamin bahwa nilai variabel yang ditaksir dari variabel responnya akan bernilai non-negatif. Fungsi penghubung $\log$ berbentuk:

$$g(\mu_i) = \log(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} \quad (7)$$

Dengan demikian, hubungan antara rata-rata variabel respon dan prediktor linear dapat ditulis sebagai:

$$\begin{aligned} \log(\mu_i) &= \mathbf{x}_i^T \boldsymbol{\beta} \\ e^{\log(\mu_i)} &= e^{\mathbf{x}_i^T \boldsymbol{\beta}} \\ \mu_i &= e^{\mathbf{x}_i^T \boldsymbol{\beta}} \end{aligned} \quad (8)$$

Sehingga fungsi penghubung untuk model regresi Poisson mempunyai logartima sebagai berikut:

$$\begin{aligned} \log(\mu_i) &= \mathbf{x}_i^T \boldsymbol{\beta} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} \\ \mu_i &= \exp(\mathbf{x}_i^T \boldsymbol{\beta}) = \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}) \end{aligned} \quad (9)$$

Rata-rata untuk model regresi Poisson adalah sebagai berikut:

$$\mu_i = \exp(\mathbf{x}_i^T \boldsymbol{\beta}) \quad (10)$$

Variansi untuk model regresi Poisson adalah sebagai berikut:

$$\text{var}(y_i) = \mu_i = \exp(\mathbf{x}_i^T \boldsymbol{\beta}) \quad (11)$$

Dengan demikian, model regresi Poisson dengan fungsi penghubung $\log$ dapat ditulis sebagai:

$$\begin{aligned} y_i &= \mu_i + \varepsilon_i \\ &= \exp(\mathbf{x}_i^T \boldsymbol{\beta}) + \varepsilon_i \end{aligned} \quad (12)$$

1.6.3 Penaksiran Parameter Regresi Poisson

Pendugaan parameter dalam regresi Poisson dapat dilakukan melalui metode *maximum likelihood* (MLE). Fungsi *likelihood* dari distribusi Poisson adalah:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!} \quad (13)$$

Kemudian mengambil logaritma dari fungsi *likelihood* (13), diperoleh fungsi *log likelihood*:

$$l(\boldsymbol{\beta}) = \sum_{i=1}^n y_i \ln \mu_i - \sum_{i=1}^n \mu_i - \sum_{i=1}^n \ln y_i! \quad (14)$$

Karena $\mu_i = \exp(\mathbf{x}_i^T \boldsymbol{\beta})$, maka persamaan (14) menjadi:

$$l(\boldsymbol{\beta}) = \sum_{i=1}^n y_i (\mathbf{x}_i^T \boldsymbol{\beta}) - \sum_{i=1}^n \exp(\mathbf{x}_i^T \boldsymbol{\beta}) - \sum_{i=1}^n \ln y_i \quad (15)$$

Untuk mendapatkan estimasi parameter $\boldsymbol{\beta}$, kita mencari turunan pertama dari fungsi *log-likelihood* (15) terhadap $\boldsymbol{\beta}$ dan menyamakannya dengan nol.

$$\frac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^n y_i \mathbf{x}_i^T - \sum_{i=1}^n \mathbf{x}_i^T \exp(\mathbf{x}_i^T \boldsymbol{\beta}) = 0 \quad (16)$$

Karena terdapat suku $\exp(\mathbf{x}_i^T \boldsymbol{\beta})$ di dalam persamaan (16) maka bentuk pasti dari $\boldsymbol{\beta}$ sulit ditemukan, sehingga pendekatan iteratif digunakan dengan metode

Newton-Raphson (Rahmayanti dan Setyo, 2018). Berikut prosedur iterasi *Newton-Raphson* adalah sebagai berikut (Zakiah et al., 2015):

- 1) Memilih taksiran awal dari β , yaitu $\hat{\beta}_{(0)} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \end{bmatrix}$, penentuan nilai awal ini diperoleh

dengan metode *Ordinary Least Square* (OLS), yaitu:

$$\hat{\beta}_{(0)} = (X^T X)^{-1} X^T Y$$

- 2) Membentuk fungsi *likelihood* dari $\beta_0, \beta_1, \dots, \beta_p$ yaitu $l(\beta)$ yang berdasarkan pada persamaan (15).

- 3) Membentuk vektor gradien $g(\hat{\beta}_{(m)}) = \begin{bmatrix} \frac{\partial l(\beta)}{\partial \beta_0} \\ \frac{\partial l(\beta)}{\partial \beta_1} \\ \vdots \\ \frac{\partial l(\beta)}{\partial \beta_p} \end{bmatrix}$

dengan $g(\hat{\beta}_{(m)})$ adalah vektor berukuran $(p + 1) \times 1$ yang elemen-elemennya merupakan turunan pertama yang diturunkan secara parsial $l(\beta)$ yang dihitung pada $\hat{\beta}_{(m)} = \hat{\beta}_{(m+1)}$.

- 4) Membentuk matriks Hessian $H(\hat{\beta}_{(m)})$, yaitu:

$$H(\hat{\beta}_{(m)}) = \begin{bmatrix} \frac{\partial^2 l(\beta)}{\partial \beta_0^2} & \frac{\partial^2 l(\beta)}{\partial \beta_0 \partial \beta_1} & \dots & \frac{\partial^2 l(\beta)}{\partial \beta_0 \partial \beta_p} \\ \frac{\partial^2 l(\beta)}{\partial \beta_1 \partial \beta_0} & \frac{\partial^2 l(\beta)}{\partial \beta_1^2} & \dots & \frac{\partial^2 l(\beta)}{\partial \beta_1 \partial \beta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 l(\beta)}{\partial \beta_p \partial \beta_0} & \frac{\partial^2 l(\beta)}{\partial \beta_p \partial \beta_1} & \dots & \frac{\partial^2 l(\beta)}{\partial \beta_p^2} \end{bmatrix}$$

dimana $H(\hat{\beta}_{(m)})$ adalah matriks berukuran $(p + 1)(p + 1)$ yang elemen-elemennya merupakan turunan kedua yang diturunkan secara parsial dari $l(\beta)$.

- 5) Menentukan taksiran dari β dengan menggunakan iterasi ke- $m + 1$ dengan ($m = 1, 2, \dots$) yaitu $\hat{\beta}_{(m+1)}$ dan persamaan iterasi sebagai berikut:

$$\hat{\beta}_{(m+1)} = \hat{\beta}_{(m)} - [H(\hat{\beta}_{(m)})]^{-1} g(\hat{\beta}_{(m)})$$

dengan $\hat{\beta}_{(m+1)}$ adalah vektor taksiran parameter dari β pada iterasi ke- $m + 1$ dan $\hat{\beta}_{(m)}$ adalah vektor taksiran parameter dari β pada iterasi ke- m .

- 6) Proses iterasi akan berhenti jika $\hat{\beta}_{(m+1)} \approx \hat{\beta}_{(m)}$ dimana $\|\hat{\beta}_{(m)} - \hat{\beta}_{(m+1)}\| < \varepsilon$

(misal $\varepsilon = 10^{-5}$), sehingga $\hat{\beta}_{(m+1)}$ dikatakan sebagai estimasi dari $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$.

1.6.4 Uji Overdispersi

Dalam pemodelan regresi Poisson, salah satu asumsi utama yang harus dipenuhi adalah asumsi ekuidispersi, yaitu kondisi di mana rata-rata dari variabel respon (Y) sama dengan variansinya. Namun, dalam banyak kasus, asumsi ini tidak terpenuhi, sehingga terjadi overdispersi. Overdispersi merupakan kondisi di mana varians dari data variabel respon lebih besar dibandingkan dengan nilai rata-ratanya (Saputro et al., 2021). Keberadaan overdispersi dalam regresi Poisson dapat berdampak serius terhadap hasil analisis, terutama dalam interpretasi model. Hal ini dapat menyebabkan estimasi *standar error* yang terlalu rendah, sehingga menghasilkan kesimpulan yang keliru terkait signifikansi parameter regresi (Nurfajrin et al., 2023). Oleh karena itu, diperlukan pengujian khusus untuk mendeteksi adanya overdispersi dalam model.

Berdasarkan penelitian yang dilakukan oleh Dewi et al. (2023) dan Darnah (2011), langkah-langkah dalam melakukan uji overdispersi adalah sebagai berikut:

1. Hipotesis

H_0 : Tidak terjadi Overdispersi

H_1 : Terjadi Overdispersi

2. Statistik Uji

Statistik uji yang digunakan adalah:

$$\theta = \frac{D^2}{df} ; D^2 = 2 \sum_{i=1}^n y_i \ln \left(\frac{y_i}{\mu_i} \right) \quad (16)$$

dengan

θ : parameter dispersi

D^2 : nilai devians

df : derajat kebebasan ($df = n - p$) dimana n adalah banyaknya pengamatan dan p adalah jumlah parameter termasuk konstanta)

y_i : nilai variabel respon untuk pengamatan ke- i

μ_i : taksiran rata-rata dari banyaknya kasus pada pengamatan ke- i

3. Kriteria Uji

Tolak H_0 apabila nilai $\theta > 1$ yang berarti data mengalami overdispersi.

1.6.5 Model Generalized Poisson Regression (GPR)

Model *Generalized Poisson Regression* (GPR) merupakan model regresi yang dikembangkan untuk mengatasi masalah overdispersi maupun underdispersi dalam data *count*. Berbeda dengan regresi Poisson standar yang mengasumsikan ekuidispersi, model GPR memperkenalkan parameter dispersi θ sebagai tambahan selain parameter rata-rata μ . Misal $y = 0, 1, 2, \dots$ merupakan variabel respon, maka fungsi distribusi *Generalized Poisson* (GP) dapat dinyatakan sebagai berikut (Famoye et al., 2004):

$$f(y_i, \mu_i; \theta) = \left(\frac{\mu_i}{1 + \theta \mu_i} \right)^{y_i} \frac{(1 + \theta y_i)^{y_i-1}}{y_i!} \exp \left\{ -\frac{\mu_i(1 + \theta y_i)}{1 + \theta \mu_i} \right\} \quad (17)$$

Dalam model GPR, nilai rata-rata dan varians dari variabel respon dinyatakan sebagai:

$$E(y_i) = \mu_i \text{ dan } var(y_i) = \mu_i(1 + \theta \mu_i)^2 \quad (18)$$

Model GPR memiliki bentuk yang serupa dengan regresi Poisson, tetapi dengan asumsi bahwa komponen acak (*random component*) mengikuti distribusi GP. Dengan kata lain, model ini tidak memerlukan asumsi ekuidispersi, sehingga lebih fleksibel dalam menganalisis data *count* yang memiliki variabilitas lebih besar. Persamaan model GPR dapat dinyatakan sebagai berikut (Ismail & Jemain, 2007):

$$\mu_i = \exp(x_i^T \beta) \quad (19)$$

1.6.6 Penaksiran Parameter Model *Generalized Poisson Regression* (GPR)

Untuk menaksir parameter dalam Model GPR, digunakan metode *Maximum Likelihood Estimation* (MLE). Metode MLE merupakan pendekatan yang digunakan untuk mencari nilai parameter yang memaksimalkan kemungkinan (*likelihood*) bahwa data yang diamati berasal dari distribusi tertentu. Fungsi *likelihood* untuk model GPR adalah sebagai berikut:

$$\begin{aligned} L(\beta, \theta) &= \prod_{i=1}^n f(y_i | x_{i1}, x_{i2}, \dots, x_{ip}; \beta_0, \beta_1, \dots, \beta_p, \theta) \\ &= \prod_{i=1}^n \left(\frac{\mu_i}{1 + \theta \mu_i} \right)^{y_i} \frac{(1 + \theta y_i)^{y_i - 1}}{y_i!} \exp \left(- \frac{\mu_i (1 + \theta y_i)}{1 + \theta \mu_i} \right) \end{aligned} \quad (20)$$

Untuk mempermudah perhitungan, digunakan fungsi *log likelihood* sebagai berikut:

$$\begin{aligned} l(\beta, \theta) &= \log L(\beta, \theta) \\ l(\beta, \theta) &= \sum_{i=1}^n \left[y_i \log \left(\frac{\mu_i}{1 + \theta \mu_i} \right) + (y_i - 1) \log(1 + \theta y_i) - \frac{\mu_i (1 + \theta y_i)}{1 + \theta y_i} - \log y_i! \right] \\ l(\beta, \theta) &= \sum_{i=1}^n \left[y_i (\beta_0 + \sum_{j=1}^p \beta_j x_{ji}) - y_i \log(1 + \theta \exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ji})) \right. \\ &\quad \left. + (y_i - 1) \log(1 + \theta y_i) - \frac{\exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ji}) (1 + \theta y_i)}{1 + \theta \exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ji})} - \log y_i! \right] \end{aligned} \quad (21)$$

Untuk mencari taksiran maksimum *likelihood* $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p, \theta$, fungsi *log likelihood* pada Persamaan (21) diturunkan secara parsial terhadap parameter $\beta_0, \beta_1, \dots, \beta_p, \theta$ dan disamakan dengan nol. Didapatkan persamaan sebagai berikut:

$$\begin{aligned} \bullet \quad \frac{\partial l(\beta, \theta)}{\partial \beta_0} &= 0 \\ \sum_{i=1}^n \left[y_i - y_i \frac{\theta \exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ij})}{1 + \theta \exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ij})} - \frac{(1 + \theta y_i) \exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ij})}{1 + \theta \exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ij})} \right] &= 0 \\ \sum_{i=1}^n \left[\frac{y_i (1 + \theta \exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ij})) - (1 + \theta y_i) \exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ij})}{(1 + \theta \exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ij}))^2} \right] &= 0 \\ \sum_{i=1}^n \left[\frac{(y_i - \exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ij}))}{(1 + \theta \exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ij}))^2} \right] &= 0 \end{aligned} \quad (22)$$

$$\begin{aligned} \bullet \quad \frac{\partial l(\beta, \theta)}{\partial \beta_1} &= 0 \\ \sum_{i=1}^n \left[y_i x_{i1} - y_i x_{i1} \frac{\theta \exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ij})}{1 + \theta \exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ij})} - \frac{(1 + \theta y_i) \exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ij})}{1 + \theta \exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ij})} \right] &= 0 \end{aligned}$$

$$\begin{aligned}
& \sum_{i=1}^n \left[y_i x_{i1} \left(1 - \frac{\theta \exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ij})}{1 + \theta \exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ij})} \right) - \left(\frac{(1 + \theta y_i) \exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ij})}{1 + \theta \exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ij})} \right) \right] = 0 \\
& \sum_{i=1}^n \left[\frac{y_i x_{i1}}{1 + \theta \exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ij})} - \frac{(1 + \theta y_i) \exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ij})}{1 + \theta \exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ij})} \right] = 0 \\
& \sum_{i=1}^n \left[\frac{y_i x_{i1} (1 + \theta \exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ij})) - (1 + \theta y_i) \exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ij})}{\left(1 + \theta \exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ij}) \right)^2} \right] = 0 \\
& \sum_{i=1}^n \left[\frac{x_{i1} (y_i - \exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ij}))}{\left(1 + \theta \exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ij}) \right)^2} \right] = 0
\end{aligned} \tag{23}$$

Dengan menggunakan cara yang sama maka akan diperoleh:

$$\bullet \quad \frac{\partial l(\beta, \theta)}{\partial \beta_k} = \sum_{i=1}^n \left[\frac{x_{ik} (y_i - \exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ij}))}{\left(1 + \theta \exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ij}) \right)^2} \right] = 0, \text{ untuk } k = 2, \dots, p \tag{24}$$

Turunan terhadap θ diperoleh:

$$\bullet \quad \frac{\partial l(\beta, \theta)}{\partial \theta} = 0$$

$$\frac{\partial l(\beta, \theta)}{\partial \theta} = \sum_{i=1}^n \left[\frac{-\frac{y_i \exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ij})}{1 + \theta \exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ij})} + \frac{y_i (y_i - 1)}{1 + \theta y_i} - \frac{\exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ij}) (y_i - \exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ij}))}{\left(1 + \theta \exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ij}) \right)^2}}{\left(1 + \theta \exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ij}) \right)^2} \right] = 0 \tag{25}$$

Persamaan-persamaan di atas merupakan bentuk persamaan eksponensial, sehingga persamaan-persamaan tersebut tidak linear dalam parameter, sehingga untuk mencari taksirannya akan dilakukan pendekatan dengan menggunakan metode *Newton-Raphson* (Islamiyah, 2020).

1.6.7 Multivariate Adaptive Regression Spline (MARS)

Multivariate Adaptive Regression Spline (MARS) merupakan metode regresi nonparametrik yang dikembangkan oleh Friedman (1991) untuk menangani data berdimensi tinggi serta mengatasi diskontinuitas tanpa memerlukan asumsi bentuk hubungan antara variabel prediktor dan variabel respon. Metode ini sangat fleksibel dalam mengidentifikasi pola hubungan antar variabel serta dapat digunakan untuk analisis regresi maupun klasifikasi (Rahmaniah et al., 2016).

Menurut Pintowati dan Otok (2012), terdapat beberapa komponen penting dalam model MARS yang perlu diperhatikan, di antaranya:

- Knot adalah titik pada *domain* variabel prediktor yang digunakan untuk membagi data menjadi beberapa *subregion*. Fungsi basis (*basis function*) diterapkan di sekitar titik-titik ini untuk menangkap perubahan pola hubungan antar variabel. Di setiap titik knot, diharapkan adanya kontinuitas fungsi basis agar model tetap halus di seluruh domain variabel. Friedman (1991) menyarankan agar minimum observasi antara knot (MO) adalah 0, 1, 2, dan 3.

- b. *Basis Functions* (BF) adalah fungsi yang digunakan untuk menggambarkan hubungan antara variabel respon dan variabel prediktor. Fungsi basis ini umumnya berbentuk polinomial dengan turunan yang kontinu pada setiap titik knot. Friedman (1991) menyarankan agar jumlah maksimum *basis function* (BF) adalah sekitar 2-4 kali jumlah variabel prediktor.
- c. Interaksi (*Interaction*) adalah hasil perkalian silang antara variabel yang saling berkorelasi. Friedman (1991) menyarankan bahwa jumlah maksimum interaksi (MI) yang diperbolehkan adalah 1, 2, atau 3. Apabila MI lebih besar dari 3, maka model yang dihasilkan akan semakin kompleks dan sulit untuk diinterpretasikan.

MARS merupakan pengembangan dari pendekatan *Recursive Partition Regression* (RPR). Pendekatan RRR masih memiliki kelemahan dimana model yang dihasilkan tidak kontinu pada knot sehingga model MARS digunakan untuk mengatasi kelemahan RRR yaitu menghasilkan model yang kontinu pada knot dan dapat mengidentifikasi adanya fungsi linear aditif (Damayanti & Sunendiari, 2021). Perbaikan yang dilakukan untuk mengatasi keterbatasan RRR menghasilkan basis fungsi sebagai berikut (Sita, 2015):

$$B_{mi}(x_i) = \prod_{k=1}^{k_m} [S_{kM} \cdot (x_{v(k,m)} - t_{kM})]_+ \quad (26)$$

Hasil modifikasi fungsi basis pada model RPR dengan kombinasi spline, diperoleh model MARS sebagai berikut:

$$\begin{aligned} \hat{f}(x_i) &= a_0 + \sum_{m=1}^M a_m B_{mi}(x_i) \\ &= a_0 + \sum_{m=1}^M a_m \prod_{k=1}^{k_m} [S_{kM} \cdot (x_{v(k,m)} - t_{kM})]_+ \end{aligned} \quad (27)$$

dengan

- a_0 : konstanta regresi dari fungsi basis
- a_m : koefisien dari fungsi basis ke- m , $m = 1, 2, \dots, M$
- M : maksimum fungsi basis (*nonconstant* fungsi basis)
- k_m : derajat interaksi untuk fungsi basis ke- m
- S_{kM} : nilai berupa +1 atau -1 yang menunjukkan positif relatif data terhadap titik knot. $S_{kM} = +1$, jika data berada disebelah kanan knot ($X_{v(k,M)} \geq t_{kM}$) dan $S_{kM} = -1$, jika data berada disebelah kiri titik knot ($X_{v(k,M)} < t_{kM}$)
- $x_{v(k,M)}$: variabel prediktor
- t_{kM} : nilai *knots* dari variabel prediktor $X_{v(k,M)}$
- v : banyaknya variabel bebas
- k : banyaknya interaksi
- m : banyaknya basis fungsi
- i : banyaknya pengamatan ($1, 2, \dots, n$)

Berdasarkan Persamaan (27) model MARS dapat dijabarkan menjadi:

$$\hat{f}(x_i) = a_0 + \sum_{m=1}^M a_m (S_{1m}(x_{v(1,m)} - t_{1m})) + \sum_{m=1}^M a_m (S_{1m}(x_{v(1,m)} - t_{1m})) (S_{2m}(x_{v(2,m)} - t_{2m})) + \sum_{m=1}^M a_m (S_{1m}(x_{v(1,m)} - t_{1m})) + (S_{2m}(x_{v(2,m)} - t_{2m})) (S_{2m}(x_{v(2,m)} - t_{2m})) + \dots \quad (28)$$

Dalam bentuk matriks, persamaan MARS dapat dinyatakan sebagai berikut:

$$\mathbf{Y} = \mathbf{B}\mathbf{a} + \boldsymbol{\varepsilon} \quad (29)$$

dimana $\mathbf{Y}$ adalah variabel respon yang berukuran $(n \times 1)$, $\mathbf{B} = [1, (X_{v(k,m)} - t_{km})_1^K]$ merupakan basis fungsi yang berukuran $(n \times (M + 1))$, $\mathbf{a}$ merupakan koefisien regresi berukuran $((M + 1) \times 1)$ dan $\boldsymbol{\varepsilon}$ merupakan *error* yang berukuran $(n \times 1)$.

dengan:

$\mathbf{Y} = (y_1, y_2, y_3, \dots, y_n)^T$, $\mathbf{a} = (a_0, \dots, a_M)^T$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^T$, dan Matriks $\mathbf{B}$ dapat ditulis sebagai berikut:

$$\mathbf{B} = \begin{bmatrix} 1 & \prod_{k=1}^{k_1} [s_{k1}(X_{v(k,1)1} - t_{k1})]_+ & \dots & \prod_{k=1}^{k_M} [s_{kM}(X_{v(k,M)1} - t_{kM})]_+ \\ 1 & \prod_{k=1}^{k_1} [s_{k1}(X_{v(k,1)2} - t_{k1})]_+ & \dots & \prod_{k=1}^{k_M} [s_{kM}(X_{v(k,M)2} - t_{kM})]_+ \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \prod_{k=1}^{k_1} [s_{k1}(X_{v(k,1)n} - t_{k1})]_+ & \dots & \prod_{k=1}^{k_M} [s_{kM}(X_{v(k,M)n} - t_{kM})]_+ \end{bmatrix} \quad (30)$$

1.6.8 Penaksiran Parameter Model *Multivariate Adaptive Regression Spline* (MARS)

Model MARS pada persamaan (29), memiliki parameter $\mathbf{a}$ yang perlu diestimasi. Untuk memperoleh estimator $\hat{\mathbf{a}}$, digunakan metode kuadrat terkecil, yang pada prinsipnya meminimumkan *Sum Square Error* (SSE), dinyatakan sebagai berikut:

$$\mathbf{Y} = \mathbf{B}\mathbf{a} + \boldsymbol{\varepsilon}$$

$$\boldsymbol{\varepsilon} = \mathbf{Y} - \mathbf{B}\mathbf{a} \quad (31)$$

$$\begin{aligned} SSE &= \sum_{i=1}^n \varepsilon_i^2 = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} = (\mathbf{Y} - \mathbf{B}\mathbf{a})^T (\mathbf{Y} - \mathbf{B}\mathbf{a}) \\ &= \mathbf{Y}^T \mathbf{Y} - \mathbf{a}^T \mathbf{B}^T \mathbf{Y} - \mathbf{Y}^T \mathbf{B}\mathbf{a} + \mathbf{a}^T \mathbf{B}^T \mathbf{B}\mathbf{a} \\ &= \mathbf{Y}^T \mathbf{Y} - 2\mathbf{a}^T \mathbf{B}^T \mathbf{Y} + \mathbf{a}^T \mathbf{B}^T \mathbf{B}\mathbf{a} \end{aligned} \quad (32)$$

Dengan menurunkan SSE secara parsial terhadap $\mathbf{a}$, diperoleh persamaan normal sebagai berikut:

$$\begin{aligned} \frac{\partial SSE}{\partial \mathbf{a}} &= -2\mathbf{B}^T \mathbf{Y} + 2\mathbf{B}^T \mathbf{B}\mathbf{a} = 0 \\ -\mathbf{B}^T \mathbf{Y} + \mathbf{B}^T \mathbf{B}\mathbf{a} &= 0 \\ \mathbf{B}^T \mathbf{B}\mathbf{a} &= \mathbf{B}^T \mathbf{Y} \end{aligned} \quad (33)$$

Karena $\mathbf{B}$ merupakan matriks *nonsingular*, maka estimasi $\mathbf{a}$ dapat diperoleh (Otok et al., 2008):

$$\hat{\mathbf{a}} = (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \mathbf{Y} \quad (34)$$

1.6.9 Pemilihan Model MARS

Pemilihan model terbaik dalam *Multivariate Adaptive Regression Splines* (MARS) dilakukan dengan dua pendekatan utama. Pendekatan pertama menggunakan metode *forward stepwise* untuk menentukan jumlah fungsi basis yang optimal.

Pendekatan kedua adalah *backward stepwise*, yang bertujuan untuk memilih basis fungsi terbaik dari model yang dihasilkan oleh *forward stepwise* sambil meminimalkan nilai *Generalized Cross Validation* (GCV), sebagaimana dijelaskan oleh Friedman (1991).

Pada tahap *forward stepwise*, model dibangun secara bertahap dengan menambahkan fungsi basis satu per satu berdasarkan kontribusinya dalam mengurangi kesalahan model. Namun, metode ini dapat menghasilkan model dengan banyak fungsi basis yang mungkin terlalu kompleks dan rentan terhadap *overfitting*. Oleh karena itu, metode *backward stepwise* digunakan untuk mengurangi kompleksitas model dengan menghapus fungsi basis yang kurang signifikan.

Dalam praktiknya, jumlah maksimum fungsi basis sering dibatasi untuk menghindari kompleksitas berlebih. Selain itu, derajat interaksi antara variabel prediktor juga biasanya dibatasi hingga derajat tiga (Dukalang, 2017). Model terbaik dipilih berdasarkan nilai GCV terkecil. GCV merupakan sebuah ukuran untuk menilai kualitas model yang dapat dihitung dengan rumus sebagai berikut (Wicaksono et al., 2014):

$$GCV = \frac{MSE}{\left[1 - \frac{C(\tilde{M})}{n}\right]^2}; \quad MSE = \frac{1}{n} \sum_{i=1}^n [y_i - \hat{f}_M(x_i)]^2 \quad (35)$$

dimana $C(\tilde{M}) = C(M) + d \cdot M$ dan $\hat{f}_M(x_i) = \hat{\alpha}_0 + \sum_{m=1}^M \hat{\alpha}_m \prod_{k=1}^{K_m} [s_{km}(x_{v(k,m)i} - t_{km})]^+$ dengan

- MSE : kuadrat tengah eror
- y_i : variabel respon ke- i
- M : jumlah fungsi basis
- x_i : variabel prediktor ke- i
- n : banyaknya pengamatan
- d : derajat interaksi
- $C(\tilde{M})n$: fungsi kompleks
- $C(M)$: $trace[(\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T] + 1$
- $\hat{f}_M(x_i)$: nilai taksiran variabel respon pada fungsi basis M .

Model MARS terbaik juga dapat dipilih berdasarkan nilai GCV terkecil dan *R-Square* terbesar (Prastika et al., 2021). *R-Square* (R^2) merupakan salah satu ukuran yang digunakan untuk menilai kebaikan model. Nilai ini menunjukkan seberapa besar proporsi variasi dari variabel respon (Y) yang dapat dijelaskan oleh variabel prediktor (X) dalam model. R^2 didefinisikan sebagai berikut (Prastika et al., 2021):

$$R^2 = 1 - \frac{SSE}{SST} \quad (36)$$

1.6.10 Multivariate Adaptive Generalized Poisson Regression Spline (MAGPRS)

Model MAGPRS adalah metode pengembangan dari MARS. Model ini merupakan gabungan dari metode *Generalized Poisson Regression* (GPR) dengan MARS.

Variabel responnya berdistribusi generalized poisson $Y_i \sim GP(\mu_i, \theta)$ dengan $E(Y) = \mu$, dan $Var(Y) = \mu(1 + \theta\mu)^2$ (Hidayati et al. (2019). Asumsinya dapat ditulis sebagai:

$$\mathbf{y} = \exp(\mathbf{B}\boldsymbol{\alpha}) + \boldsymbol{\varepsilon} \quad 37$$

dimana $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$, $\mathbf{a} = (a_0, a_1, \dots, a_M)^T$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^T$

Sebuah fungsi kepadatan atau massa dikatakan termasuk dalam keluarga eksponensial jika dapat dituliskan dalam bentuk (Otok et al., 2019):

$$P(y|\eta) = h(y) \exp\{\eta^T T(y) - A(\eta)\} \quad (38)$$

dengan

$h(y)$: fungsi yang hanya bergantung pada y ,

η : natural parameter,

$T(y)$: *sufficient statistics*,

$A(\eta)$: fungsi partisi log.

Distribusi Generalized Poisson (GP) dengan parameter μ dan θ , dalam bentuk eksponensial dituliskan sebagai:

$$\begin{aligned} P(y|\eta) &= \exp\left\{\log\left(\left(\frac{\mu}{1+\theta\mu}\right)^y \frac{(1+\theta y)^{y-1}}{y!} \exp\left(-\frac{\mu(1+\theta y)}{1+\theta y}\right)\right)\right\} \\ &= \exp\left\{\log\left(\frac{\mu}{1+\theta\mu}\right)^y + \log\frac{(1+\theta y)^{y-1}}{y!} + \log\left(\exp\left(-\frac{\mu(1+\theta y)}{1+\theta y}\right)\right)\right\} \\ &= \frac{1}{y!} \frac{(1+\theta y)^{y-1}}{(1+\theta\mu)^y} \exp\left\{y \log \mu - \frac{\mu(1+\theta y)}{1+\theta y}\right\} \end{aligned} \quad (39)$$

dimana $\mu = \exp\{\eta\}$ dan $\eta = \log \mu$, $\eta = \mathbf{B}\boldsymbol{\alpha}$

Berikut model MAGPRS, yaitu:

$$\begin{aligned} \mu &= \exp\{\eta\} \\ \mu &= \exp(\mathbf{B}\boldsymbol{\alpha}) \\ \mu &= \exp\left(\alpha_0 + \sum_{m=1}^M \alpha_m B_{mi}(x_i)\right) \end{aligned} \quad (40)$$

1.6.11 Pengujian Parameter Model *Multivariate Adaptive Generalized Poisson Regression Spline* (MAGPRS)

Pengujian simultan parameter model MAGPRS dilakukan dengan menggunakan *Maximum Likelihood Ratio Test* (MLRT). Adapun hipotesis yang digunakan dalam pengujian ini adalah sebagai berikut:

$$H_0 : \alpha_1 = \dots = \alpha_M = 0$$

$$H_1 : \text{paling sedikit terdapat satu } \alpha_m \neq 0, m = 1, \dots, M$$

Statistik uji yang digunakan, yaitu (Otok et al., 2019):

$$\hat{G}^2 = -2 \ln \Lambda = 2(\ln L(\hat{\Omega}) - \ln L(\hat{\omega})) \quad (41)$$

dimana

$$L(\hat{\Omega}) = \max L(\Omega)$$

$$\ln L(\hat{\Omega}) = \sum_{i=1}^n \{y_i(\mathbf{B}\hat{\alpha}) - y_i \ln(1 + \hat{\theta} \exp(\mathbf{B}\hat{\alpha})) + (y_i - 1) \ln(1 + \hat{\theta} y_i) - \ln(y_i!) - \exp(\mathbf{B}\hat{\alpha})(1 + \hat{\theta} y_i)(1 + \hat{\theta}(\mathbf{B}\hat{\alpha}))^{-1}\} \quad (42)$$

dan

$$L(\hat{\omega}) = \max L(\omega)$$

$$\ln L(\hat{\omega}) = \sum_{i=1}^n \{y_i \ln(\hat{\alpha}_0) - y_i \ln(1 + \hat{\theta} \exp(\hat{\alpha}_0)) + (y_i - 1) \ln(1 + \hat{\theta} y_i) - \ln(y_i!) - \exp(\hat{\alpha}_0)(1 + \hat{\theta} y_i)(1 + \hat{\theta}(\hat{\alpha}_0))^{-1}\} \quad (43)$$

Sehingga:

$$\hat{G}^2 = 2 \left[\begin{aligned} &\sum_{i=1}^n y_i(\mathbf{B}\hat{\alpha}) - \sum_{i=1}^n y_i \ln(\hat{\alpha}_0) + \sum_{i=1}^n y_i \frac{\ln(1 + \hat{\theta} \exp(\hat{\alpha}_0))}{\ln(1 + \hat{\theta} \exp(\mathbf{B}\hat{\alpha}))} \\ &- \sum_{i=1}^n \exp(\mathbf{B}\hat{\alpha})(1 + \hat{\theta} y_i)(1 + \hat{\theta} \exp(\mathbf{B}\hat{\alpha}))^{-1} \\ &+ \sum_{i=1}^n \exp(\hat{\alpha}_0)(1 + \hat{\theta} y_i)(1 + \hat{\theta} \exp(\hat{\alpha}_0))^{-1} \end{aligned} \right] \quad (44)$$

Karena $\hat{G}^2$ berdistribusi X^2 , maka $\hat{G}^2 \sim X^2_{\alpha, v}$ dimana $v = n(\Omega) - n(\omega)$. Dengan keputusan, tolak H_0 jika nilai statistik uji $\hat{G}^2 > X^2_{\alpha, v}$. Setelah pengujian serentak, dilakukan pengujian secara parsial untuk mengetahui variabel prediktor mana yang memiliki pengaruh signifikan terhadap variabel respon. Pengujian ini dilakukan untuk setiap koefisien α_m secara individu.

Hipotesis untuk pengujian parsial adalah sebagai berikut.

$$H_0 : \alpha_m = 0$$

$$H_1 : \alpha_m \neq 0, m = 1, \dots, M$$

Statistik uji yang digunakan adalah sebagai berikut:

$$t_{hit} = \left(\frac{\hat{\alpha}_m}{se(\hat{\alpha}_m)} \right)^2; se(\hat{\alpha}_m) = \sqrt{\left(\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-k-1} \right) C_{mm}} \quad (45)$$

dengan

$\hat{\alpha}_m$: estimasi koefisien α_m

$se(\hat{\alpha}_m)$: kesalahan standar dari α_m

n : jumlah observasi

k : jumlah parameter dalam model

C_{mm} : elemen-elemen pada diagonal utama matriks

y_i : nilai observasi

$\hat{y}_i$: nilai prediksi untuk observasi ke- i

Dengan keputusan, tolak H_0 apabila $|t_{hit}| > t_{\frac{\alpha}{2}, n-k}$ atau $p - value < \alpha$ yang

berarti bahwa semua fungsi basis yang terpilih dalam model *Multivariate Adaptive Generalized Poisson Regression Spline* (MAGPRS) memiliki pengaruh terhadap variabel respon (Prastika et al., 2021).

1.6.12 Maximum Likelihood Estimasion (MLE)

Maximum Likelihood Estimation (MLE) adalah metode statistik yang digunakan untuk mengestimasi parameter suatu distribusi probabilitas dengan memaksimalkan *likelihood function* berdasarkan data yang diamati. Metode ini bertujuan untuk menemukan nilai parameter yang membuat data yang diamati paling mungkin terjadi (Rossi, 2018).

Misalkan $X_1, X_2, \dots, X_n$ adalah sampel acak independen dari distribusi probabilitas dengan fungsi densitas (atau fungsi massa probabilitas) $f(x; \theta)$, dimana θ adalah parameter yang ingin diestimasi. Fungsi *Likelihood* didefinisikan sebagai:

$$L(\theta) = \prod_{i=1}^n f(x_i; \theta) \quad (46)$$

Untuk kemudahan analisis, biasanya digunakan *log-likelihood*:

$$l(\theta) = \log L(\theta) = \sum_{i=1}^n \log f(x_i; \theta) \quad (47)$$

Dalam prinsip MLE, MLE mencari nilai $\hat{\theta}$ yang memaksimalkan *log-likelihood*:

$$\hat{\theta} = \arg \max_{\theta} l(\theta) \quad (48)$$

Untuk mencarinya, dilakukan turunan pertama dari *log-likelihood* terhadap parameter dan disamadengankan dengan nol:

$$\frac{\partial l(\theta)}{\partial \theta} = 0 \quad (49)$$

Nilai θ yang memenuhi ini adalah kandidat estimator maksimum *likelihood*. Kemudian dicek apakah titik tersebut adalah maksimum dengan melihat turunan keduanya.

1.6.13 Kematian Ibu

Menurut Organisasi Kesehatan Dunia (WHO), kematian ibu didefinisikan sebagai kematian seorang wanita selama kehamilan dan atau periode 42 hari setelah berakhirnya kehamilan, akibat semua yang terkait dengan atau diperberat oleh kehamilan atau penanganannya, tetapi bukan disebabkan oleh kecelakaan atau cedera. Kematian ibu merupakan salah satu indikator yang menggambarkan kesejahteraan masyarakat di suatu negara. Selain itu, kematian ibu juga merupakan indikator derajat kesehatan perempuan. Jumlah kematian ibu dihitung dari kematian yang terjadi selama masa kehamilan, persalinan, dan nifas atau pengelolaannya, tetapi bukan disebabkan oleh sebab-sebab lain seperti kecelakaan atau terjatuh. Angka Kematian Ibu (AKI) merujuk pada jumlah kematian ibu yang terjadi di setiap 100.000 kelahiran hidup.

Adapun faktor-faktor yang mempengaruhi kematian Ibu pada penelitian ini, yaitu:

1. Persentase Ibu Hamil Melaksanakan Program K4

Program K4 mencakup empat kali kunjungan antenatal selama kehamilan, yang bertujuan untuk mendeteksi potensi masalah kesehatan sejak dini. Kunjungan ini memberi kesempatan bagi tenaga medis untuk memonitor kondisi ibu hamil, mengidentifikasi faktor risiko, serta memberikan penanganan lebih awal jika

diperlukan. Kunjungan yang rendah (rendahnya cakupan K4) mengurangi kesempatan untuk mendeteksi dan menangani masalah serius seperti hipertensi, anemia, atau gangguan obstetrik lainnya. Tanpa pemantauan yang memadai, komplikasi yang tidak terdeteksi dapat menyebabkan kematian ibu. Oleh karena itu, rendahnya angka kunjungan K4 berkaitan langsung dengan tingginya risiko kematian ibu (Indriyani & Puspitasari, 2016).

2. Persentase Ibu Hamil Melaksanakan Program K1

Kunjungan pertama kali ibu hamil (K1) adalah pemeriksaan awal yang sangat penting untuk menentukan kondisi kesehatan ibu dan janin. Melalui pemeriksaan K1, tenaga medis dapat mengidentifikasi faktor risiko seperti anemia, infeksi, atau penyakit kronis yang dapat memperburuk kondisi selama kehamilan dan persalinan. Pemeriksaan yang dilakukan secara rutin dan sesuai standar dapat menurunkan risiko komplikasi, sehingga mengurangi kemungkinan terjadinya kematian ibu. Pemeriksaan yang terlambat atau tidak memadai dapat mengakibatkan risiko tinggi yang tidak terdeteksi, yang dapat berujung pada kematian ibu (Ayu et al., 2022).

3. Persentase Persalinan di Fasilitas Layanan Kesehatan

Menjalani persalinan di fasilitas kesehatan dengan tenaga medis terlatih sangat penting untuk mengurangi risiko komplikasi persalinan, seperti perdarahan, infeksi, atau obstruksi. Kehamilan dan persalinan yang tidak ditangani dengan baik dapat menyebabkan komplikasi yang berpotensi fatal. Fasilitas kesehatan dengan tenaga medis yang terlatih mampu memberikan penanganan yang cepat dan tepat jika terjadi komplikasi. Oleh karena itu, semakin tinggi persentase persalinan yang terjadi di fasilitas kesehatan, semakin rendah risiko kematian ibu, karena komplikasi dapat ditangani dengan lebih efektif (Perdana et al., 2023).

4. Persentase Ibu Hamil Mendapat Td3

Imunisasi Td (Tetanus Diphtheria) merupakan vaksin lanjutan dari vaksin DT dengan dosis ke-3 yang direkomendasikan untuk diberikan pada ibu hamil di usia kehamilan 27-36 minggu atau segera setelah bayi lahir agar anak kebal akan penyakit infeksi. Imunisasi Td3 (Tetanus Diphtheria) memberikan perlindungan bagi ibu hamil dari risiko tetanus dan difteri, yang dapat berbahaya bagi ibu dan bayi. Selain itu, vaksin ini juga meningkatkan daya tahan tubuh ibu terhadap infeksi, yang sangat penting selama masa kehamilan dan persalinan. Tanpa imunisasi ini, ibu hamil lebih rentan terhadap infeksi yang dapat menyebabkan komplikasi serius dan meningkatkan risiko kematian. Dengan demikian, mendapatkan imunisasi Td3 adalah salah satu cara untuk mengurangi risiko kematian ibu akibat infeksi yang dapat dicegah (Jones et al., 2024).

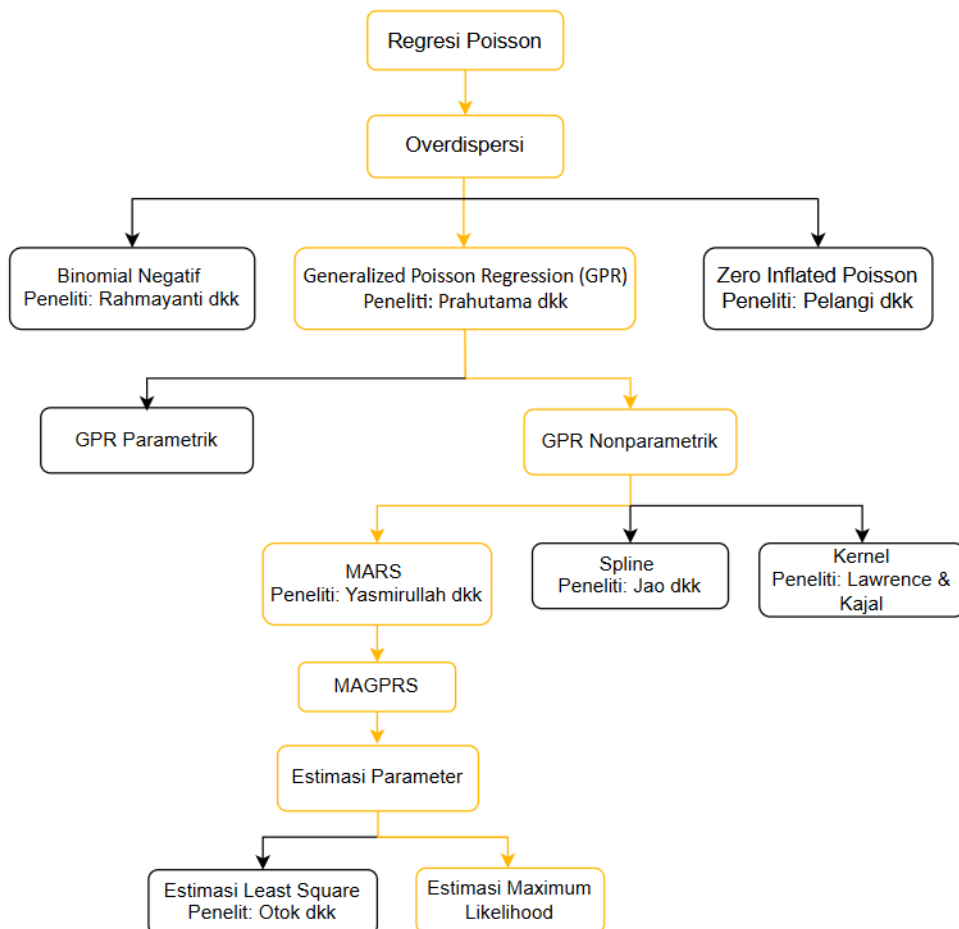
5. Persentase ibu nifas mendapat vitamin A

Ibu nifas adalah ibu yang baru melahirkan sampai 6 minggu setelah melahirkan/setelah kelahiran bayi (0-42 hari). Ibu nifas harus diberikan 2 kapsul vitamin A dosis tinggi untuk meningkatkan kandungan dalam ASI selama 60 hari. Pemberian kapsul vitamin A dosis tinggi dengan takaran 200.00 UI, kapsul pertama diberikan segera setelah melahirkan dan kapsul kedua diberikan pada hari berikutnya tidak lebih dari 6 minggu. Vitamin A berperan penting dalam menjaga kesehatan ibu nifas dan mencegah komplikasi setelah persalinan. Pemberian vitamin

A dosis tinggi dapat mencegah masalah penglihatan seperti rabun senja, serta meningkatkan kualitas ASI, yang berdampak pada kesehatan bayi. Selain itu, vitamin A juga mencegah anemia pada ibu nifas, yang jika dibiarkan, dapat menyebabkan kelelahan berlebihan dan mempengaruhi pemulihan ibu setelah melahirkan. Tanpa pemberian vitamin A yang cukup, risiko komplikasi pasca-persalinan bisa meningkat, yang pada akhirnya berkontribusi pada tingginya angka kematian ibu (Sari et al., 2023).

1.6.14 Kerangka Konseptual

Kerangka konsep dalam penelitian menggambarkan hubungan antara satu konsep dengan konsep lainnya yang terkait dengan masalah yang ingin diteliti. Dalam menyusun kerangka konsep, peneliti perlu memahami variabel yang akan diukur, karena kerangka konsep memberikan dasar konseptual bagi penelitian. Selain itu, kerangka konsep juga berfungsi untuk mengidentifikasi hubungan antara variabel-variabel yang dianggap relevan dengan studi masalah tertentu. Berikut ini adalah kerangka konsep dalam penelitian ini:



Gambar 1. Kerangka Konseptual

BAB II

METODE PENELITIAN

2.1 Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder, yaitu data kematian ibu beserta faktor-faktor yang diduga mempengaruhinya tahun 2021. Data ini bersumber dari publikasi oleh Dinas Kesehatan Provinsi Sulawesi Selatan melalui situs <https://dinkes.sulselprov.go.id/> yang diakses pada tanggal 14 Februari 2025 pukul 10.00 WITA.

2.2 Variabel Penelitian

Variabel penelitian dibagi menjadi 2 variabel yaitu variabel respon dan variabel prediktor. Variabel respon yaitu jumlah kematian ibu di provinsi Sulawesi Selatan dan variabel prediktor, yaitu persentase ibu hamil melaksanakan program K4 (X_1), persentase ibu hamil melaksanakan program K1 (X_2), persentase persalinan di layanan fasilitas kesehatan (X_3), persentase ibu nifas mendapat vitamin A (X_4), dan persentase ibu hamil mendapat Td3 (X_5) seperti pada Tabel 1 berikut ini:

Tabel 1. Variabel Penelitian

Variabel	Definisi	Skala
Jumlah kematian ibu (Y)	Jumlah kematian yang terjadi pada ibu sebelum mencapai usia satu tahun di tiap kabupaten/kota di Sulawesi Selatan.	Rasio
Persentase ibu hamil melaksanakan program K4 (X_1)	Jumlah ibu hamil yang memperoleh pelayanan antenatal K4 sesuai standar di satu wilayah kerja pada kurun waktu tertentu dibagi jumlah seluruh ibu hamil di satu wilayah kerja dalam kurun waktu yang sama dikali 100 persen.	Rasio
Persentase ibu hamil melaksanakan program K1 (X_2)	Jumlah ibu hamil yang memperoleh pelayanan antenatal K1 sesuai standar di satu wilayah kerja pada kurun waktu tertentu dibagi jumlah seluruh ibu hamil di satu wilayah kerja dalam kurun waktu yang sama dikali 100 persen	Rasio
Persentase persalinan di layanan fasilitas kesehatan (X_3)	Jumlah ibu bersalin yang ditolong oleh tenaga kesehatan yang memiliki kompetensi kebidanan (dokter kandungan dan kebidanan, dokter umum, dan bidan) di satu wilayah kerja pada kurun tertentu dibagi dengan jumlah ibu bersalin di satu wilayah kerja pada waktu yang sama dikali 100 persen.	Rasio
Persentase ibu nifas mendapat vitamin A (X_4)	Jumlah ibu nifas mendapat vitamin A pada wilayah dan kurun waktu tertentu dibagi dengan jumlah ibu nifas pada suatu wilayah dan kurun waktu yang sama dikali 100 persen.	Rasio

Variabel	Definisi	Skala
Persentase ibu hamil mendapat Td3 (X_5)	Jumlah ibu hamil yang mendapat imunisasi Td3 pada wilayah dan kurun waktu tertentu dibagi dengan jumlah ibu hamil yang mendapat imunisasi Td3/Td4 pada suatu wilayah dan kurun waktu yang sama dikali 100 persen.	Rasio

2.3 Metode Analisis Data

Metode analisis data yang digunakan dalam penelitian ini adalah *Multivariate Adaptive Generalized Poisson Regression Spline* (MAGPRS). Adapun langkah-langkah yang dilakukan adalah sebagai berikut:

1. Mengestimasi model MAGPRS dengan maximum likelihood terboboti dengan langkah-langkah sebagai berikut:
 - a. Menentukan model MAGPRS
 - b. Mencari nilai pembobot w_i
 - c. Mengestimasi parameter model dengan menggunakan *Maximum Likelihood* Terboboti, dengan cara sebagai berikut:
 - 1) Menentukan fungsi *likelihood* terbobot model MAGPRS
 - 2) Mencari fungsi *log-likelihood* terboboti model MAGPRS
 - 3) Mencari turunan pertama dan kedua fungsi *likelihood* terboboti model MAGPRS
 - d. Melakukan penyelesaian estimasi dengan metode iterasi Newton-Raphson.
2. Menerapkan model MAGPRS pada data Kematian Ibu di Provinsi Sulawesi Selatan dengan langkah-langkah sebagai berikut:
 - a. Menyajikan statistika deskriptif data jumlah kematian ibu beserta faktor-faktor yang diduga mempengaruhinya.
 - b. Melakukan pengujian distribusi Poisson dengan uji Kolmogorov-Smirnov
 - c. Melakukan pengujian overdispersi pada data.
 - d. Membuat *scatterplot* antara variabel respon dengan semua variabel prediktor secara parsial untuk melihat pola sebaran data.
 - e. Melakukan pemodelan MAGPRS dengan:
 - 1) Menentukan maksimum jumlah basis fungsi (BF). Dalam penelitian ini digunakan 5 variabel prediktor sehingga maksimum jumlah basis fungsi adalah 10, 15, dan 20.
 - 2) Menentukan jumlah maksimum interaksi, dalam penelitian ini jumlah maksimum interaksi (MI) yaitu 1, 2, dan 3.
 - 3) Menentukan minimum observasi (MO) yaitu 0, 1, 2 dan 3.
 - f. Melakukan estimasi model MAGPRS dan menentukan model MAGPRS terbaik berdasarkan nilai *Generalized Cross Validation* (GCV) yang terkecil yang diperoleh dari kombinasi antara BF, MI, dan MO.

- g. Menguji signifikansi model MAGPRS untuk mengevaluasi kecocokan model dengan uji koefisien regresi secara simultan (Uji F) maupun secara parsial (Uji t).
- h. Melakukan interpretasi model MAGPRS terbaik.
- i. Menarik kesimpulan.