

BAB I PENDAHULUAN

1.1 Latar Belakang

Tetanus neonatorum merupakan penyakit yang terjadi pada bayi yang berusia kurang dari 28 hari. Penyakit ini disebabkan oleh adanya bakteri yang bernama *Clostridium Tetani* yang dapat menyebabkan kematian. Faktor risiko yang dapat mempengaruhi terjadinya tetanus neonatorum terdiri dari faktor medis dan non medis. Faktor medis meliputi standar perawatan prenatal (kurangnya perawatan antenatal dan pengetahuan ibu hamil mengenai imunisasi tetanus toksoid), perawatan perinatal (kurang tersedianya fasilitas persalinan dan tenaga medis), dan perawatan neonatal (bayi yang baru lahir dalam keadaan tidak steril) sedangkan untuk faktor non medis berhubungan dengan adat istiadat masyarakat setempat (Handoko, 2011). Adapun upaya yang dapat dilakukan untuk mengurangi risiko terjadinya tetanus neonatorum yaitu peningkatan imunisasi tetanus toksoid, kunjungan antenatal, dan peningkatan fasilitas pelayanan kesehatan (Rusdiana, 2017). Oleh karena itu, agar upaya tersebut tepat guna, maka perlu dilakukan pemodelan hubungan jumlah kasus tetanus neonatorum dengan faktor-faktor yang diduga mempengaruhinya.

Analisis regresi merupakan salah satu metode yang digunakan untuk memodelkan hubungan antara variabel respon dengan satu atau lebih variabel prediktor. Analisis regresi terbagi menjadi analisis regresi linier dan analisis regresi non linier. Analisis regresi linier merupakan analisis regresi yang digunakan untuk menganalisis variabel respon yang berupa data kontinu dan mengikuti distribusi normal. Regresi linier juga membentuk hubungan antara variabel prediktor terhadap variabel respon yang bentuk fungsinya linier. Namun, pada beberapa data sering ditemukan data yang tidak memenuhi asumsi dari regresi linier sehingga dapat diatasi dengan menggunakan analisis regresi non linier. Regresi non linier merupakan suatu metode analisis regresi untuk memperoleh model non linier yang digunakan untuk mengetahui hubungan antara variabel respon dan prediktor. Apabila hubungan antara variabel respon dan prediktor bersifat non linier maka kurva yang dihasilkan tidak membentuk garis lurus.

Regresi Poisson merupakan salah satu analisis regresi non linier untuk mengetahui hubungan antara variabel respon yang berdistribusi Poisson dengan satu atau lebih variabel prediktor dan merupakan data diskrit. Dalam regresi Poisson asumsi yang harus dipenuhi yaitu nilai variansi dan nilai rata-rata pada variabel respon bernilai sama (equidispersi). Namun, dalam penerapannya sering ditemukan nilai variansi lebih besar dari nilai rata-rata (overdispersi) atau nilai variansi lebih kecil dari nilai rata-rata (underdispersi). Data yang tidak memenuhi asumsi equidispersi jika dianalisis dengan menggunakan regresi Poisson maka menyebabkan adanya pendugaan galat baku yang rendah sehingga menghasilkan pendugaan parameter yang bias kebawah (underestimate). Hal ini akan memberikan kesimpulan yang tidak sesuai atau tidak valid (Pontoh, 2015).

Pada beberapa kasus penyebab terjadinya overdispersi yaitu data mengalami excess zeros kondisi dimana proporsi nilai nol pada variabel respon lebih besar dari nilai lainnya. Hal ini menyebabkan adanya berbagai pengembangan metode statistik yang dapat mengatasi masalah tersebut. Berbagai pendekatan yang dilakukan untuk

mengatasi masalah data cacah yang menyebabkan overdispersi. Salah satu metode analisis yang dapat digunakan untuk menangani kasus overdispersi karena *excess zeros* yaitu *Hurdle Negative Binomial*. Pada model *hurdle* dapat dilakukan dua jenis pemodelan, pertama memodelkan observasi bernilai nol dengan menggunakan model logit, kedua memodelkan observasi yang bernilai positif dengan menggunakan *truncated binomial*. Kelebihan dari model regresi *Hurdle Negative Binomial* yaitu dapat menampung semua observasi baik observasi yang bernilai nol maupun bulat non-negatif. Pada *Hurdle Negative Binomial* dapat dilakukan penaksiran parameter secara terpisah sehingga lebih mudah dalam penginterpretasiannya (Hilbe, 2011). Adapun model yang dikaji dalam penelitian ini menggunakan model regresi *Hurdle Negative Binomial* (HNB) pada data kasus tetanus neonatorum di Indonesia pada tahun 2019.

Selain pemilihan model statistik yang tepat, seleksi variabel juga merupakan langkah penting dalam penelitian. Dengan banyaknya potensi variabel prediktor, pemilihan variabel yang relevan diperlukan untuk meningkatkan kemampuan prediksi model, mengurangi kompleksitas, serta mempermudah interpretasi hasil analisis. Model yang terlalu banyak memuat variabel yang tidak signifikan dapat menyebabkan ketidakefisienan dan menurunkan ketepatan estimasi parameter. Salah satu metode seleksi variabel yang digunakan adalah *Backward Elimination*. *Backward Elimination* merupakan metode seleksi variabel dengan pendekatan pemilihan mundur, yang diawali dengan memasukkan seluruh variabel prediktor ke dalam model (model penuh). Selanjutnya, dilakukan pengujian signifikansi terhadap masing-masing variabel. Variabel dengan nilai signifikansi terbesar (*p-value* tertinggi) dan melebihi tingkat signifikansi yang ditetapkan akan dieliminasi dari model. Setelah variabel tersebut dihapus, model diestimasi kembali dan proses pengujian diulangi hingga seluruh variabel yang tersisa dalam model signifikan secara statistik. Dengan demikian, metode ini menghasilkan model yang lebih sederhana tanpa mengurangi kemampuan penjelasan model secara signifikan (Bode, 2017).

Penelitian sebelumnya yang dilakukan oleh Desjardins (2013) mengenai perbandingan antara model *Zero Inflated Negative Binomial* (ZINB) dan *Hurdle Negative Binomial* (HNB) dengan menggunakan data simulasi menyatakan bahwa model regresi *Hurdle Negative Binomial* merupakan model regresi terbaik yang dapat menangani kasus overdispersi karena *excess zeros* dibandingkan model regresi *Zero Inflated Negative Binomial* (ZINB). Model ini telah diterapkan oleh beberapa peneliti, Faidah dan Pontoh (2015) menerapkan regresi *Hurdle Negative Binomial* ada kasus penyakit difteri di Provinsi Jawa Barat, Drajat (2021) memodelkan banyaknya penggunaan rokok di Provinsi Sulawesi Tengah menggunakan regresi *Hurdle Negative Binomial* dan Anisa Haura (2022) memodelkan angka kematian bayi di Kota Makassar tahun 2017 menggunakan regresi *Hurdle Negative Binomial*.

Berdasarkan uraian di atas, maka pada penelitian ini akan dilakukan penerapan regresi *Hurdle Negative Binomial* menggunakan *Maximum Likelihood Estimation* dengan seleksi variabel *Backward Elimination* pada data kasus tetanus neonatorum di Indonesia tahun 2019. Sehingga penulis mengangkat judul ***"Penerapan Regresi Hurdle Negative Binomial Dengan Seleksi Variabel Backward Elimination Pada Data Tetanus Neonatorum di Indonesia"***.

1.2 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Data yang digunakan pada penelitian ini adalah data kasus tetanus neonatorum di Indonesia Tahun 2019.
2. Metode estimasi parameter menggunakan *Maximum Likelihood Estimation* (MLE).
3. Menggunakan metode seleksi variabel *Backward Elimination*.

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Membuat model Regresi *Hurdle Negative Binomial* pada kasus tetanus neonatorum di Indonesia Tahun 2019.
2. Mengetahui faktor-faktor yang mempengaruhi kasus tetanus neonatorum di Indonesia Tahun 2019 berdasarkan model Regresi *Hurdle Negative Binomial*.

1.4 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah sebagai berikut:

1. Diharapkan menambah wawasan mengenai model Regresi *Hurdle Negative Binomial* dalam penanganan masalah *excess zeros* dengan metode *Maximum Likelihood Estimation* dan seleksi variabel *Backward Elimination*.
2. Memberikan informasi tambahan kepada pemerintah mengenai faktor-faktor yang berpengaruh signifikan terhadap jumlah kasus tetanus neonatorum.
3. Sebagai bahan pertimbangan dan bahan referensi untuk melakukan penelitian selanjutnya tentang permasalahan tetanus neonatorum di Indonesia.

1.5 Teori

1.5.1 Regresi Poisson

Menurut Cahyandari (2014), regresi Poisson yaitu salah satu analisis regresi yang berasal dari distribusi Poisson yang dapat digunakan untuk memodelkan data peubah respon berupa data diskrit dengan peubah prediktor berupa data diskrit, kontinu, dan campuran. Apabila Y yang merupakan variabel acak Poisson dengan parameter $\mu > 0$, maka fungsi massa peluangnya (Cameron dan Trivedi, 1998):

$$f(y_i, \mu) = \frac{e^{-\mu} \mu^{y_i}}{y_i!}, y_i = 0, 1, 2, \dots \quad (1)$$

dalam *Generalized Linier Model* diketahui bahwa kombinasi linier dari variabel prediktor X dengan koefisien parameter β yang menghasilkan penduga linier η , yaitu:

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} \quad (2)$$

dalam distribusi Poisson diketahui bahwa $\mu_i = E(y_i)$. Model yang menghubungkan antara μ_i dan η_i , dinyatakan dengan $g(\mu_i) = \eta_i$. Apabila menggunakan *ln link*, maka berdasarkan *Generalized Linier Model* untuk distribusi Poisson, hubungan antara μ_i dan η_i dapat dinyatakan dengan $\ln(\mu_i) = \eta_i$. Dengan menggunakan fungsi penghubung log natural tersebut maka didapatkan model regresi Poisson yaitu:

$$\begin{aligned} \ln \mu_i &= \eta_i \\ \ln \mu_i &= \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} \\ \mu_i &= \exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}) \end{aligned} \quad (3)$$

Menurut Cameron dan Trivedi (1998), dalam regresi Poisson asumsi yang harus dipenuhi yaitu nilai varians dan nilai rata-rata (*mean*) dari variabel respon harus sama yang ditunjukkan dalam persamaan berikut:

$$E(Y_i) = V(Y_i) = \mu_i \quad (4)$$

Model regresi Poisson dengan penghubung log natural dapat dituliskan pada Persamaan

$$y_i = \mu_i + \varepsilon_i = \exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}) + \varepsilon_i \quad (5)$$

dengan:

y_i = jumlah kejadian ke- i

μ_i = *mean* jumlah kejadian

ε_i = galat *error*, $i = 1, 2, \dots, n$

Menurut Daniel (1989), pengujian suatu kesesuaian distribusi Poisson dapat digunakan uji *Kolmogrov-Smirnov*. Adapun hipotesis yang digunakan sebagai berikut:

H_0 = data mengikuti distribusi Poisson

H_1 = data tidak mengikuti distribusi Poisson

Statistik uji:

$$D_{hitung} = \max |F_n(Y) - F_0(Y)| \quad (6)$$

dengan:

$F_0(Y)$ = fungsi distribusi kumulatif yang dihipotesiskan

$F_n(Y)$ = fungsi distribusi kumulatif yang diamati

Adapun kriteria pengujiannya adalah tolak H_0 apabila $D_{hitung} > D_{tabel}(\alpha)$.

1.5.2 Overdispersi

Dalam regresi Poisson terdapat asumsi yang harus dipenuhi yaitu asumsi equidispersi atau nilai variansi dan nilai rata-rata pada variabel respon bernilai sama. Namun, dalam penerapannya sering ditemukan nilai variansi lebih besar dari nilai rata-rata yang disebut overdispersi atau nilai variansi lebih kecil dari nilai rata-rata yang disebut underdispersi. Data yang mengandung overdispersi jika dianalisis dengan menggunakan regresi Poisson maka akan menyebabkan adanya pendugaan galat baku yang rendah sehingga menghasilkan pendugaan parameter yang bias kebawah (*underestimate*). Hal ini akan memberikan hasil yang tidak sesuai atau tidak valid (Pontoh, 2015). Salah satu penyebab adanya overdispersi pada data yaitu *excess zeros* pada variabel respon. Menurut Agresti (2002), untuk melihat adanya overdispersi pada data maka dapat dilakukan dengan mendeteksi nilai rasio antara *deviance* dengan derajat bebas. Adapun rasio dispersi dapat dinyatakan dalam persamaan berikut:

$$\phi = \frac{dev}{db} \quad (7)$$

nilai *deviance* dinyatakan dalam Persamaan

$$dev = 2 \sum_{i=1}^n \left\{ y_i \log \frac{y_i}{\hat{y}_i} - (y_i - \hat{y}_i) \right\} \text{ dengan } i = 1, 2, \dots, n \quad (8)$$

dengan:

dev = nilai *deviance*

db = derajat bebas, $db = n - (p + 1)$

y_i = nilai variabel respon ke- i

$\hat{y}_i$ = estimasi regresi Poisson

apabila nilai pada Persamaan (7) lebih besar dari 1, maka dapat dikatakan bahwa dalam model terjadi adanya overdispersi.

1.5.3 Multikolinearitas

Salah satu asumsi yang harus terpenuhi yaitu dengan tidak adanya multikoliniearitas antar variabel prediktor. Uji multikoliniearitas bertujuan untuk melihat interaksi atau hubungan dengan karakteristik linier (korelasi) antara beberapa variabel prediktor dalam sebuah model regresi. Hubungan antara variabel prediktor akan berdampak pada nilai taksir parameter bernilai besar dari errornya. Pendeteksian multikolinearitas dapat dilakukan dengan menghitung nilai *Variance Inflation Factor* (VIF), nilai VIF dinyatakan dalam Persamaan berikut.

$$VIF_j = \frac{1}{1 - R_j^2} \text{ dengan } j = 1, 2, \dots, p \quad (9)$$

dimana R_j^2 merupakan koefisien determinasi dari X_j sebagai variabel respon dan X_j sebagai variabel prediktor. Nilai R_j^2 berkisar antara 0 – 1 sehingga nilai dari VIF akan terus bertambah seiring dengan kenaikan koefisien determinasi. Nilai VIF yang lebih dari 10 dinyatakan adanya multikolinearitas (Hocking, 1996).

1.5.4 Excess Zeros

Permasalahan yang dapat ditemukan dalam regresi Poisson yaitu nilai nol yang berlebih atau dapat disebut *excess zeros*. *Excess zeros* dapat dilihat pada proporsi variabel respon yang bernilai nol lebih besar dari data diskrit lainnya. Menurut Winkleman (2008), banyaknya proporsi pada data yang bernilai nol dari proporsi data lainnya akan berakibat pada ketepatan dalam pengambilan keputusan. Salah satu akibat adanya *excess zeros* yaitu terjadinya overdispersi yang menyebabkan ragam data lebih besar dari rata-rata.

1.5.5 Regresi Hurdle Negative Binomial

Regresi *Hurdle Negative Binomial* (HNB) merupakan regresi yang digunakan untuk data yang variabel responnya berupa data diskrit, dan memiliki proporsi nilai nol yang lebih besar dari nilai lainnya atau disebut dengan *excess zeros*, serta mengalami overdispersi (Desjardins, 2013). Fungsi peluang model regresi *Hurdle Negative Binomial* sebagai berikut (Saffari *et al.*, 2012):

$$P(Y_i = y_i) = \begin{cases} \pi & , y_i = 0 \\ (1 - \pi) \frac{\Gamma(y_i + \frac{1}{k})}{\Gamma(y_i + 1) \Gamma(\frac{1}{k})} \left(\frac{k\mu}{1 + k\mu} \right)^{y_i} \left(\frac{1 + k\mu}{1 - (1 + k\mu)} \right)^{-k-1} & , y_i > 0 \end{cases} \quad (10)$$

Nilai dari variabel respon yang muncul pada pengamatan terbagi dalam dua keadaan yang terpisah. Keadaan pertama disebut *zero hurdle* yang terjadi dengan peluang π , sedangkan keadaan kedua disebut *truncated* binomial negatif yang terjadi dengan peluang $1 - \pi$, dengan $0 < \pi < 1$, μ merupakan rata-rata distribusi binomial negatif dengan $\mu > 0$ dan k merupakan parameter dispersi yang tidak bergantung pada variabel prediktor dengan $k > 0$.

Dapat diketahui bahwa π dan μ merupakan parameter yang bergantung pada vektor dari variabel prediktor, sehingga dapat didefinisikan sebagai berikut:

a. Model Zero Hurdle

Diketahui bahwa π_i dan μ_i bergantung pada vektor dari variabel prediktor yang dapat didefinisikan sebagai berikut:

$$\begin{aligned}
 \text{logit}(\pi_i) &= \log\left(\frac{\pi_i}{1 - \pi_i}\right) = \sum_{j=1}^p x_{ij}\delta_j \\
 \frac{\pi_i}{1 - \pi_i} &= \exp\left(\sum_{j=1}^p x_{ij}\delta_j\right) \\
 \pi_i &= (1 - \pi_i) \exp\left(\sum_{j=1}^p x_{ij}\delta_j\right) \\
 \pi_i &= \exp\left(\sum_{j=1}^p x_{ij}\delta_j\right) - (\pi_i) \exp\left(\sum_{j=1}^p x_{ij}\delta_j\right) \\
 \pi_i \left(1 + \exp\left(\sum_{j=1}^p x_{ij}\delta_j\right)\right) &= \exp\left(\sum_{j=1}^p x_{ij}\delta_j\right)
 \end{aligned}$$

sehingga didapatkan:

$$\pi_i = \frac{\exp\left(\sum_{j=1}^p x_{ij}\delta_j\right)}{1 + \exp\left(\sum_{j=1}^p x_{ij}\delta_j\right)} \quad (11)$$

Adapun model untuk *zero hurdle* sebagai penghubung logit yaitu dapat dinyatakan sebagai berikut:

$$\text{logit}(\pi_i) = \hat{\delta}_0 + \sum_{j=1}^p x_{ij}\hat{\delta}_j \quad (12)$$

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

Jika dituliskan dalam bentuk matriks, maka bentuk Persamaan (12) dapat ditulis:

$$\begin{bmatrix} \text{logit } \pi_1 \\ \text{logit } \pi_2 \\ \vdots \\ \text{logit } \pi_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} \hat{\delta}_0 \\ \hat{\delta}_1 \\ \vdots \\ \hat{\delta}_p \end{bmatrix}$$

b. Model *truncated* binomial negatif

Nilai μ didapat dari model log linier sebagai berikut:

$$\begin{aligned}
 \ln(\mu_i) &= \sum_{j=1}^p x_{ij}\beta_j \\
 \mu_i &= \exp\left(\sum_{j=1}^p x_{ij}\beta_j\right)
 \end{aligned} \quad (13)$$

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

Model untuk *truncated* binomial negatif dengan fungsi penghubung log natural dinyatakan sebagai berikut:

$$\ln(\mu_i) = \hat{\beta}_0 + \sum_{j=1}^p x_{ij} \hat{\beta}_j \quad (14)$$

dan apabila dituliskan dalam bentuk matriks, maka Persamaan (14) dapat ditulis:

$$\begin{bmatrix} \ln \mu_1 \\ \ln \mu_2 \\ \vdots \\ \ln \mu_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \end{bmatrix}$$

Selanjutnya, berdasarkan fungsi probabilitas y_i yang diperoleh pada Persamaan (10) nilai π_i dan μ_i yang disubstitusikan pada Persamaan (11) dan (13), sehingga didapatkan fungsi probabilitas model regresi *Hurdle Negative Binomial* (HNB) pada Persamaan (15) berikut:

$$P(Y_i = y_i) = \begin{cases} \frac{\exp(\sum_{j=1}^p x_{ij} \delta_j)}{1 + \exp(\sum_{j=1}^p x_{ij} \delta_j)} & y_i = 0 \\ \frac{1}{1 + \exp(\sum_{j=1}^p x_{ij} \delta_j)} \frac{\Gamma(y_i + \frac{1}{k})}{\Gamma(y_i + 1) \Gamma(\frac{1}{k})} \left(\frac{k \exp(\sum_{j=1}^p x_{ij} \beta_j)}{1 + k \exp(\sum_{j=1}^p x_{ij} \beta_j)} \right)^{y_i} \frac{(1 + k \exp(\sum_{j=1}^p x_{ij} \beta_j))^{-k-1}}{1 - (1 + k \exp(\sum_{j=1}^p x_{ij} \beta_j))^{-k-1}} & y_i > 0 \end{cases} \quad (15)$$

1.5.6 Estimasi Parameter Regresi *Hurdle Negative Binomial*

Parameter model regresi HNB dapat diestimasi dengan menggunakan metode *Maximum Likelihood Estimation* (MLE). MLE merupakan metode estimasi yang memaksimumkan fungsi *likelihood*. Berdasarkan fungsi probabilitas untuk y_i pada Persamaan (15), maka fungsi *likelihood* pada regresi HNB dibedakan menjadi dua yaitu $y_i = 0$ dan $y_i > 0$. Maka fungsi *likelihood* model regresi HNB dapat dinyatakan dalam Persamaan (16) dan (17) berikut:

untuk $y_i = 0$

$$L_0(\kappa, \delta, \beta) = \prod_{\substack{i=1 \\ y_i=0}}^n \frac{\exp(\sum_{j=1}^p x_{ij} \delta_j)}{1 + \exp(\sum_{j=1}^p x_{ij} \delta_j)} \quad (16)$$

untuk $y_i > 0$

$$L_{\neq 0}(\kappa, \delta, \beta) = \prod_{\substack{i=1 \\ y_i > 0}}^n \frac{1}{1 + \exp(\sum_{j=1}^p x_{ij} \delta_j)} \frac{\Gamma(y_i + \frac{1}{k})}{\Gamma(y_i + 1) \Gamma(\frac{1}{k})} \left(\frac{k \exp(\sum_{j=1}^p x_{ij} \beta_j)}{1 + k \exp(\sum_{j=1}^p x_{ij} \beta_j)} \right)^{y_i} \frac{(1 + k \exp(\sum_{j=1}^p x_{ij} \beta_j))^{-k-1}}{1 - (1 + k \exp(\sum_{j=1}^p x_{ij} \beta_j))^{-k-1}} \quad (17)$$

Selanjutnya membuat fungsi $\ln$ *likelihood* dari persamaan dinyatakan Persamaan (18) dan (19) berikut:

untuk $y_i = 0$

$$\begin{aligned}
 l_0(\kappa, \delta, \beta) &= \ln L_0(\kappa, \delta, \beta) \\
 &= \sum_{y_i=0} \left[\ln \exp \left(\sum_{j=1}^p x_{ij} \delta_j \right) - \ln \left(1 + \exp \left(\sum_{j=1}^p x_{ij} \delta_j \right) \right) \right] \\
 &= \sum_{y_i=0} \left[\left(\sum_{j=1}^p x_{ij} \delta_j \right) - \ln \left(1 + \exp \left(\sum_{j=1}^p x_{ij} \delta_j \right) \right) \right]
 \end{aligned} \tag{18}$$

untuk $y_i > 0$

$$\begin{aligned}
 l_{\neq 0}(\kappa, \delta, \beta) &= \ln L_{\neq 0}(\kappa, \delta, \beta) \\
 &= \sum_{y_i > 0} \left[-\ln(1 + \exp(\sum_{j=1}^p x_{ij} \delta_j)) + \ln \Gamma \left(y_i + \frac{1}{\kappa} \right) - \ln \Gamma(y_i + 1) \right. \\
 &\quad - \ln \Gamma \left(\frac{1}{\kappa} \right) + y_i \ln \left(\exp \left(\sum_{j=1}^p x_{ij} \beta_j \right) \right) + y_i \ln(\kappa) \\
 &\quad - \left(\frac{1}{\kappa} + y_i \right) \ln \left(1 + \kappa \exp \left(\sum_{j=1}^p x_{ij} \beta_j \right) \right) \\
 &\quad \left. - \ln \left(1 - (1 + \kappa \exp(\sum_{j=1}^p x_{ij} \beta_j))^{-\kappa^{-1}} \right) \right]
 \end{aligned} \tag{19}$$

Fungsi $\ln$ *likelihood* model regresi HNB dapat dinyatakan dalam Persamaan (20) berikut yang merupakan gabungan dari dua fungsi $y_i = 0$ dan $y_i > 0$

$$\begin{aligned}
 l(\kappa, \delta, \beta) &= \ln l_0(\kappa, \delta, \beta) + \ln l_{\neq 0}(\kappa, \delta, \beta) \\
 &= \sum_{y_i=0}^n \left[\left(\sum_{j=1}^p x_{ij} \delta_j \right) - \ln \left(1 + \exp(\sum_{j=1}^p x_{ij} \delta_j) \right) \right] + \\
 &\quad \sum_{y_i > 0}^n \left[-\ln(1 + \exp(\sum_{j=1}^p x_{ij} \delta_j)) + \ln \Gamma \left(y_i + \frac{1}{\kappa} \right) - \ln \Gamma(y_i + 1) \right. \\
 &\quad - \ln \Gamma \left(\frac{1}{\kappa} \right) + y_i \ln \left(\exp \left(\sum_{j=1}^p x_{ij} \beta_j \right) \right) + y_i \ln(\kappa) \\
 &\quad - \left(\frac{1}{\kappa} + y_i \right) \ln(1 + \kappa \exp(\sum_{j=1}^p x_{ij} \beta_j)) \\
 &\quad \left. - \ln \left(1 - (1 + \kappa \exp(\sum_{j=1}^p x_{ij} \beta_j))^{-\kappa^{-1}} \right) \right]
 \end{aligned} \tag{20}$$

Estimator parameter model regresi HNB yang didapatkan dari turunan pertama pada Persamaan (20) terhadap κ , δ , dan β yang disamadengankan nol. Jika hasil persamaan yang didapatkan dari turunan pertama fungsi $\ln$ *likelihood* terhadap masing-masing parameter tidak *close form*, maka dapat dilakukan metode iteratif *Newton Raphson*. Rumus umum metode *Newton Raphson* dinyatakan pada Persamaan (21) berikut:

$$\theta^{(m+1)} = \theta^{(m)} - [H(\theta^{(m)})]^{-1} q(\theta^{(m+1)}) \tag{21}$$

dengan $\theta^{(m)} = [\kappa, \delta, \beta]^T$

$q(\theta^{(m)})$ merupakan syarat vektor yang elemennya berisi turunan pertama fungsi *ln likelihood* terhadap masing-masing parameter dan $H(\theta^{(m)})$ merupakan syarat dengan matriks yang setiap elemennya berisi turunan kedua fungsi *ln likelihood* terhadap parameter. Dalam notasi matriks dapat dituliskan sebagai berikut:

$$q(\theta^{(m)}) = \begin{bmatrix} \frac{\partial l(\kappa, \delta, \beta)}{\partial \kappa} & \frac{\partial l(\kappa, \delta, \beta)}{\partial \delta^T} & \frac{\partial l(\kappa, \delta, \beta)}{\partial \beta^T} \end{bmatrix}^T$$

$$H(\theta^{(m)}) = \begin{bmatrix} \frac{\partial^2 l(\kappa, \delta, \beta)}{\partial \kappa^2} & \frac{\partial^2 l(\kappa, \delta, \beta)}{\partial \delta \partial \kappa} & \frac{\partial^2 l(\kappa, \delta, \beta)}{\partial \beta \partial \kappa} \\ \frac{\partial^2 l(\kappa, \delta, \beta)}{\partial \delta \partial \delta^T} & \frac{\partial^2 l(\kappa, \delta, \beta)}{\partial \beta \partial \delta^T} & \frac{\partial^2 l(\kappa, \delta, \beta)}{\partial \beta \partial \beta^T} \\ \text{simetris} & & \end{bmatrix}$$

Iterasi akan berhenti apabila kondisi konvergen terpenuhi, yaitu $\|\theta^{(m+1)} - \theta^{(m)}\| \leq \varepsilon$ dimana ε merupakan bilangan yang sangat kecil.

1.5.7 Backward Elimination

Untuk menentukan variabel yang layak dipertahankan dalam model salah satu metode yang digunakan yaitu *Backward Elimination*. *Backward Elimination* merupakan prosedur pemilihan variabel yang signifikan dimulai dengan memasukkan seluruh variabel prediktor ($X_1, X_2, \dots, X_p$) kedalam model. Setelah itu, variabel yang dianggap paling tidak signifikan akan dikeluarkan satu per satu sehingga variabel yang tersisa dalam model memenuhi kriteria signifikan yang telah ditentukan. Langkah pertama yang dilakukan saat melakukan *Backward Elimination* yaitu dengan menentukan nilai signifikan sebesar 5% pada umumnya yang berarti *p-value* menjadi 0,05. Kedua, memasukkan semua variabel dalam model. Langkah ketiga, cari nilai variabel dengan *p-value* terbesar dan bandingkan dengan nilai signifikan. Langkah keempat, apabila *p-value* terbesar yang ada dalam model > 0.05 maka variabel tersebut akan dihapus. Langkah selanjutnya kembali ke langkah kedua hingga nilai *p-value* tertinggi dalam model < 0.05 (S. Srinidhi, 2019).

1.5.8 Pengujian Hipotesis

Setelah model terbentuk, dilakukan pemeriksaan peranan variabel-variabel prediktor dalam model, hal ini perlu dilakukan untuk pengujian terhadap parameter model (δ_j dan β_j , dengan $j = 1, 2, \dots, p$). Pengujian terhadap parameter dilakukan baik itu simultan maupun parsial. Pengujian parameter secara simultan dilakukan dengan menggunakan uji *Likelihood Ratio* dan pengujian parameter secara parsial dilakukan dengan menggunakan statistik uji *t*.

a. Pengujian Simultan

Statistik uji G merupakan uji *likelihood ratio* yang digunakan untuk mengetahui apakah variabel prediktor secara bersama-sama mempengaruhi variabel respon secara signifikan. Hipotesis uji simultan sebagai berikut (Agresti, 2002):

$H_0 : \delta_1 = \dots = \delta_p = \beta_1 = \dots = \beta_p = 0$ (variabel prediktor tidak memiliki pengaruh terhadap variabel respon)

$H_1 : \text{paling sedikit ada satu } \delta_j \neq 0 \text{ atau } \beta_j \neq 0 \text{ dimana } j = 1, 2, \dots, p$

Statistik uji G dinyatakan pada Persamaan (22) berikut:

$$G = -2 \ln \left[\frac{L(\kappa, \delta_0, \beta_0)}{L(\kappa, \delta, \beta)} \right] = -2(l(\kappa, \delta_0, \beta_0) - l(\kappa, \delta, \beta)) \quad (22)$$

dengan $L(\kappa, \delta_0, \beta_0)$ merupakan fungsi *ln likelihood* tanpa variabel prediktor (model konstan) dan $L(\kappa, \delta, \beta)$ merupakan fungsi *ln likelihood* dengan variabel prediktor (model penuh). Statistik uji G mengikuti sebaran distribusi *Chi Square* dengan derajat bebas ν yaitu banyaknya parameter model dibawah populasi dikurangi banyaknya parameter dibawah H_0 . Daerah penolakan H_0 yaitu jika $G > \chi^2_{(\alpha, \nu)}$ atau $p - \text{value} < \alpha$ yang dapat dikatakan bahwa minimal terdapat satu variabel prediktor yang mempengaruhi variabel respon.

b. Pengujian Parsial

Pengujian parameter regresi secara parsial dilakukan dengan menggunakan uji statistik *t*. Statistik uji *t* merupakan statistik uji yang sering digunakan untuk menguji signifikansi parameter secara parsial pada setiap variabel prediktor. Pengujian parameter secara parsial pada masing-masing variabel bagian *zero hurdle* dan *truncated negative binomial* sebagai berikut (Hosmer dan Lemeshow, 2000):

Hipotesis model *zero hurdle*

$H_0 : \delta_j = 0$

$H_1 : \delta_j \neq 0 \text{ dengan } j = 1, 2, \dots, p$

Statistik uji *t* model *zero hurdle* sebagai berikut:

$$t_j = \frac{\hat{\delta}_j}{SE(\hat{\delta}_j)} \quad (23)$$

dimana $\hat{\delta}_j$ merupakan estimator dari δ_j dan $SE(\hat{\delta}_j)$ merupakan *standard error* dari estimasi parameter $\hat{\delta}_j$ yang disebut matriks varian kovarian dari $\hat{\delta}_j$ yang didapatkan melalui invers negatif matriks Hessian.

Kriteria pengujian H_0 ditolak apabila $|t| > t_{(\frac{\alpha}{2}, n-1)}$ atau $p - \text{value} < \alpha$ dimana $t_{\alpha/2}$ merupakan nilai *t* pada taraf α yang ditentukan dan $n - 1$ merupakan derajat bebas.

Hipotesis model *truncated negative binomial*

$H_0 : \beta_j = 0$

$H_1 : \beta_j \neq 0 \text{ dengan } j = 1, 2, \dots, p$

Statistik uji *t* model *truncated negative binomial* sebagai berikut:

$$t_j = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \quad (24)$$

dimana $\hat{\beta}_j$ merupakan estimator dari β_j dan $SE(\hat{\beta}_j)$ merupakan *standard error* dari estimasi parameter $\hat{\beta}_j$ yang disebut matriks varian kovarian dari $\hat{\beta}_j$ yang didapatkan dari melalui invers negatif dari matriks Hessian.

Kriteria pengujian H_0 ditolak apabila $|t| > t_{(\frac{\alpha}{2}, n-1)}$ atau $p\text{-value} < \alpha$ dimana $t_{\alpha/2}$ merupakan nilai t pada taraf α yang ditentukan dan $n - 1$ merupakan derajat bebas.

1.5.9 Tetanus Neonatorum

Tetanus neonatorum merupakan penyakit tetanus yang terjadi pada bayi yang baru lahir. Penyakit ini disebabkan oleh *Clostridium Tetani* pada bayi usia kurang dari 28 hari yang dapat menyebabkan kematian. Infeksi ini terjadi akibat saat pemotongan tali pusat menggunakan alat yang tidak steril. Upaya yang dapat dilakukan untuk mencegah kembali adanya kasus tetanus neonatorum yaitu dengan melakukan imunisasi Tetanus Toksoid (TT) yang lengkap pada ibu hamil, perawatan persalinan dan pasca persalinan yang bersih (Sari S. N, 2017).

Menurut Rusdiana (2017), terdapat beberapa faktor yang dapat dilakukan untuk mengurangi risiko terjadinya tetanus neonatorum yaitu sebagai berikut:

a. Imunisasi Tetanus Toksoid (TT)

1) Wanita Usia Subur (WUS)

Imunisasi Tetanus Toksoid (TT) yang diberikan kepada Wanita Usia Subur (WUS) yaitu wanita yang berusia 15-39 tahun. Imunisasi ini dilakukan sebanyak 5 kali dengan jarak waktu tertentu. Untuk presentase cakupan imunisasi TT5 pada wanita usia subur dapat diperoleh dengan rumus berikut:

$$\frac{\text{Jumlah WUS yang melakukan imunisasi TT5}}{\text{seluruh jumlah WUS}} \times 100\%$$

2) Ibu Hamil

Kelompok ibu hamil yang sudah melakukan TT2 hingga TT5 dapat dikatakan mendapatkan imunisasi TT2+. Adapun manfaat dari adanya pemberian imunisasi pada ibu hamil yaitu dapat mencegah tetanus pada ibu dan juga bayinya yaitu pada saat proses persalinan terdapat luka pada rahim ataupun pada tali pusat bayi.

a) Cakupan imunisasi TT1

Untuk mengukur presentase cakupan imunisasi TT1 dapat dilakukan dengan rumus berikut:

$$\frac{\text{Jumlah ibu hamil yang melakukan imunisasi TT1}}{\text{seluruh jumlah ibu hamil}} \times 100\%$$

b) Cakupan imunisasi TT2

Untuk mengukur presentase cakupan imunisasi TT2 dapat dilakukan dengan rumus berikut:

$$\frac{\text{Jumlah ibu hamil yang melakukan imunisasi TT2}}{\text{seluruh jumlah ibu hamil}} \times 100\%$$

c) Cakupan imunisasi TT3

Untuk mengukur presentase cakupan imunisasi TT3 dapat dilakukan dengan rumus berikut:

$$\frac{\text{Jumlah ibu hamil yang melakukan imunisasi TT3}}{\text{seluruh jumlah ibu hamil}} \times 100\%$$

d) Cakupan imunisasi TT4

Untuk mengukur presentase cakupan imunisasi TT4 dapat dilakukan dengan rumus berikut:

$$\frac{\text{Jumlah ibu hamil yang melakukan imunisasi TT4}}{\text{seluruh jumlah ibu hamil}} \times 100\%$$

e) Cakupan imunisasi TT5

Untuk mengukur presentase cakupan imunisasi TT5 dapat dilakukan dengan rumus berikut:

$$\frac{\text{Jumlah ibu hamil yang melakukan imunisasi TT5}}{\text{seluruh jumlah ibu hamil}} \times 100\%$$

f) Cakupan imunisasi TT2+

Untuk mengukur presentase cakupan imunisasi TT2+ dapat dilakukan dengan rumus berikut:

$$\frac{\text{Jumlah ibu hamil yang melakukan imunisasi TT2 hingga TT5}}{\text{seluruh jumlah ibu hamil}} \times 100\%$$

b. Kunjungan antenatal empat kali (K4)

Tujuan dari K4 yaitu untuk menurunkan kesakitan dan kematian ibu dan perinatal, dimana kasus tetanus neonatorum terjadi pada periode perinatal. Pelayanan antenatal sesuai dengan standar paling sedikit empat kali, yaitu minimal satu kali kontak pada trimester pertama (K1), minimal satu kali kontak pada trimester kedua (K2) dan minimal dua kali kontak pada trimester ketiga (K3 dan K4). Untuk mengukur presentase dari cakupan kunjungan antenatal empat kali (K4) dapat dilakukan dengan rumus berikut:

$$\frac{\text{Jumlah ibu hamil yang melakukan kunjungan antenatal 4 kali (K4)}}{\text{seluruh jumlah ibu hamil}} \times 100\%$$

c. Cakupan persalinan di fasilitas pelayanan kesehatan

Adapun persalinan di fasilitas kesehatan yaitu dengan mengirim ibu bersalin ke fasilitas kesehatan seperti puskesmas dan rumah sakit yang memiliki perlengkapan yang memadai untuk proses persalinan. Perawatan persalinan dan pasca persalinan yang bersih dan steril dapat menurunkan infeksi perinatal termasuk tetanus neonatorum. Untuk mengukur presentase dari cakupan persalinan di fasilitas pelayanan kesehatan dapat dilakukan dengan rumus berikut:

$$\frac{\text{Jumlah ibu hamil yang melakukan persalinan di fasyankes}}{\text{seluruh jumlah ibu hamil}} \times 100\%$$

BAB II

METODE PENELITIAN

2.1 Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yaitu mengenai faktor-faktor yang diperkirakan mempengaruhi kasus tetanus neonatorum di Indonesia pada tahun 2019. Data ini diperoleh dari publikasi Profil Kementerian Kesehatan Indonesia pada situs <http://pusdatin.kemendes.go.id> dengan jumlah pengamatan sebanyak 34 provinsi yang ada di Indonesia.

2.2 Variabel Penelitian

Variabel dalam penelitian ini terdiri dari variabel respon (Y) dan variabel prediktor (X). Adapun variabel yang digunakan dalam penelitian ini adalah:

Tabel 1. Variabel Respon dan Prediktor

Variabel	Keterangan	Skala
Y	Jumlah Kasus Tetanus Neonatorum di Indonesia Tahun 2019	Rasio
X_1	Presentase cakupan imunisasi TT1 pada ibu hamil	Rasio
X_2	Presentase cakupan imunisasi TT2+ pada ibu hamil	Rasio
X_3	Presentase cakupan imunisasi TT5 pada wanita subur	Rasio
X_4	Presentase cakupan kunjungan antenatal empat kali	Rasio
X_5	Presentase cakupan persalinan di fasilitas pelayanan Kesehatan	Rasio

2.3 Metode Analisis

Adapun langkah-langkah analisis yang dilakukan berdasarkan tujuan penelitian dengan bantuan *software R-Studio*.

1. Melakukan analisis statistik dekriptif pada variabel penelitian.
2. Mengidentifikasi hubungan linier antar variabel respon dan prediktor.
3. Melakukan pengujian kecocokan distribusi Poisson dengan menggunakan uji *Kolmogorov-Smirnov*.

- a. Hipotesis

H_0 : data mengikuti distribusi Poisson

H_1 : data tidak mengikuti distribusi Poisson

- b. Statistik Uji

$$D_{hitung} = \max |F_n(Y) - F_0(Y)|$$

dengan:

$F_0(Y)$ = fungsi distribusi kumulatif yang dihipotesiskan

$F_n(Y)$ = fungsi distribusi kumulatif yang diamati

- c. Kriteria Pengujian

H_0 ditolak apabila $D_{hitung} > D_{tabel}$

4. Melakukan pengujian dispersi dengan uji *deviance*.

a. Hipotesis

H_0 : dispersi = 1 (variabel respon tidak mengalami overdispersi)

H_1 : dispersi > 1 (variabel respon mengalami overdispersi)

b. Statistik Uji

$$\phi = \frac{dev}{db}$$

c. Kriteria Pengujian

H_0 ditolak jika dispersi > 1

5. Memeriksa kelebihan nilai nol (*excess zeros*) pada variabel respon.
6. Mendapatkan model regresi *Hurdle Negative Binomial* (HNB) dengan menggunakan *Maximum Likelihood Estimation* pada data jumlah kasus tetanus neonatorum di Indonesia tahun 2019.
7. Menyeleksi variabel yang didapatkan dengan menggunakan fitur seleksi *Backward Elimination*.
8. Melakukan pengujian signifikansi parameter model regresi *Hurdle Negative Binomial* (HNB) secara simultan dengan uji *likelihood ratio*.

$$G = -2 \ln \left[\frac{L(\kappa, \delta_0, \beta_0)}{L(\kappa, \delta, \beta)} \right] = -2(l(\kappa, \delta_0, \beta_0) - l(\kappa, \delta, \beta))$$

dan secara parsial dengan uji *t*

$$t(\hat{\delta}_j) = \frac{\hat{\delta}_j}{SE(\hat{\delta}_j)} ; t(\hat{\beta}_j) = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}$$

9. Menginterpretasi model yang didapatkan dan menarik kesimpulan.