BAB II

LANDASAN TEORI

2.1 Kriminalitas

Kriminalitas merupakan tindakan ilegal yang dilakukan oleh individu atau kelompok yang bertentangan dengan hukum pidana (Kitab Undang-Undang Hukum Pidana/KUHP) dan peraturan perundang-undangan lainnya di Indonesia (Azhar, 2011). Dalam konteks kebijakan publik, kriminalitas sering diukur sebagai jumlah kasus kejahatan per 100.000 penduduk dalam satu periode tertentu (Hermanto et al., 2020). Tingkat kriminalitas tidak hanya mencerminkan intensitas tindakan kriminal tetapi juga indikator ketidakadilan sosial, ketimpangan ekonomi, dan lemahnya sistem kontrol sosial serta penegakan hukum (Prawidyah & Syahputri, 2016).

Kriminalitas dipengaruhi oleh dua dimensi utama, yaitu faktor internal dan faktor eksternal. Faktor internal meliputi kondisi psikologis pelaku, motivasi ekonomi, dan latar belakang keluarga. Sementara itu, faktor eksternal lebih terkait dengan kondisi lingkungan sosial, ekonomi, budaya, dan struktur pemerintahan (Santoso & Zulfa, 2022).

Beberapa faktor utama yang sering dikaitkan dengan meningkatnya tingkat kriminalitas adalah :

a. Kemiskinan

Kemiskinan menjadi salah satu penyebab utama perilaku kriminal karena mengurangi kemampuan seseorang untuk memenuhi kebutuhan dasar secara legal. Hal ini mendorong individu melakukan tindakan ilegal seperti pencurian, pemerasan, dan kejahatan jalanan lainnya (Rahmalia et al., 2019; Yusuf & Zanudin, 2025). Menurut Chantiqa et al. (2024) menunjukkan bahwa kurangnya akses terhadap lapangan kerja dan rendahnya kualitas pendidikan semakin memperparah dampak kemiskinan terhadap kriminalitas. Menurut BPS, Indeks Keparahan Kemiskinan (*Severity of Poverty Index*) adalah ukuran statistik yang digunakan untuk menggambarkan tingkat atau kedalaman kemiskinan di suatu wilayah atau populasi.

Rumus yang digunakan adalah :

$$P_1 = \frac{1}{q} \sum_{i=1}^q \left(\frac{z - y_i}{z} \right) \quad (2.1.1)$$

dimana :

- P_1 : Indeks Kearifan Kemiskinan (*Poverty Severity Index*)
- z : Garis kemiskinan (misalnya: pengeluaran per kapita per bulan yang layak)
- y_i : Pendapatan/pengeluaran individu miskin ke-i
- q : Jumlah penduduk miskin

b. Tingkat Pengangguran Terbuka

Menurut BPS, pengangguran terbuka adalah kondisi di mana seseorang tidak memiliki pekerjaan sama sekali, sedang mencari pekerjaan, dan secara aktif berusaha memperoleh pekerjaan yang layak. Pengangguran jenis ini termasuk dalam kategori pengangguran riil (nyata) karena memang tidak ada hubungan kerja atau aktivitas produktif yang dilakukan oleh individu tersebut. Pengangguran merupakan masalah struktural yang berpotensi memicu tindakan kriminal. Individu yang tidak memiliki pekerjaan resmi cenderung mencari penghasilan alternatif melalui cara ilegal (Dona & Setiawan, 2015). Menurut Hidayat (2020), daerah dengan tingkat pengangguran tinggi cenderung memiliki tingkat kriminalitas yang lebih tinggi dibandingkan daerah dengan tingkat pengangguran rendah.

Rumus yang digunakan adalah :

$$TPT = \left(\frac{\text{Jumlah pengangguran terbuka}}{\text{jumlah angkatan kerja}} \right) \times 100\% \quad (2.1.2)$$

c. Ketimpangan Ekonomi

Ketimpangan pendapatan antarkelompok masyarakat merupakan faktor kuat yang memicu konflik sosial dan tindakan kriminal. Teori anomie yang dikembangkan oleh Merton (1938) menjelaskan bahwa ketika ada kesenjangan antara harapan hidup dan realita ekonomi, individu akan mencari cara alternatif untuk mencapai tujuan tersebut, termasuk melalui tindakan kriminal (Santoso & Zulfa, 2022). Penelitian di

Sumatera Utara membuktikan bahwa variabel Gini Ratio memiliki hubungan positif signifikan dengan tingkat kriminalitas, artinya semakin tinggi ketimpangan ekonomi, semakin tinggi pula angka kejahatan (Prawidyah & Syahputri, 2016). Pengeluaran per kapita menggambarkan berapa besar rata-rata uang yang dikeluarkan oleh satu orang untuk memenuhi kebutuhan hidupnya, seperti makanan, pendidikan, kesehatan, transportasi, dan lain-lain. Dengan rumus :

$$\text{Pengeluaran Per Kapita} = \frac{\text{total pengeluaran seluruh penduduk}}{\text{jumlah penduduk}} \quad (2.1.3)$$

d. Tingkat Pendidikan

Pendidikan berperan penting dalam membentuk karakter dan moral individu. Rendahnya tingkat pendidikan berdampak pada minimnya kesadaran hukum dan norma sosial, sehingga meningkatkan risiko terlibat dalam aktivitas kriminal (Rahmalia et al., 2019). Kuciswara et al. (2021) menyatakan bahwa pendidikan yang berkualitas dapat menjadi solusi jangka panjang dalam menekan angka kriminalitas melalui peningkatan kesadaran dan partisipasi masyarakat dalam pembangunan. Rata-rata Lama Sekolah (RLS) adalah indikator yang menunjukkan jumlah rata-rata tahun pendidikan yang telah ditempuh oleh penduduk berusia 15 tahun ke atas, baik yang masih sekolah maupun yang sudah tidak sekolah. Dihitung dengan cara :

$$RLS = \frac{\sum(P \times t)}{q} \quad (2.1.4)$$

dimana :

- RLS* : rata-rata lama sekolah
- P* : jumlah penduduk pada tingkat pendidikan tertentu
- t* : tahun yang ditempuh
- q* : jumlah total penduduk usia 15 tahun ke atas

e. Kepadatan Penduduk

Daerah perkotaan dengan kepadatan penduduk tinggi sering menjadi pusat kriminalitas karena adanya heterogenitas sosial, kompetisi sumber daya, dan potensi anomie (kehilangan nilai normatif) akibat urbanisasi cepat (Adi, 2022). *Federal*

Bureau of Investigation (FBI) menyebutkan bahwa variasi komposisi penduduk, mobilitas tinggi, dan model transportasi menjadi faktor pendukung kenaikan tingkat kriminalitas di wilayah perkotaan. Rumus kepadatan penduduk adalah :

$$\text{Kepadatan penduduk} = \left(\frac{\text{Jumlah penduduk}}{\text{luas wilayah (km}^2\text{)}} \right) \quad (2.1.5)$$

f. Stabilitas Keluarga dan Perceraian

Stabilitas keluarga merupakan faktor penting dalam pembentukan kepribadian anak. Tingkat perceraian yang tinggi berkontribusi pada melemahnya pengawasan orang tua, sehingga meningkatkan risiko anak terlibat dalam tindakan kriminal.

g. Pembangunan Manusia

Indeks Pembangunan Manusia (IPM) yang rendah menggambarkan kualitas hidup masyarakat yang buruk, termasuk dalam aspek kesehatan, pendidikan, dan standar hidup. IPM yang rendah berkorelasi positif dengan tingkat kriminalitas (Triani et al). Rumus perhitungan IPM adalah :

$$IPM = \sqrt[3]{I_{HLS} \times I_{pendidikan} \times I_{PPK}} \quad (2.1.6)$$

- I_{HLS} : Harapan hidup saat lahir
- $I_{pendidikan}$: Indeks pendidikan
- I_{PPK} : Indeks pengeluaran

2.2 Regresi Poisson

Regresi Poisson merupakan salah satu bentuk regresi nonlinear dari distribusi Poisson yang di gunakan untuk model data cacah (Hilbe, 2011). Model regresi Poisson termasuk kedalam *Generalized Linear Models* karena variabel respon Y memiliki sebaran dari keluarga eksponensial yaitu sebaran poisson, tidak terdapat multikolinearitas antar variabel predictor dan mengasumsikan ekuidispersi.

Analisis regresi Poisson mengasumsikan bahwa distribusi yang mendasari variabel respon Y yang digunakan adalah poisson. Maka fungsi peluang distribusi poisson dengan parameter μ dinyatakan sebagai berikut (Kleinbaum dan Kupper, 1998) :

$$f(y; \mu) = \frac{e^{-\mu} \mu^y}{y!} \quad (2.2.1)$$

dimana :

- μ : rata-rata banyaknya sukses dalam selang waktu atau daerah tertentu
($\mu > 0$)
- y : banyaknya sukses dalam selang waktu atau daerah tertentu
($y = 0, 1, 2, \dots$)

Misalkan data diambil dari bentuk :

$$\begin{bmatrix} y_1 & x_{11} & x_{21} & \dots & x_{k1} \\ y_2 & x_{12} & x_{22} & \dots & x_{k2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ y_n & x_{1n} & x_{2n} & \dots & x_{kn} \end{bmatrix}$$

dengan y_i adalah observasi ke- i dari variabel respon Y , dan x_{ji} adalah nilai variabel prediktor X_j ($j = 1, 2, \dots, k$), maka model regresi poisson dapat dinyatakan sebagai berikut:

$$E(y_i) = \mu(x_i, \beta) = \exp(x_i^T \beta); \quad i = 1, 2, \dots, n \quad (2.2.2)$$

dengan $x_i^T = (x_{1i}, x_{2i}, \dots, x_{ki})$ atau matriks transformasi dari variabel bebasnya dan $\beta^T = (\beta_1, \beta_2, \dots, \beta_k)$ atau matriks transformasi dari parameter regresi. Fungsi $\mu(x_i, \beta)$ merupakan model regresi Poisson yang merupakan fungsi dari x_i sebagai variabel predictor dan β sebagai parameter regresi yang akan ditaksir.

Bentuk umum model regresi Poisson dengan fungsi *Link log*, dimana jika $Y_i \sim \text{Poisson}(\mu_i)$, maka model dapat dituliskan dalam bentuk :

$$\ln(\mu(X_1, X_2, \dots, X_j)) = \beta_0 + \beta_1 x_2 + \dots + \beta_j X_j \quad (2.2.3)$$

atau

$$\begin{aligned} \mu(X_1, X_2, \dots, X_j) &= \exp(\beta_0 + \beta_1 x_2 + \dots + \beta_j X_j) \\ &= \exp(\beta_0 + \sum_{j=1}^p \beta_j X_j) \end{aligned} \quad (2.2.4)$$

Taksiran parameter koefisien regresi Poisson dapat dilakukan dengan menggunakan *Maximum Likelihood Estimation* (MLE). Fungsi *log likelihood* untuk regresi Poisson dapat ditulis sebagai :

$$\ell(y_i; \mu_i) = \sum_{i=1}^n [y_i \ln(\mu_i) - \mu_i - \ln(y_i!)] \quad (2.2.5)$$

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

$$\ell(y_i; \boldsymbol{\beta}) = \sum_{i=1}^n [y_i (\mathbf{x}_i^T \boldsymbol{\beta}) - \exp(\mathbf{x}_i^T \boldsymbol{\beta}) - \ln(y_i!)] \quad (2.2.6)$$

Dengan demikian, penaksiran *maximum likelihood* dapat diselesaikan dengan memaksimumkan model *log likelihood* $\ell(y_i; \boldsymbol{\beta})$, yaitu :

$$\frac{\partial \ell(y_i; \boldsymbol{\beta})}{\partial \beta_j} = \sum_{i=1}^n (y_i - \mu_i) x_{ij} = 0, \quad j = 1, 2, \dots, p \quad (2.2.7)$$

dan

$$\frac{\partial^2 \ell(y_i; \boldsymbol{\beta})}{\partial \beta_j \partial \beta_s} = \sum_{i=1}^n \mu_i x_{ij} x_{is}, \quad j, s = 1, 2, \dots, p \quad (2.2.8)$$

2.3 Overdispersi

Salah satu asumsi yang harus dipenuhi dalam model regresi Poisson adalah equidispersi yaitu, kondisi dimana kesamaan antara rata-rata dan variansi dari variabel dependen. Namun, dalam analisis data cacah seringkali dijumpai data yang variansinya lebih besar dari rata-ratanya atau disebut overdispersi. Fenomena overdispersi dapat dituliskan:

$$\text{Var}(Y) > E(Y)$$

Overdispersi dapat diindikasikan dengan nilai *deviance* dan *pearson chi-squares* yang dibagi dengan derajat kebebasannya. Jika kedua nilai tersebut lebih dari 1, maka dikatakan terjadinya overdispersi pada data (McCullagh, 1989).

Terdapat dua cara yang dapat digunakan untuk mendeteksi overdispersi, yaitu :

1. Deviance

$$\phi_1 = \frac{D^2}{db}, \quad (2.3.1)$$

dimana :

$$D^2 = 2 \sum_{i=1}^n \left\{ y_i \ln \left(\frac{y_i}{\hat{\mu}_i} \right) - (y_i - \mu_i) \right\}$$

$$db = n - k$$

dengan k merupakan banyaknya parameter termasuk konstanta, n merupakan banyaknya pengamatan dan D^2 adalah nilai *Deviance*.

2. *Pearson Chi-Squares*

$$\phi_2 = \frac{x^2}{db}, \quad (2.3.2)$$

dimana :

$$x^2 = \sum_{i=1}^n \frac{(y_i - \mu_i)^2}{var(y_i)}$$

$$db = n - k$$

dengan k merupakan banyaknya parameter termasuk konstanta, n merupakan banyaknya pengamatan dan x^2 adalah nilai *Pearson Chi-Squares*.

2.4 Multikolinearitas

Regresi Poisson juga memiliki asumsi bahwa tidak ada multikolinieritas antar variabel prediktor. Multikolinieritas merupakan adanya korelasi yang tinggi diantara variabel-variabel bebas dalam model. Salah satu metode untuk mendeteksi multikolinieritas adalah dengan nilai *Variance Inflation Factors* (VIF) (Kutner et all, 2004). Kasus Multikolinearitas dapat diketahui melalui *Variance Inflation Factors* (VIF) yang bernilai lebih besar dari 10, dengan nilai VIF yang dinyatakan sebagai berikut.

$$VIF_j = \frac{1}{1 - R_j^2} \quad (2.4.1)$$

Berdasarkan rumus tersebut, dapat diketahui bahwa R_j^2 adalah koefisien determinasi antara x_j dengan variabel prediktor lainnya. Nilai R_j^2 akan sama dengan nol dan VIF akan bernilai satu apabila variabel prediktor tidak saling linier pada model regresi. Nilai VIF lebih lebih dari 10 mengindikasikan adanya multikolinieritas diantara variabel-variabel prediktor.

2.5 Regresi Binomial Negatif

Regresi binomial negatif merupakan suatu model regresi yang digunakan untuk menganalisis hubungan antara suatu variabel respon yang berupa data cacah dengan satu atau lebih variabel bebas. Regresi binomial negatif dapat digunakan baik dalam keadaan equidispersi ataupun overdispersi (Fadil, Raupong, & Ilyas, 2024).

Jika y_i menyebar Poisson dengan parameter $\lambda_i u_i$ dan terdapat variabel *random* u_i berdistribusi gamma dengan parameter v dan rata-rata 1. Maka fungsi peluang bersama Poisson Gamma adalah :

$$\begin{aligned} f(y_i; u_i) &= \int_0^{\infty} \frac{e^{-\lambda_i u_i} (\lambda_i u_i)^{y_i}}{y_i!} \frac{v^v}{\Gamma(v)} u_i^{v-1} e^{-v u_i} du_i \\ &= \frac{\lambda_i^{y_i}}{\Gamma(y_i + 1)} \frac{v^v}{\Gamma(v)} \int_0^{\infty} e^{-(\lambda_i + v) u_i} u_i^{(y_i + v - 1)} du_i \end{aligned}$$

Misal $t = (\lambda_i + v) u_i$

maka diperoleh

$$\begin{aligned} u_i &= \frac{t}{(\lambda_i + v)} \\ du_i &= \frac{dt}{(\lambda_i + v)} \end{aligned}$$

sehingga diperoleh,

$$f(y_i; u_i) = \frac{\Gamma(y_i + v)}{\Gamma(y_i + 1)\Gamma(v)} \left(\frac{\lambda_i}{\lambda_i + v}\right)^{y_i} \left(\frac{v}{\lambda_i + v}\right)^v$$

Jika *inverse* dari v adalah α dan $\lambda_i = \mu_i$. Sehingga fungsi kepadatan peluang Binomial Negatif menjadi :

$$f(y_i; \mu_i; \alpha) = \frac{\Gamma(y_i + \frac{1}{\alpha})}{\Gamma(y_i + 1)\Gamma(\frac{1}{\alpha})} \left(\frac{\alpha \mu_i}{1 + \alpha \mu_i}\right)^{y_i} \left(\frac{v}{1 + \alpha \mu_i}\right)^{\frac{1}{\alpha}} \quad (2.5.1)$$

Rataan dan variansi dari Binomial Negatif adalah $E(y_i) = \mu_i$ dan $Var(y_i) = (1 + \alpha \mu_i)$, dengan α merupakan parameter dispersi. Jika $\alpha = 0$, maka rata-rata dan variansinya akan sama,

$$E(y_i) = Var(y_i).$$

Jika $\alpha > 0$, maka variansinya akan melebihi rata-ratanya,

$$\text{Var}(y_i) > E(y_i).$$

Jika diasumsikan bahwa $E(y_i|\mathbf{x}_i) = \mu_i = \exp(\mathbf{x}_i^T \boldsymbol{\beta})$. Dengan $\mathbf{x}_i$ adalah vektor yang berukuran $p \times 1$ yang menjelaskan variabel *independent* dan $\boldsymbol{\beta}$ adalah vektor berukuran $p \times 1$ merupakan parameter regresi. Maka model *log likelihood* untuk regresi Binomial Negatif dapat dituliskan sebagai:

$$\begin{aligned} \ell(y_i; \mu_i; \alpha) = \sum_{i=1}^n y_i \ln \left(\frac{\alpha \mu_i}{1 + \alpha \mu_i} \right) - \frac{1}{\alpha} \ln(1 + \alpha \mu_i) + \ln \Gamma \left(y_i + \frac{1}{\alpha} \right) - \ln \Gamma(y_i + 1) \\ - \ln \Gamma \left(\frac{1}{\alpha} \right) \end{aligned} \quad (2.5.2)$$

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

$$\begin{aligned} \ell(y_i; \mu_i; \alpha) = \sum_{i=1}^n y_i \ln \left(\frac{\alpha \exp(\mathbf{x}_i^T \boldsymbol{\beta})}{1 + \alpha \exp(\mathbf{x}_i^T \boldsymbol{\beta})} \right) - \frac{1}{\alpha} \ln(1 + \alpha \exp(\mathbf{x}_i^T \boldsymbol{\beta})) \\ + \ln \Gamma \left(y_i + \frac{1}{\alpha} \right) - \ln \Gamma(y_i + 1) - \ln \Gamma \left(\frac{1}{\alpha} \right) \end{aligned} \quad (2.5.3)$$

Secara umum bentuk model regresi Binomial Negatif sebagai berikut:

$$\ln(\mu(X_1, X_2, \dots, X_j)) = (\beta_0 + \beta_1 x_2 + \dots + \beta_j X_j) \quad (2.5.4)$$

atau

$$\begin{aligned} \mu(X_1, X_2, \dots, X_j) &= \exp(\beta_0 + \beta_1 x_2 + \dots + \beta_j X_j) \\ &= \exp\left(\beta_0 + \sum_{j=1}^p \beta_j X_j\right) \end{aligned} \quad (2.5.5)$$

Dengan demikian, penaksir *maximum likelihood* dapat diselesaikan dengan memaksimumkan model $\ell(y_i; \boldsymbol{\beta}; \alpha)$, yaitu :

$$\frac{\partial \ell(y_i; \boldsymbol{\beta}, \alpha)}{\partial \beta_j \beta_s} = \sum_{i=1}^n \frac{(y_i - \mu_i) x_{ij}}{1 + \alpha \mu_i} = 0, \quad j = 1, 2, \dots, p \quad (2.5.6)$$

$$\frac{\partial^2 \ell(y_i; \boldsymbol{\beta}, \alpha)}{\partial \beta_j \beta_s} = \sum_{i=1}^n \frac{\mu_i x_{ij} x_{is} (1 + \alpha y_i)}{(1 + \alpha \mu_i)^2} = 0, \quad j = 1, 2, \dots, p \quad (2.5.7)$$

dan

$$\frac{\partial \ell(y_i; \boldsymbol{\beta}, \alpha)}{\alpha} = \sum_{i=1}^n \left[\frac{1}{\alpha^2} \left(\ln(1 + \alpha \mu_i) + \frac{\alpha(y_i - \mu_i)}{1 + \alpha \mu_i} \right) + \psi \left(y_i + \frac{1}{\alpha} \right) - \psi \left(\frac{1}{\alpha} \right) \right] = 0 \quad (2.5.8)$$

$$\frac{\partial^2 \ell(y_i; \boldsymbol{\beta}, \alpha)}{\partial \alpha^2} = \sum_{i=1}^n \left[-\frac{1}{\alpha^3} \left(\frac{\alpha(1+2\alpha)(y_i - \mu_i) - \alpha\mu_i(1+\alpha\mu_i)}{(1+\alpha\mu_i)^2} \right) + \psi' \left(y_i + \frac{1}{\alpha} \right) - \psi' \left(\frac{1}{\alpha} \right) \right] \quad (2.5.9)$$

Fungsi digamma ψ merupakan turunan pertama dari fungsi log-gamma $\ln\Gamma()$. Fungsi trigamma ψ' merupakan turunan kedua dari $\ln\Gamma()$.

2.6 Penaksiran Parameter

Menurut Hilbe (2011), iterasi Newton-Raphson digunakan untuk penaksiran parameter regresi Poisson dan Binomial Negatif. Penurunan algoritma didasarkan pada modifikasi dua orde deret Taylor dengan fungsi *log likelihood*. Bentuk dari deret Taylor yaitu :

$$f(X_r) = f(X_{r-1}) + (X_r - X_{r-1})f'(X_{r-1}) + \frac{(X_r - X_{r-1})^2}{2!} f''(X_{r-1}) + \frac{(X_r - X_{r-1})^3}{3!} f'''(X_{r-1}) + \dots \quad (2.6.1)$$

Jika dipotong sampai orde kedua menjadi :

$$f(X_r) = f(X_{r-1}) + (X_r - X_{r-1})f'(X_{r-1}) \quad (2.6.2)$$

Untuk menentukan nilai X_r dapat menggunakan perluasan deret Taylor berupa $f(X_r) = 0$

$$0 = f(X_{r-1}) + (X_r - X_{r-1})f'(X_{r-1}) \quad (2.6.3)$$

atau

$$X_r = (X_{r-1}) - \frac{f(X_{r-1})}{f'(X_{r-1})} \quad (2.6.4)$$

Metode Newton-Raphson menerapkan estimasi di atas dengan menggunakan nilai turunan pertama dan kedua dari fungsi *log likelihood* sebagai dasar estimasi parameter, yaitu :

$$\beta_r = (\beta_{r-1}) - \frac{\partial \ell(\beta_{r-1})}{\partial^2 \ell(\beta_{r-1})}, \quad r = 1, 2, \dots, k \quad (2.6.5)$$

dengan β_r dan (β_{r-1}) adalah vektor untuk $\beta^T = [\beta_0, \beta_1, \dots, \beta_p]$ pada iterasi ke r dan ke $r - 1$ untuk iterasi Newton-Raphson. Turunan pertama dari fungsi *log likelihood*, $\partial \ell(\beta_{r-1})$, untuk regresi Poisson dan Binomial Negatif terdapat pada

Persamaan (2.2.7) dan (2.5.3). Turunan kedua dari fungsi *log likelihood*, $\partial^2 \ell(\beta_{r-1})$, untuk regresi Poisson dan Binomial Negatif terdapat pada Persamaan (2.2.8) dan (2.5.4)

Penaksiran parameter dispersi, untuk memaksimumkan $\partial \ell(y_i; \beta, \alpha)$ terhadap α pada regresi Binomial Negatif dapat menggunakan iterasi Newton-Raphson.

$$\alpha_r = (\alpha_{r-1}) - \frac{\partial \ell(\alpha_{r-1})}{\partial^2 \ell(\alpha_{r-1})} \quad (2.6.6)$$

Turunan pertama dari fungsi $\partial \ell(\alpha_{r-1})$ untuk regresi Binomial Negatif terdapat pada Persamaan (2.5.4). Turunan kedua dari fungsi $\partial^2 \ell(\alpha_{r-1})$ untuk regresi Binomial Negatif terdapat pada Persamaan (2.5.5).

Secara umum, model regresi poisson dan Binomial Negatif untuk menganalisis data adalah :

$$\ln(\mu(X_1, X_2, \dots, X_j)) = \beta_0 + \beta_1 x_2 + \dots + \beta_j X_j \quad (2.6.7)$$

atau

$$\begin{aligned} \mu(X_1, X_2, \dots, X_j) &= \exp(\beta_0 + \beta_1 x_2 + \dots + \beta_j X_j) \\ &= \exp(\beta_0 + \sum_{j=1}^p \beta_j X_j) \end{aligned} \quad (2.6.8)$$

Nilai *residual* didefinisikan sebagai perbedaan antara nilai yang diamati (Y) dengan nilai hasil prediksi ($\hat{Y}$) atau (μ). Sehingga, *Residual* = $(Y_i - \hat{Y}_i)$ atau *Residual* = $(Y_i - \mu_i)$.

2.7 Pengujian Signifikansi Model

Menurut Hosmer dan Lemeshow (2000), pengujian parameter regresi secara serentak dilakukan untuk menguji signifikansi dari koefisien regresi secara bersama-sama. Pengujian yang dilakukan adalah uji rasio *likelihood*. Statistik uji ini juga dapat digunakan sebagai *Goodness Of Fit* untuk kesesuaian model. Hipotesis yang digunakan :

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_j = 0$$

$$H_1 : \text{minimal satu } \beta_j \neq 0, j : 1, 2, 3, \dots, p$$

Statistik uji:

$$G = 2(\hat{\ell}_1 - \hat{\ell}_0) \quad (2.7.1)$$

dengan:

- $\hat{\ell}_1$: *log likelihood* untuk model yang mengandung seluruh variabel *independent*
- $\hat{\ell}_0$: *log likelihood* untuk model yang tidak mengandung variabel *independent*.

Pada kondisi H_0 statistik uji G mendekati distribusi *chi-square* (χ^2) dengan derajat bebas p . Aturan keputusannya adalah H_0 ditolak pada tingkat signifikansi 0.05 jika $G > \chi_{0.05,p}^2$. Penolakan H_0 pada tingkat signifikansi 0.05 berarti bahwa terdapat paling sedikit satu parameter diantara $\beta_1, \beta_2, \dots, \beta_j$ yang signifikan pada tingkat signifikansi 0.05 atau dapat dikatakan bahwa model regresi tersebut cocok untuk menjelaskan hubungan antara variabel *independent* dan *dependent* pada tingkat signifikansi 0.05.

2.8 Pengujian Signifikan Parameter Dalam Model

Pengujian signifikan parameter dilakukan untuk memeriksa pengaruh parameter regresi dari masing-masing variabel bebas secara individu pada model (Hosmer & Lemeshow, 2000). Pengujian yang dilakukan adalah uji *Wald*, dengan hipotesis :

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0, j : 1, 2, 3, \dots, p$$

Statistik uji :

$$W = \left(\frac{b_j}{\hat{S}_e(b_j)} \right)^2 \quad (2.8.1)$$

dengan,

b_j : Nilai taksiran parameter

$\hat{S}_e(b_j)$: nilai taksiran *standart error* dari b_j

Pada kondisi H_0 statistik uji W mendekati distribusi *chi-square* (χ^2) dengan derajat bebas 1. Aturan keputusannya adalah H_0 ditolak pada tingkat signifikansi 0.05 jika

$W > \chi_{0.05;1}^2$. Penolakan H_0 pada tingkat signifikansi 0.05 berarti bahwa variabel *independent* X_j untuk suatu j tertentu dengan $j = 1, 2, \dots, p$ memiliki kontribusi terhadap variabel *dependent* pada tingkat signifikansi 0.05.

2.9 Pengujian Signifikan Parameter Dispersi Pada Model

Pengujian signifikansi untuk parameter dispersi dilakukan untuk apakah model regresi Binomial Negatif baik dibandingkan dengan model regresi Poisson. Untuk pengujian parameter dispersi pada regresi Binomial Negatif hipotesis yang digunakan adalah :

$$H_0 \quad : \alpha = 0$$

$$H_1 \quad : \alpha > 0$$

dengan :

α : Parameter dispersi regresi Binomial Negatif

Misalkan taksiran *log likelihood* untuk model yang mengandung seluruh variabel *independent* pada regresi Poisson dan Binomial Negatif adalah $\hat{\ell}_P$ dan $\hat{\ell}_{BN}$. Pada model regresi Binomial Negatif, statistik uji yang digunakan adalah :

$$T = 2(\hat{\ell}_{BN} - \hat{\ell}_P) \tag{2.9.1}$$

Pada kondisi H_0 statistik uji T mendekati distribusi *chi-square* (χ^2) dengan derajat bebas 1. Aturan keputusannya adalah H_0 ditolak pada tingkat signifikansi 0.05 jika $T > \chi_{0.05;1}^2$. Penolakan H_0 pada tingkat signifikansi 0.05 berarti bahwa model regresi Binomial Negatif lebih tepat digunakan dibandingkan model regresi Poisson.

2.10 Kesesuaian Model Regresi

Pengukuran yang digunakan untuk membandingkan model tersebut untuk dicari model yang terbaik dapat menggunakan *likelihood ratio tests*, *Akaike Information Criteria* (AIC) dan *Bayesian Schwartz Information Criteria* (BIC) (Ismail, 2007).

Uji *likelihood ratio tests* (T) biasa digunakan untuk uji perbandingan model. Uji nilai *likelihood ratio tests* (T) yang menunjukkan penolakan H_0 berarti model tersebut lebih baik, karena mengindikasikan terjadinya overdispersi pada data. Nilai AIC dapat didefinisikan sebagai :

$$AIC = -2(\hat{\ell}_1 - k) \quad (2.10.1)$$

Nilai *log likelihood* untuk model yang mengandung seluruh variabel independen adalah $\hat{\ell}_1$. Banyaknya parameter termasuk konstanta dinyatakan dengan k . Dalam pengukuran ini, jika nilai $AIC_A < AIC_B$, maka model A lebih baik.

Nilai BIC dapat didefinisikan sebagai :

$$BIC = -2\hat{\ell}_1 + k \ln(n) \quad (2.10.2)$$

Nilai *log likelihood* untuk model yang mengandung seluruh variabel independen adalah $\hat{\ell}_1$. Banyaknya parameter termasuk konstanta dinyatakan dengan k . Dalam pengukuran ini, jika nilai $BIC_A < BIC_B$ maka model A lebih baik.