# II. LANDASAN TEORI

## 2.1 Model Regresi Poisson

Analisis regresi merupakan metode statistika yang populer digunakan untuk menyatakan hubungan antara variabel respon Y dengan variabel-variabel prediktor X. Regresi poisson adalah salah satu regresi yang dapat menggambarkan hubungan antara variabel respon Y dimana variabel respon berdistribusi poisson dengan variabel prediktor X. Model regresi poisson merupakan model standar untuk data diskrit dan termasuk dalam model linier. Regresi poisson adalah suatu bentuk model linear umum dimana variabel respon dimodelkan sebagai distribusi poisson. Regresi poisson merupakan suatu bentuk analisis menggunakan regresi untuk menduga model data seperti jumlah, perubahan nilai atau mengelompokan data ke tabel. Regresi poisson dapat dimodelkan mengunakan kombinasi *non*-linier $\exp(x'_i\beta)$ dari variabel-variabel yang diberikan:

$$E(Y_i|x_i) = \exp(x'_i\beta)$$

Penggunaan fungsi eksponensial untuk memastikan bahwa bagian sebelah kanan selalu positif, seperti yang kita harapankan dari nilai Y yang merupakan penjumlahan tidak mugkin negatif. Pengunaan fungsi eksponensial atau bisa kita sebut fungsi *link*, hanya untuk kemudahan. Pada prinsipnya dengan cara ini akan selalu menghasilkan nilai positif, tetapi dengan adanya eksponensial ini tidak ada hubungannya dengan model poisson. Dari model ini nilai β, yang merupakan parameter yang tidak diketahui. Nilai dugaan dari parameter-parameter dapatdiperoleh dengan metode *maximum likelihood*. Sebagai catatan bahwa dengan mengestimasi β maka dapat diestimasi juga keseluruhan dari distibusi dari Y terhadap x. Dengan ini regresi poisson memberikan suatu model yang realistis untuk berbagai macam fenomena acak poisson berupa bilangan bulat non negatif. Distribusi poisson disebut juga distribusi peristiwa yang jarang terjadi dalam suatu interval waktu tertentu, ditemukan oleh S.D. Poisson (1781-1841), seorang ahli matematika berkebangsaan perancis. Distribusi poisson termasuk distribusiteoritis yang memakai variabel random diskrit.Misalkan Y peubah acak yang terjadi dalam suatu interval waktu tertentu, (Hogg and Tanis,1997).

Maka fungsi peluang dari distribusi poisson diberikan sebagai berikut (Cameron dan trevedi, 1998).

$$\mathrm{p}(\mathrm{Y}, \lambda) = \frac{\lambda^{y} e^{-\lambda}}{y!} \; ; \mathrm{y} = 0,1,2, \ldots$$

Dimana λ adalah means distribusi poisson, means dan variansnya adalah:

$$\mathrm{E}(Y) = var(Y) = \lambda$$

Peluang banyaknya peubah acak Y dalam periode waktu t diberikan oleh:

$$\mathrm{p}(\mathrm{Y}, \lambda \mathrm{t}) = \frac{(\lambda \mathrm{t})^{y} e^{-\lambda \mathrm{t}}}{y!} \; ; \mathrm{y} = 0,1,2, \ldots$$

Persamaan diatas digunakan untuk menghitung peluang peubah acak Y, means jumlah kejadian $\mathrm{E}(\mathrm{y}) = \mathrm{var}(\mathrm{y}) = \lambda \mathrm{t}$, berdasarkan asumsi bahwa mean jumlah kejadian per periode waktu adalah konstan. Model regresi poisson dapat ditulis sebagai berikut:

$$\hat{y}_i = \mu_i + \varepsilon_i = t_i \exp(x'_i\beta) + \varepsilon_i$$

Dimana:

$y_i$ =jumlah kejadian ke ij, $i = 1,2, \ldots$

$\mu_i$ =means jumlah kejadian dalam periode $t_i$

$\varepsilon_i$ =galat error atau residual

## 2.2 Distribusi Gamma

Berdasarkan Hogg dan Craig (1995), suatu variabel acak kontinu $Y$ dikatakan berdistribusi Gamma dengan parameter $\alpha$ dan $\beta$ jika variabel tersebut mempunyai fungsi peluang sebagai berikut:

$$f(y) = \begin{cases} \dfrac{\beta^{\alpha}}{\Gamma(\alpha)} y^{\alpha-1} e^{-y\beta} \,; & 0 < y < \infty \\ 0 & ; \text{ untukylainnya} \end{cases}$$

Nilai mean dan variansnya adalah

$$E(y) = \frac{\alpha}{\beta}$$

$$V(y) = \frac{\alpha}{\beta^2}$$

dimana $\alpha, \beta > 0$.

Definisi fungsi gamma dari $\theta$ dan $\Gamma(\theta)$ yaitu:

$$\Gamma(\theta) = \int_0^\infty t^{\theta-1} e^{-t}\, dt = (\theta - 1)!$$

Untuk $\theta > 0$ dan nilai dari integral tersebut adalah bilangan positif.

Beberapa nilai dari fungsi gamma adalah

(i). Jika $\theta = 1$, maka $\Gamma(1) = \int_0^\infty e^{-t}\, dt = 1$.

(ii). Jika $\theta$ adalah suatu bilangan bulat positif yang lebih besar dari satu, maka diperoleh $\Gamma(\theta) = (\theta - 1)!$.

(iii). Jika $\theta > 1$, maka $\Gamma(\theta) = \int_0^\infty t^{\theta-1} e^{-t}\, dt = (\theta - 1)\Gamma(\theta - 1)$.

(iv). Untuk $c > 0$, maka $\frac{\Gamma(\theta+c)}{\Gamma(c)} = \frac{(c+\theta-1)(c+\theta-2)\ldots(c+1)c\Gamma(c)}{\Gamma(c)}$

$$= (c + \theta - 1)(c + \theta - 2) \ldots (c + 1)c$$

**2.3 Distribusi Binomial negatif**

Distribusi binomial negatif merupakan distribusi yang memiliki banyak sekali cara dalam hal pendekatannya. Pendekatan klasik dari distribusi binomial negatif yang sering digunakan adalah distribusi binomial negatif sebagai barisan percobaan Bernoulli, yaitu jumlah Bernoulli yang dibutuhkan sampai terjadi $k$ buah sukses, dimana setiap pengulangan saling bebas, dan peluang sukses setiap percobaan konstan yaitu $p$ sedangkan peluang gagal yaitu $1 - p$. Misalkan variabel acak $Y$ menyatakan jumlah percobaan yang dibutuhkan sampai terjadi

$k$buah sukses, maka $Y$ berdistribusi binomial negatif dengan fungsi peluang sebagai berikut:

$$f(y) = \binom{y-1}{r-1} p^r (1-p)^{y-1} \; ; y = k, k+1, \ldots$$

$$= 0 \qquad\qquad ; untuk\ y\ lainnya$$

Mean, varians, dan fungsi pembangkit momen dari distribusi binomial negatif adalah sebagai berikut:

1. $\mu = \frac{r}{p}$
2. $\sigma^2 = \frac{r(1-p)}{p^2}$
3. $3. M_y(t) = \frac{(pe^t)^r}{[1-(1-p)e^t]^r} \; ; (1-p)e^t < 1$

**2.4 Model regresi Binomial Negatif**

Distribusi binomial negatif merupakan model untuk menghitung jumlah suatu kejadian. Biasanya distribusi binomial negatif digunakan untuk menghitung probabilitas dari jumlah kegagalan yang terjadi sebelum berhasil. Tetapi karena merupakan kebalikan dari binomial maka dapat juga digunakan untuk menghitung jumlah kejadian, karena percobaan akan dilakukan terus menerus sampai berhasil.

Dalam distribusi poisson kita ketahui bahwa nilai mean sama dengan nilai variansnya. Namun dalam beberapa kasus, sering ditemukan bahwa nilai varians dari data yang teramati lebih besar dari pada meannya yang biasanya disebut *overdispersi*. Maka distribusi binomial negatif mempunyai peranan yang cukup penting dalam analisis statistika parametrik untuk mengatasi data yang mengandung *overdispersi*. Distribusi binomial negatif ini diperoleh dari proses pengintegralan dari distribusi campuran poisson-gamma terhadap $\delta_i$. Misalkan bahwa variabel acak $y_i$ berdistribusi poisson dengan parameter $\mu_i$ atau poisson($\mu_i$). Akan tetapi, $\mu_i$ itu sendiri merupakan peubah acak dan diasumsikan berdistribusi gamma yaitu:

$$y_i \sim poisson(\mu_i)$$

$$\mu_i \sim gamma(\alpha, \beta)$$

Jika suatu distribusi poisson ($\mu_i$) dimana $\mu_i$ merupakan nilai variabel random yang berdistribusi gamma, maka akan dihasilkan distribusi campuran yang dinamakan distribusi binomial negatif. Model regresi binomial negatif mengasumsikan terdapat peubah $\delta_i$ yang menyebar gammadengan nilai tengah 1 dan ragam $1/\theta_i$.Sehingga untuk memperoleh $E(\delta_i) = 1$jika parameter$\alpha = \beta = \theta_i$dalam nilai tengah sebaran poisson. Misalkan $v_i$ adalah sumber keragaman yang tidak teramati, sehingga nilai tengah sebaran campuran poisson-gamma adalah

$$E(y_i) = \tilde{\mu}_i = exp(x'_i\beta + v_i) = exp(x'_i\beta)exp(v_i) = \mu_i\delta_i$$

Dengan $\mu_i = exp(x'_i\beta)$ adalah nilai tengah model poisson dan $\delta_i = exp(v_i)$. Dengandiasumsikan$E(\delta_i) = 1$, maka model poisson dan binomial negatif memiliki nilai tengah yang sama, yaitu $E(y_i) = \mu_i E(\delta_i) = \mu_i$. Fungsi peluang sebaran campuran poisson-gamma dapat ditulis sebagai berikut:

$$f(y_i = \frac{(\mu_i\delta_i)^{y_i} exp(\mu_i\delta_i)}{y_i!}$$

peubah $\delta_i$ menyebar gamma dengan parameter $\alpha$ dan $\beta$. Fungsi peluang gamma adalah

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

dengan nilai harapan $E(\delta_i) = \frac{\alpha}{\beta}$ , sehingga untuk memperoleh $E(\delta_i) = 1$ maka parameter $\alpha = \beta$ditentukan sebesar $\theta_i$. Diasumsikan fungsi peluang gamma menjadi,

$$g(\delta_i) = \frac{{\theta_i}^{\theta_i}}{\Gamma(\theta_i)} \delta_i^{\theta_i - 1} exp(-\delta_i \theta_i)$$

Sehingga dapat ditulis bentuk fungsi marjinal dari distribusi campuran poisson-gamma adalah:

$$f(y_i) = \int_0^\infty f(y_i | \mu_i, \delta_i) \, . \, g(\delta_i) d\delta_i$$

Dari hasil integral untuk fungsi marjinal distribusi campuran poisson-gamma, maka diperoleh bentuk umum dari model regresi binomial negatif sebagai berikut:

$$f(y_i) = \frac{\Gamma(y_i + \theta_i)}{y_i! \, \Gamma(\theta_i)} \left( \frac{\theta_i}{(\mu_i + \theta_i)} \right)^{\theta_i} \left( \frac{\mu_i}{(\mu_i + \theta_i)} \right)^{y_i}$$

## 2.5Fungsi Link

Fungsi link adalah suatu fungsi yang menghubungkan fungsi prediktor linear $\eta$ dengan nilai tengah respon $\mu$. Dalam model linear klasik, fungsi link bisa berupa fungsi yang identik atau kanonik. Suatu fungsi link dikatakan fungsi link kanonik bila parameter kanoniknya sama dengan fungsi linknyayaitu :

$\eta = \theta$(2.4.1)

Dimana $\theta$ adalah parameter kanonik.

Berikut fungsi link kanonik untuk beberapa distribusi :

| **Distribusi** | **Fungsi link kanonik** |
|---|---|
| Normal | $\eta = \mu$ |
| Poisson | $\eta = log\ \mu$ |
| Binomial | $\eta = log\ \left( {}^{\mu}/_{(1-\mu)} \right)$ |
| Gamma | $\eta = \mu^{-1}$ |

Terdapat dua fungsi penghubung yang biasa digunakan dalam regresi binomial negatif yaitu penghubung identitas (*identity link)* dan penghubung log (*log link).*

Fungsi penghubung identitas berbentuk :

$g(\mu_i) = \mu_i = x_i^T \beta$(2.4.2)

Sedangkan fungsi penghubung log berbentuk :

$$g(\mu_i) = ln\left(\frac{\mu_i}{1-\mu_i}\right) = x_i^T \beta \qquad (2.4.3)$$

Fungsi penghubung log adalah fungsi yang paling cocok digunakan, karena fungsi log menjamin bahwa nilai variabel yang diharapkan dari variabel responnya akan bernilai non negatif.

**2.6Metode Kemungkinan Maksimum (*Method of Maximum Likelihood*)**

**Definisi 2.6.1:** (Hogg and Craig, 1995)

Fungsi densitas bersama dari variabel random $X_1, X_2, \ldots, X_n$ yang bernilai $x_1, x_2, \ldots, x_n$ adalah$L(\theta) = f(x_1, x_2, \ldots, x_n; \theta)$ yang merupakan fungsi likelihood. Untuk $x_1, x_2, \ldots, x_n$ tetap, fungsi likelihood merupakan fungsi dari $\theta$ dan dilambangkan dengan $L(\theta)$. Jika $X_1, X_2, \ldots, X_n$ mewakili sebuah sampel random dari $f(x; \theta)$, maka $L(\theta) = f(x_1; \theta). f(x_2; \theta) \ldots f(x_n; \theta)$ dapat dituliskan sebagai berikut:

$$
\begin{aligned}
L(\theta) &= f(\tilde{x}; \theta) \\
&= f(x_1, x_2, \ldots, x_n; \theta) \\
&= f(x_1; \theta). f(x_2; \theta) \ldots f(x_n; \theta) \\
&= \prod_{i=1}^{n} f(x_i; \theta)
\end{aligned}
$$

$L(\theta) = f(x_1, x_2, \ldots, x_n)$, $\theta \epsilon \Omega$ merupakan fungsi densitas probabilitas dari $x_1, x_2, \ldots, x_n$. Untuk hasil pengamatan $x_1, x_2, \ldots, x_n$, nilai $\hat{\theta}$ berada dalam $\Omega(\hat{\theta} \epsilon \Omega)$, dimana $L(\theta)$ maksimum yang disebut sebagai *maximum likelihood estimation* (MLE) dari $\theta$. Jadi, $\hat{\theta}$ merupakan nilai dugaan dari $\theta$.

Jika $f(x_1, x_2, \dots, x_n) = max\, f(x_1, x_2, \dots, x_n; \theta)$; $\theta \epsilon \Omega$, maka untuk memperoleh nilai $\hat{\theta}$ tersebut yang memaksimumkan $L(\theta)$ harus diderivatifkan dengan langkah-langkah sebagai berikut:

1. Nilai $\hat{\theta}$ diperoleh dari derivatif pertama jika:

$$\frac{\partial}{\partial \theta} L(\theta)|_{\theta=\hat{\theta}} = 0$$

2. Nilai $\hat{\theta}$ dikatakan memaksimumkan $L(\theta)$ jika:

$$\frac{\partial^2}{\partial^2 \theta} L(\theta)|_{\theta=\hat{\theta}} < 0$$

Selain dengan memaksimumkan fungsi likelihood, nilai $\hat{\theta}$ juga dapat diperoleh dengan memaksimumkan fungsi *log-likelihood*, karena dengan memaksimumkan fungsi *log-likelihood*, juga akan memaksimumkan fungsi likelihood, sebab log $L(\theta)$ merupakan fungsi yang menoton naik, maka untuk memperoleh $\hat{\theta}$ dengan memaksimumkan fungsi *log-likelihood* dapat dilakukan dengan langkah-langkah yang sama yaitu:

1. Nilai $\hat{\theta}$ diperoleh dari derivatif pertama jika:

$$\frac{\partial}{\partial \theta} log\, L(\theta)|_{\theta=\hat{\theta}} = 0$$

2. Nilai $\hat{\theta}$ dikatakan memaksimumkan $L(\theta)$ jika:

$$\frac{\partial^2}{\partial^2 \theta} log\, L(\theta)|_{\theta=\hat{\theta}} < 0$$

## 2.7Metode Iterasi *Newton Rhapson*

Apabila dalam proses estimasi parameter didapat persamaan akhir yang non linear maka tidak mudah memperoleh estimasi parameter tersebut, sehingga diperlukan suatu metode numerik untuk memecahkan persamaan non linear tersebut. Salah satu metode yang sangat populer digunakan untuk memecahkan sistem persamaan non linear adalah metode *Newton Rhapson.* Metode *Newton Rhapson* adalah metode untuk menyelesaikan persamaan non linear secara iteratif seperti persamaaan *likelihood* yang mencari lokasi yang memaksimalkan suatu fungsi.

Dasar dari metode ini adalah pendekatan deret taylor linear :

$$f(x) = f(x^0) + \sum_{i=1}^{p} \frac{1}{i!} \frac{\partial^p f(x^0)}{\partial (x^0)^i} (x - x^0)^i$$

Perluasan dari bentuk orde 1 :

$$\frac{\partial F(\theta)}{\partial \theta} = 0$$

Diperoleh :

$$\frac{\partial f(\theta)}{\partial \theta} = f(\theta^0) + H^0(\theta - \theta^0)$$

$$= G^0 + H^0(\theta - \theta^0)$$

Jika $\theta^0$ merupakan nilai awal (inisialisasi) dari θ atau $\theta^0$merupakan nilai ke-1 dari θ, maka dapat dimisalkan $\theta^0 = \theta^t$ dan $\theta = \theta^{t+1}$ dengan $t$ awal $= 0$. Begitu pula dengan G dan H. Maka diperoleh iterasi sebagai berikut:

$$\theta^{t+1} = \theta^t - (H^t)^{-1} G^t$$

Dengan indeks *t* menyatakan ukuran iterasi.

Adapun langkah-langkah metode iterasi *Newton Rhapson* adalah sebagai berikut:

1. Ambil estimasi awal dari θ, misal $\theta^0$.
2. $\hat{\theta}^1 = \theta^0 - \frac{G(\hat{\theta}^0)}{H(\hat{\theta}^0)}$ , $G(\hat{\theta}^0)$ merupakan derivative pertama dari $f(\theta)$ pada $\theta = \hat{\theta}^t$.
3. $\hat{\theta}^{t+1} = \hat{\theta}^t - \frac{G(\hat{\theta}^t)}{H(\hat{\theta}^t)}$, misal $H\left(\hat{\theta}^t\right) = H^t$ dan $G\left(\hat{\theta}^t\right) = G^t$, maka :

   $\hat{\theta}^{t+1} = \hat{\theta}^t - (H^t)^{-1} G^t$
4. Estimator $\hat{\theta}^t$ diiteratif terus sampai diperoleh jarak antara $\hat{\theta}^{t+1}$ dengan $\hat{\theta}^t$ nilainya sangat kecil atau $\hat{\theta}^{t+1} - \hat{\theta}^t \approx \varepsilon$

Metode ini dapat diperluas untuk menyelesaikan sistem persamaan dengan lebih dari satu parameter. Misal $\theta_1, \theta_2, \dots, \theta_p$ maka iterasinya sebagai berikut :

$$\hat{\theta}^{t+1} = \hat{\theta}^t - (H^t)^{-1} G^t$$

dimana $\hat{\theta}^{t+1}$ dan $\hat{\theta}^t$ dalam bentuk vektor yaitu :

$$\hat{\theta}^{t+1} = \begin{bmatrix} \hat{\theta}_1^{\ t+1} \\ \vdots \\ \hat{\theta}_p^{\ t+1} \end{bmatrix} \text{dan } \hat{\theta}^t = \begin{bmatrix} \hat{\theta}_1^{\ t} \\ \vdots \\ \hat{\theta}_p^{\ t} \end{bmatrix}$$

$$H = \begin{bmatrix} \frac{\partial^2 F(\theta)}{\partial \theta_1^2} & \frac{\partial^2 F(\theta)}{\partial \theta_1 \partial \theta_2} & \cdots & \frac{\partial^2 F(\theta)}{\partial \theta_1 \partial \theta_p} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 F(\theta)}{\partial \theta_p^2} & \frac{\partial^2 F(\theta)}{\partial \theta_p \partial \theta_2} & \cdots & \frac{\partial^2 F(\theta)}{\partial \theta_p^2} \end{bmatrix} \text{dan} \quad G = \begin{bmatrix} \frac{\partial F(\theta)}{\partial \theta_1} \\ \vdots \\ \frac{\partial F(\theta)}{\partial \theta_p} \end{bmatrix}$$