BAB II
LANDASAN TEORI

2.1 Data Statistik

Data adalah sekumpulan fakta-fakta yang biasa disajikan dengan angka-angka yang saling berhubungan satu sama lain (Qudratullah dkk, 2012). Menurut Kamus Besar Bahasa Indonesia (KBBI) pengertian dari data adalah keterangan yang benar dan nyata. Agar mendapat kesimpulan tepat dan benar makan data yang dikumpulkan harus dalam pengamatan yang nyata dan benar, berikut syarat-syaratnya:

1. Data harus objektif
2. Data harus mewakili (representative)
3. Data harus update
4. Data harus relevan

Berikut jenis-jenis data:

1. Dari segi bentuknya data

Ada 2 (dua) jenis data, yaitu data *kualitatif* dan data *kuantitatif*. Data *kualitatif* adalah data yang dinyatakan dalam bentuk bukan angka, sedangkan data kuantitatif adalah data yang dinyatakan dalam bentuk angka, yang kemudian data kuantitatif terbagi lagi menjadi 2 (dua) yaitu data cacahan (diskrit) dan data ukuran (kontinu).

2. Dari segi pengumpulannya

Data dibagi menjadi 3 (tiga) jenis,yaitu data yang dikumpulkan pada suatu waktu tertentu yang disebut data *cross section*, data yang dikumpulkan dari waktu ke waktu yang disebut data berkala atau runtun waktu (*time series*), dan gabungan antara data *cross section* dan runtun waktu yang disebut data panel.

3. Dari segi sumbernya

Data di bagi menjadi 2 (dua) yaitu data interen dan data eksteren. Data interen adalah data yang dikumpulkan oleh suatu badan itu sendiri, sedangkan data eksteren adalah data yang diperoleh atau bersumber dari luar badan tersebut.

Sementara itu data eksteren dibagi menjadi 2 (dua) jenis, yaitu data primer dan data sekunder. Data primer yaitu data yang diperoleh langsung dari objek penelitian melalui proses perhitungan atau pengukuran secara langsung, sedangkan data sekunder data yang tidak langsung dikumpulkan melainkan data dari media yang sudah diterbitkan atau dipublikasikan.

Pada analisis statistik terdapat empat skala pengukuran yang harus diketahui, yaitu skala nominal, ordinal, interval, rasio.

1. Skala Nominal

Skala nominal merupakan skala pengelompokan, kategorisasi, identifikasi kejadian atau fenomena ke dalam kelas-kelas sehingga yang masuk satu kelas atau kategori merupakan sama dalam hal sifat. Kelas atau kategori hanya nama yang membedakan suatu kejadian/peristiwa satu dengan yang lain. Skala ini bersifat kuantitatif.

2. Skala Ordinal

Tergolong jenis data kuantitatif, data diperoleh dari observasi, pengamatan, atau angket berskala dari suatu variabel. Berbeda halnya dengan nominal, ordinal mengenal suatu urutan menurut kualitas atributnya.

3. Skala Interval

Tergolong jenis data kuantitatif, data diperoleh dari hasil pengukuran , berbentuk bilangan kontinu, dan tidak memiliki nilai mutlak.

4. Skala Rasio

Tergolong jenis data kuantitatif, data diperoleh dari hasil pengukuran, bentuk bilangan kontinu. Perbedaan dengan data interval adalah data rasio memiliki nilai nol mutlak.

2.2 Statistik Deskriptif

Statistik deskriptif adalah statistik yang berfungsi untuk mendeskripsikan atau memberi gambaran terhadap obyek yang diteliti melalui data sampel atau populasi sebagaimana adanya, tanpa melakukan analisis dan membuat kesimpulan yang berlaku untuk umum (Sugiyono, 2015).

Pada statistik deskriptif akan dikemukakan cara cara penyajian data, dengan tabel biasa maupun distribusi frekuensi; grafik garis maupun batang;

diagram lingkaran; piktogram; penjelasan kelompok melalui modus, median, *mean* dan variansi kelompok melalui rentang dan simpangan baku (Sugiyono,2015).

Mean (Rataan) didefinisikan sebagai jumlah seluruh data dibagi dengan banyaknya data (Walpole, 14).

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} \tag{2.1}$$

Variansi adalah suatu nilai yang menunjukkan ukuran variabilitas yang dihitung dengan cara mengkuadratkan standar deviasi (Walpole, 14).

$$S^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1} \tag{2.2}$$

2.3 Variabel Acak (Variabel Random)

2.3.1 Pengertian Variabel Acak

Variabel adalah karakteristik yang bervariasi. Sedangkan variabel acak adalah nilai yang didapatkan dari percobaan atau eksperimen secara acak (random), dimana peneliti tidak tahu variasi apakah yang akan keluar. Artinya nilai tersebut tidak dapat diperkirakan nilai yang keluar dan tidak selalu nilai yang keluar pada setiap percobaan atau eksperimen adalah sama atau berbeda. Jadi variabel acak disebut sebagai variabel acak jika variabel tersebut menghasilkan nilai yang bisa berbeda pada setiap peristiwa atau percobaan dan perubahan hasil di setiap peristiwa tidak dapat diperkirakan (Budiarto, 2012)

Definisi 2.3.1.1

Suatu fungsi bernilai real yang harganya ditentukan oleh setiap anggota dalam ruang sampel disebut suatu peubah acak. peubah acak (random variabel) dinyatakan dengan huruf kapital, sedangkan nilainya dinyatakan dengan huruf kecil. Biasanya ditandai dengan huruf " X " yang memiliki sebuah nilai numerik tunggal untuk setiap keluaran dari sebuah eksperimen probabilitas.

Jika X variabel acak, maka nilainya dinyatakan dengan x. jika peluang kejadian X bernilai kurang dari atau sama dengan x dinyatakan dengan

$P(X \le x)$. Jadi X dapat bernilai berapapun tergantung pada keluaran yang mungkin dihasilkan dari eksperimen.

Misal dalam suatu percobaan yang acak/random, outcome (keluaran) yang muncul dapat diberi nilai misal 0, 1, 2. Keluarnya bilangan 0, 1, 2 dalam setiap percobaan tersebut selalu terjadi secara acak.

Definisi 2.3.1.2

Jika suatu ruang sampel mengandung titik yang berhingga banyaknya atau suatu deretan anggota yang banyaknya sama dengan banyaknya bilangan bulat, maka ruang sampel itu disebut ruang sampel diskrit dan peubah acak yang didefinisikan pada ruang sampel tersebut adalah peubah acak diskrit.

Definisi 2.3.1.3

Bila ruang sampel mengandung titik sampel yang berhingga banyaknya dan sama banyaknya dengan banyak titik pada sepotong garis, maka ruang sampel itu disebut ruang sampel kontinu dan peubah acak yang didefinisikan diatasnya disebut peubah acak kontinu

2.3.2 Distribusi Peluang Diskrit

Suatu peubah acak diskrit mendapat tiap nilai dengan peluang tertentu

Definisi 2.3.2.1

Fungsi $f(x)$ adalah suatu fungsi peluang atau distribusi peluang suatu peubah acak diskrit X jika untuk setiap hasil x yang mungkin:

i. $f(x) \ge 0$

ii. $\sum_x f(x) = 1$

iii. $P(X = x) = f(x)$

Definisi 2.3.2.2

Distribusi kumulatif $f(x)$ suatu peubah acak X dengan distribusi peluang $f(x)$ dinyatakan oleh $F(x) = P(X \le x) = \sum_{t \le x} f(t)$

2.3.3 Distribusi peluang kontinu

Suatu peubah acak kontinu mempunyai peluang nol pada setiap titik x . karena itu, distribusi peluangnya tidak mungkin disajikan dalam bentuk tabel. Bila x kontinu maka:

$$\begin{aligned} P(a \le X \le b) &= P(X = a) + P(a < X < b) + P(X = b) \\ &= P(a < X < b) \\ &= \int_{-\infty}^{\infty} f(x)dx \end{aligned}$$

Definisi 2.3.3.1

Fungsi $f(x)$ adalah fungsi padat peluang peubah acak kontinu X yang didefinisikan di atas himpunan semua bilangan real R bila:

i. $f(x) \ge 0$ *untuk semua* $x \in R$

ii. $\int_{-\infty}^{\infty} f(x)dx = 1$

iii. $P(a < x < b) = \int_a^b f(x)dx$

2.3.4 Distribusi Peluang Gabungan

Bila X dan Y dua peubah acak, distribusi peluang terjadinya secara serentak dapat dinyatakan dengan fungsi $f(x, y)$. Biasanya $f(x, y)$ dinamakan distribusi peluang gabungan X dan Y . Jadi dalam kasus diskrit yang dapat didaftar $f(x, y) = P(X = x, Y = y)$

Definisi 2.3.4.1

Fungsi $f(x, y)$ adalah fungsi peluang gabungan peubah acak diskrit X dan Y bila:

i. $f(x, y) \ge 0$ *untuk semua* (x, y)

ii. $\sum_x \sum_y f(x, y) = 1$

iii. $P[(X, Y) \in A] = \sum_A \sum f(x, y)$ *untuk setiap daerah* A *di bidang* xy

Definisi 2.3.4.2

Fungsi $f(x,y)$ adalah fungsi peluang gabungan peuabh acak kontinu X dan Y bila:

i. $f(x,y) \geq 0 \text{ untuk semua } (x,y)$

ii. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dx\,dy = 1$

iii. $P[(X,\mathrm{Y}) \in \mathrm{A}] = \int_A\int f(x,y)dx\,dy \text{ untuk tiap daerah } A \text{ di bidang } xy$

2.3.5 Distribusi Marginal

Jika $f(x,y)$ diketahui maka kita dapat mencari distribusi peluang X saja dan Y saja, yaitu:

$$g(x) = \begin{cases} \sum_y f(x,y), & jika\ diskrit \\ \int_{-\infty}^{\infty} f(x,y)\,dy, & jika\ kontinu \end{cases} \quad \text{(Distribusi marginal X)}$$

$$h(y) = \begin{cases} \sum_x f(x,y), & jika\ diskrit \\ \int_{-\infty}^{\infty} f(x,y)\,dy, & jika\ kontinu \end{cases} \quad \text{(Distribusi marginal Y)}$$

2.3.6 Distribusi Bersyarat

Nilai dari variabel random sebenarnya adalah kejadian yang merupakan himpunan bagian dari ruang sampel, sehingga jika A dan B merupakan kejadian yang ditentukan oleh masing-masing $X = x$, $Y = y$ maka dari definisi peluang bersyarat $P(A|B) = \frac{P(A \cap B)}{P(A)}$, didapat

$$P(Y = y|X = x) = \frac{P(X = x, Y = x)}{P(X = x)} = \frac{f(x,y)}{g(x)} \tag{2.3}$$

Atau sering ditulis $f(x|y) = \frac{f(x,y)}{h(y)}$, $f(y|x) = \frac{f(x,y)}{g(x)}$

Perhatikan jika X dan Y bebas maka x tidak tergantung sehingga

$f(x|y) = f(x)$

Jadi $f\left(x|y\right)=\dfrac{f(x,y)}{h(y)}=g(x)$, diperoleh $f(x,y)=g(x)h(y)$

2.4 Distribusi Peluang Diskrit

2.4.1 Distribusi Uniform

Distribusi peluang diskrit yang paling sederhana adalah distribusi uniform (seragam). Distribusi tersebut merupakan distribusi variabel random diskrit yang mengasumsikan bahwa semua nilai mempunyai kemungkinan yang sama untuk muncul.

Definisi 2.4.1.1

Jika variabel random diskrit X mempunyai $x_1, x_2, \ldots, x_k$ maka peluang setiap titik sampel sama, sehingga

$$f(x;k)=\frac{1}{k}, untuk\ x=x_1,x_2,\ldots,x_k \tag{2.4}$$

Distribusi uniform memiliki nilai tengah dan variansi sebagai berikut:

Nilai tengah : $\mu=\dfrac{\sum_{i=1}^{k}x_i}{k}$

Variansi : $\sigma^2=\dfrac{\sum_{i=1}^{k}(x_i-\mu)^2}{k}$

2.4.2 Distribusi Binomial

Distribusi binomial adalah suatu percobaan atau eksperimen yang mempunyai dua buah kejadian, yaitu sukses atau gagal. Eksperimen binomial umumnya mempunyai 4 syarat, yaitu:

a. Banyaknya eksperimen merupakan bilangan tetap (*fixed number of trial*)
b. Setiap eksperimen mempunyai dua buah hasil yang diketegorikan menjadi sukses atau gagal. Dikatakan sukses apabila sesuai dengan harapan dan keinginan
c. Probabilitas sukses sama pada setiap eksperimen atau percobaan
d. Eskperimen tersebut harus bebas (*independent*) satu sama lain, artinya hasil eksperimen yang satu tidak mempengaruhi yang lain.

Definisi 2.4.2.1

Jika suatu eksperimen binomial memiliki peluang sukses p dan peluang gagal q maka variabel random X, yaitu banyaknya sukses dalam eksperimen tersebut adalah:

$$b(x;n,p)=\binom{n}{x}p^{x}q^{n-x}, untuk\ x=0,1,2,...,n\ dan\ q=1-p \tag{2.5}$$

Nilai tengah : $\mu = np$

Variansi : $\sigma^2 = npq$

2.4.3 Distribusi Multinomial

Distribusi Multinomial adalah perluasan dari distribusi Binomial. Bila Distribusi Binomial adalah suatu eksperimen yang mempunyai 2 kejadian, yaitu “sukses atau gagal”. Sedangkan Distribusi Multinomial adalah suatu eksperimen yang mempunyai 3 atau lebih kejadian yaitu “sukses, nyaris sukses, dan gagal”. Seperti peristiwa keadaan cuaca dapat digolongkan dengan menjadi cerah, mendung, atau hujan.

Definisi 2.4.3.1

Misal dalam sebuah eksperimen menghasilkan peristiwa $E_1, \mathrm{E}_2, ..., E_k$ dengan peluang $\pi_1 = P(E_1), \pi_2 = P(E_2), ..., \pi_k = P(E_k)$ dengan $\pi_1 + \pi_2 + ... + \pi_k = 1$. Distribusi Multinomial memberikan probabilitas terjadinya X_1 kali kejadian E_1, X_2 kali kejadian E_2, X_k kali kejadian E_k dalam n ulangan yang bebas dengan $\mathrm{X}_1 + \mathrm{X}_2 + ... + \mathrm{X}_k = n$

Sehingga probabilitas Distribusi Multinomial dinyatakan:

$$P(x_1, x_2, ..., x_k) = \frac{n!}{x_1!, x_2!, ..., x_k!}\pi_1^{x_1}\pi_2^{x_2}...\pi_k^{x_k} \tag{2.6}$$

2.4.4 Distribusi Hypergeometrik

Distribusi Hypergeometrik adalah peluang dari suatu kejadian dari 1 populasi yang berukuran N dan diantara N terdapat D kategori tertentu dan dari populasi ini diambil berukuran n . Suatu Distribusi Hypergeometrik dibentuk oleh suatu percobaan yang memenuhi kondisi-kondisi berikut:

a. Populasi berukuran N (anggota terdiri dari N objek)

b. Setiap anggota populasi dapat dinyatakan sebagai sukses atau gagal dan terdapat D buah sukses dalam populasi, jadi $p = \frac{D}{N}$

c. Suatu sample berukuran n (anggota terdiri dari n objek) dipilih dari x populasi tanpa pergantian dimana setiap himpunan bagian beranggota n yang dapat dibentuk dari populasi memiliki kesempatan yang sama untuk terpilih menjadi sampel.

Variabel random X didefinisikan sebagai banyaknya objek yang sukses dalam eksperimen hypergeometrik. Distribusi peluang hypergeometrik merupakan distribusi peluang dari variabel random X dengan rumus:

$$h(x; N, n, k) = \frac{C_x^k C_{n-x}^{n-k}}{C_n^N}, x = 0, 1, 2, ..., \min(n, k) \tag{2.7}$$

Nilai tengah : $\mu = n\frac{k}{N}$

Variansi : $\sigma^2 = n\frac{k}{N}\frac{N-n}{N-1}\left(1 - \frac{k}{N}\right)$

2.4.5 Distribusi Poisson

Distribusi Poisson merupakan suatu distribusi untuk peristiwa yang probabilitas kejadiannya kecil, dimana kejadian tergantung pada selang waktu tertentu atau di suatu daerah tertentu dengan hasil pengamatan berupa variabel diskrit dan antar variabel prediktor saling independen. Selang waktu tersebut dapat berupa berapa saja panjangnya, missal semenit, sehari, seminggu, sebulan, bahkan setahun. Daerah tertentu yang di maksudkan dapat berupa suatu garis, suatu luasan, suatu volume, atau mungkin sepotong bahan (Walpole, 1995).

Distribusi poisson memiliki ciri-ciri sebagai berikut:

1. Banyaknya percobaan yang terjadi dalam selang waktu atau suatu daerah tertentu tidak tergantung pada banyaknya hasil percobaan yang terjadi pada selang waktu atau daerah lain yang terpisah.
2. Peluang terjadinya suatu hasil percobaan selama selang waktu yang singkat sekali atau dalam suatu daerah yang kecil. Sebanding dengan panjang selang waktu tersebut atau besarnya daerah tersebut dan tidak bergantung pada banyak hasil percobaan yang terjadi diluar selang waktu dan daerah tertentu.
3. Peluang bahwa lebih dari satu hasil percobaan akan terjadi dalam selang waktu yang singkat tersebut atau dalam daerah yang kecil tersebut dapat diabaikan.

Fungsi peluang untuk data berdistribusi poisson bergantung pada parameter tunggal, yaitu μ. Fungsi peluangnya adalah sebagai berikut:

$$f(\mathrm{x};\mu)=\frac{e^{-\mu}\mu^{x}}{\mathrm{x}!} \tag{2.8}$$

Dimana:

$x=0,1,2,\dots$

$\mu > 0$, menyatakan banyaknya sukses yang terjadi dalam selang waktu atau daerah tertentu tersebut

$e=2,7183$

Distribusi Poisson mempunyai rata-rata dan variansi keduanya sama dengan μ.

Teorema 2.4.5.1

Misalkan Y variabel random berdistribusi poisson dengan parameter μ maka rata-rata dan variansi Y adalah μ

Bukti:

$$E(Y) = \sum_{x=0}^{\infty} \mathrm{x} \,.\, f(\mathrm{x}, \mu)$$
$$= \sum_{x=0}^{\infty} \mathrm{x} \,.\, \frac{e^{-\mu}\mu^{x}}{\mathrm{x}!}$$
$$= \sum_{x=0}^{\infty} \mathrm{x} \,.\, \frac{e^{-\mu}\mu^{x}}{\mathrm{x}(\mathrm{x}-1)!}$$
$$= \sum_{x=0}^{\infty} \frac{e^{-\mu}\mu^{x}}{(\mathrm{x}-1)!}$$
$$= \sum_{x=0}^{\infty} \frac{\mu e^{-\mu}\mu^{(\mathrm{x}-1)}}{(\mathrm{x}-1)!}$$
$$= \mu \sum_{x=0}^{\infty} \frac{e^{-\mu}\mu^{(\mathrm{x}-1)}}{(\mathrm{x}-1)!}$$
$$= \mu \sum\nolimits_{z=0}^{\infty} \frac{e^{-\mu}\mu^{z}}{z!} \, (misal\; z = x-1, \mathrm{x} = 1 maka\; z = 0)$$
$$= \mu \sum\nolimits_{z=0}^{\infty} f(z;\mu)$$
$$= \mu.1$$
$$= \mu$$

Sedangkan variansi dari Y yang berdistribusi poisson dengan parameter μ adalah

$$V(Y) = E(Y^2) - \left[E(Y)\right]^2$$
$$= E(Y^2 - Y + Y) - \left[E(Y)\right]^2$$
$$= E\left[Y(Y-1) + Y\right] - \left[E(Y)\right]^2$$
$$= E\left[Y(Y-1)\right] + E(Y) - \left[E(Y)\right]^2$$
$$= \sum_{x=0}^{\infty} y(y-1).\frac{e^{-\mu}\mu^{y}}{y!} + E(Y) - \left[E(Y)\right]^2$$
$$= \sum_{y=0}^{\infty} \mathrm{x}(\mathrm{x}-1).\frac{e^{-\mu}\mu^{x}}{\mathrm{x}(\mathrm{x}-1)(\mathrm{x}-2)!} + E(Y) - \left[E(Y)\right]^2$$

$$= \sum_{x=0}^{\infty} \frac{e^{-\mu}\mu^{x}}{(\mathrm{x}-2)!} + E(Y) - \left[E(Y)\right]^{2}$$

$$= \sum_{x=2}^{\infty} \frac{\mu^{x-2}e^{-\mu}\mu^{2}}{(\mathrm{x}-2)!} + E(Y) - \left[E(Y)\right]^{2}$$

$$= \mu^{2}\sum_{x=2}^{\infty} \frac{\mu^{x-2}e^{-\mu}}{(\mathrm{x}-2)!} + E(Y) - \left[E(Y)\right]^{2}$$

$$= \mu^{2}\sum_{z=0}^{\infty} \frac{\mu^{x-2}e^{-\mu}}{(\mathrm{x}-2)!} + E(Y) - \left[E(Y)\right]^{2} (misal\ z = x-2, \mathrm{x} = 2\ maka\ z = 0)$$

$$= \mu^{2} + \mu - \mu^{2} = \mu$$

Beberapa data cacah yang memungkinkan model distribusi poisson dapat digunakan

a. Jumlah kasus penyakit χ pada suatu Negara dalam interval waktu tertentu.
b. Jumlah kasus kematian ibu dan bayi pada suatu daerah dalam interval waktu tertentu
c. Jumlah kecelakaan lalu lintas dalam interval waktu tertentu.
d. Jumlah kasus HIV/AIDS pada suatu daerah dalam interval waktu tertentu.

2.5 Distribusi Normal

Distribusi peluang kontinu yang paling sering digunakan adalah distribusi normal (Distribusi Gaussian). Distribusi normal dituliskan dengan rumus:

$$n(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^{2}}, untuk -\infty < x < \infty \tag{2.10}$$

Distribusi variabel random dilambangkan dengan Z , memiliki *mean* dan variansi yang berbeda, nilai variabel random dilambangkan dengan X . Bentuk rumusnya adalah:

$$Z = \frac{X-\mu}{\sigma}$$

2.6 Pendugaan Parameter

Populasi adalah totalitas dari semua objek atau individu yang memiliki karakteristik tertentu, jelas dan lengkap yang akan diteliti. Sampel adalah bagian dari populasi yang diambil melalui cara-cara tertentu yang juga memiliki karakteristik tertentu, jelas dan lengkap yang dianggap bias mewakili populasi (Hasan, 2002)

Guna menerangkan karakteristik dari populasi dan sampel, digunakan istilah parameter dan statistic. Parameter dan statistic adalah besaran yang berupa data ringkasan atau angka ringkasan yang menunjukkan suatu ciri dari pupulasi dan sampel. Parameter dan statistik merupakan hasil hitungan nilai dari semua unit di dalam populasi dan sampel bersangkutan (Hasan, 2002)

Pendugaan (estimasi) adalah proses menggunakan sampel statistik untuk menduga atau menaksir hubungan parameter populasi yang tidak diketahui. Pendugaan merupakan suatu pernyataan mengenai parameter populasi yang diketahui berdasarkan informasi dari sampel, dalam hal ini sampel random yang diambil dari populasi bersangkutan. Jadi dengan pendugaan itu keadaan populasi dapat diketahui (Hasan, 2002)

Penduga (estimator) adalah suatu statistic (harga sampel) yang digunakan untuk menduga suatu parameter. Dengan pendugaan, dapat diketahui jauh suatu parameter populasi yang tidak diketahui berada disekitar sampel (statistik sampel).

2.7 Metode *Maximum Likelihood Estimation* (MLE)

Metode *Maximum Likelihood* adalah salah satu metode estimasi titik yang berperan dalam statistic. Pertama kali dikenal pada tahun 1920 oleh R. A. Fisher. Konsep dasar dari estimasi parameter metode Maximum Likelihood yaitu dengan memaksimumkan fungsi likelihood untuk menentukan nilai penduga (Wibisono, 2009).

Metode *Maximum Likelihood* dikenal sebagai metode sampel besar yang merupakan aplikasi yang lebih luas diaplikasikan dalam model regresi nonlinear dalam suatu parameter. (Gujarati, Porter;2010).

Diketahui n data pasangan $\{(x_1, y_1),...,(x_n, y_n)\}$ yang saling bebas (independen) dan diasumsikan bahwa untuk $\mathrm{X} = x$ fungsi kepadatan peluang Y adalah

$$Y|x \sim f(y|\theta)$$

Dengan θ adalah parameter yang merupakan fungsi dari x , yaitu $\theta = s(x)$ di mana $s(x)$ adalah fungsi penghalus, maka fungsi likelihood dari $Y|x$ adalah

$$L(\theta|x, y) = \prod_{i=1}^{n} f(y_1|s(x_i))$$

Untuk $x = \{x_1, x_2, ..., x_n\}$ dan $y = \{\mathrm{y}_1, \mathrm{y}_2, ..., \mathrm{y}_n\}$

Kemudian menduga $s(x_i)$ berdasarkan likelihood lokal dengan meggunakan pendekatan linear lokal dengan bentuk $s(x_i) = \beta_{0i} + \beta_{1i}x_i$. Sehingga diperoleh penduga likelihood lokal untuk $s(x_i)$ yaitu

$$s(x_i) = \beta_{0i} + \beta_{1i}x_i \tag{2.13}$$

Dengan β_{0i} dan β_{1i} nilai yang memaksimumkan fungsi likelihood lokal

$$\ell_i(\beta_{0i}, \beta_{1i}|x, y, h) = \sum_{i=1}^{n} \left\{ \ln(f(y_j|\beta_{0i} + \beta_{1i}x_j)) \frac{1}{h} K\left(\frac{x_i - x_j}{h}\right) \right\}; i = 1, 2, ..., n \tag{2.14}$$

2.7.1 Maximum Likelihood Uniparameter

Menurut Aziz (2010), Apabila diberikan variabel acak X kontinu dengan ukuran n berdistribusi normal dengan rata rata μ dan variansi σ^2 maka fungsi padat peluang gabungannya

$$\begin{aligned} f(X_1, X_2, ..., X_n) &= f(X_1) f(X_2) ... f(X_n) \\ &= \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \left(\frac{X_i - \mu}{\sigma} \right)^2 \right) \right) \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp\left(\sum_{i=1}^{n} \left(-\frac{1}{2} \left(\frac{X_i - \mu}{\sigma} \right)^2 \right) \right) \end{aligned} \tag{2.15}$$

Apabila parameter μ dan σ^2 diketahui, maka fungsi peluang bersyarat dari X adalah

$$f(X_1, X_2, ..., X_n \left| \mu, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp\left(\sum_{i=1}^{n} \left(-\frac{1}{2} \left(\frac{X_i - \mu}{\sigma} \right)^2 \right) \right)\right.$$

Apabila parameter μ yang tidak diketahui, maka fungsi peluang bersyarat dari X adalah

$$l(\mu) = f(\mu | X_1, X_2, ..., X_n) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp\left(\sum_{i=1}^{n} \left(-\frac{1}{2} \left(\frac{X_i - \mu}{\sigma} \right)^2 \right) \right)$$

Oleh karena bentuk fungsi tersebut eksponensial maka untuk memudahkan menyelesaikan solusi, fungsi likelihood tersebut diubah menjadi fungsi log-likelihood. Fungsi tersebut dapat di uraikan menjadi

$$\begin{aligned} L(\mu | X) &= \ln l(\mu) \\ &= \ln\left[\left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp\left(\sum_{i=1}^{n} \left(-\frac{1}{2} \left(\frac{X_i - \mu}{\sigma} \right)^2 \right) \right) \right] \\ &= \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n + \sum_{i=1}^{n} \left(-\frac{1}{2} \left(\frac{X_i - \mu}{\sigma} \right)^2 \right) \\ &= -\frac{n}{2} \ln\left(2\pi\sigma^2 \right) - \frac{1}{2} \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2 \end{aligned}$$

Kemudian fungsi tersebut dimaksimumkan dengan mencari nilai turunan parsial pertama fungsi tersebut terhadap μ dan disamadenganakan nol, sehingga menjadi

$$\begin{aligned} \frac{\partial L}{\partial \mu} &= 0 - \frac{1}{2} \left(\frac{1}{\sigma^2} \right) (2)(-1) \sum_{i=1}^{n} X_i - n\mu \\ 0 &= \frac{1}{\sigma^2} \left(\sum_{i=1}^{n} X_i - n\mu \right) \\ 0 &= \sum_{i=1}^{n} X_i - n\mu \\ n\mu &= \sum_{i=1}^{n} X_i \mu = \frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X} \end{aligned}$$

Sedangkan untuk memaksimumkan solusi tunggal dari memaksimumkan fungsi log-likelihood dapat dilakukan dengan cara melihat turunan kedua yang bernilai negative, yaitu

$$\frac{\partial^2 L}{\partial \mu^2} = -\frac{n}{\sigma^2}$$

2.7.2 Maximum Likelihood Multiparameter

Menurut Bain (1991) Apabila X merupakan variabel acak berukuran k dan diberikan fungsi Y berukuran n maka

$$\begin{aligned} Y_1 &= \beta_1 + \beta_2 X_{12} + \beta_3 X_{13} + \ldots + \beta_k X_{1k} + \varepsilon_1 \\ Y_2 &= \beta_1 + \beta_2 X_{22} + \beta_3 X_{23} + \ldots + \beta_k X_{2k} + \varepsilon_2 \\ &\vdots \\ Y_i &= \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \ldots + \beta_k X_{ik} + \varepsilon_i \end{aligned}$$

Persamaan diatas dapat disederhanakan menjadi matriks yaitu

$$\underset{(n\times 1)}{\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}} = \underset{(n\times k)}{\begin{bmatrix} 1 & X_{12} & \cdots & X_{1k} \\ 1 & X_{22} & \cdots & X_{2k} \\ 1 & \vdots & & \vdots \\ 1 & X_{n2} & \cdots & X_{nk} \end{bmatrix}} \underset{(\mathrm{k}\times 1)}{\begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}} + \underset{(n\times 1)}{\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}}$$

Sehingga dapat dinyatakan dalam model sederhana, yaitu

$$Y = X\beta + \varepsilon$$

Dengan

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

$$X = \begin{bmatrix} 1 & X_{12} & \cdots & X_{1k} \\ 1 & X_{22} & \cdots & X_{2k} \\ 1 & \vdots & & \vdots \\ 1 & X_{n2} & \cdots & X_{nk} \end{bmatrix}$$

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}$$

$$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Apabila $X_i = \begin{bmatrix} 1 & X_{i1} & X_{i2} & \cdots & X_{ik} \end{bmatrix}$ *dengan* $i = 1, 2, ..., n$ *maka* Y_i dapat dinyatakan dengan

$$Y_i = X_i \beta + \varepsilon_i$$

Apabila Y_i berdistribusi normal atau $Y_i \sim N\left(\mu, \sigma^2\right)$, maka diperoleh

$$\begin{aligned} \mu &= E(Y_i) \\ &= E\left(X_i \beta + \varepsilon_i\right) \\ &= E\left(X_i \beta\right) + E\left(\varepsilon_i\right) \\ &= E\left(X_i \beta\right) \\ &= X_i \beta \end{aligned}$$

Sehingga didapatkan fungsi likelihood sebagai berikut

$$\begin{aligned} l(\beta) &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(\sum_{i=1}^{n}\left(-\frac{1}{2}\left(\frac{X_i - \mu}{\sigma}\right)^2\right)\right) \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{1}{2\sigma^2}\left(Y - X\beta\right)^T \left(Y - X\beta\right)\right) \end{aligned} \quad (2.16)$$

Aziz (2010) menjelaskan dengan logaritma natural pada fungsi likelihood, sehingga didapatkan fungsi log-likelihood sebagai berikut

$$
\begin{aligned}
L(\beta) &= \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{1}{2\sigma^2}\left(Y - X\beta\right)^T\left(Y - X\beta\right)\right)\\
&= \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n + \left(-\frac{1}{2\sigma^2}\left(Y - X\beta\right)^T\left(Y - X\beta\right)\right)\\
&= \ln\left(2\pi\sigma^2\right)^{-\frac{n}{2}} + \left(-\frac{1}{2\sigma^2}\left(Y - X\beta\right)^T\left(Y - X\beta\right)\right)\\
&= -\frac{n}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\left(\left(Y^T - X^T\beta^T\right)\left(Y - X\beta\right)\right)\\
&= -\frac{n}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\left(Y^TY - Y^TX\beta - \beta^TX^TY + \beta^TX^TX\beta\right)
\end{aligned} \tag{2.17}
$$

Kemudian fungsi tersebut dimaksimumkan dengan melakukan turunan parsial pertama fungsi *log-likelihood* terhadap β dan menyamakan dengan nol.

$$
\begin{aligned}
\frac{\partial L}{\partial \beta} &= \frac{\partial}{\partial \beta}\left(-\frac{n}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\left(Y^TY - Y^TX\beta - \beta^TX^TY + \beta^TX^TX\beta\right)\right)\\
0 &= 0 + \frac{\partial}{\partial \beta}\left(-\frac{1}{2\sigma^2}\left(Y^TY - Y^TX\beta - \beta^TX^TY + \beta^TX^TX\beta\right)\right)\\
&= -\frac{1}{2\sigma^2}\left(\frac{\partial}{\partial \beta}(Y^TY) - \frac{\partial}{\partial \beta}(Y^TX\beta) - \frac{\partial}{\partial \beta}(\beta^TX^TY) + \frac{\partial}{\partial \beta}(\beta^TX^TX\beta)\right)\\
&= -\frac{1}{2\sigma^2}\left(0 - X^TY - X^TY + \left(\frac{\partial}{\partial \beta}(\beta^TX^TX\beta) + \frac{\partial}{\partial \beta^T}(\beta^TX^TX\beta)^T\right)\right)\\
&= -\frac{1}{2\sigma^2}\left(0 - X^TY - X^TY + X^TX\beta + X^TX\beta\right)\\
&= -\frac{1}{2\sigma^2}\left(-2X^TY + 2X^TX\beta\right)\\
&= -2X^TY + 2X^TX\beta\\
X^TX\beta &= X^TY\\
\beta &= \left(X^TX\right)^{-1}X^TY
\end{aligned}
$$

Sehingga penduga β dapat ditulis dengan

$$\beta = \left(X^TX\right)^{-1}X^TY$$

Solusi tunggal secara nyata dari pemaksimuman fungsi *log-likelihood* dapat di periksa dengan turunan kedua untuk maksimum lokal yaitu turunan kedua harus bernilai negatif.

$$\begin{aligned}\frac{\partial^2 L}{\partial \beta^2} &= \frac{\partial}{\partial \beta}\left(\frac{\partial L}{\partial \beta}\right)\\ &= \frac{\partial}{\partial \beta}\left(-\frac{1}{\sigma^2}\left(-X^T Y + X^T X \beta\right)\right)\\ &= -\frac{1}{\sigma^2}\left(0 + X^T X\right)\\ &= -\frac{1}{\sigma^2} X^T X\end{aligned}$$

2.8 Regresi Linear Berganda

Analisis yang memiliki variabel prediktor lebih dari satu disebut analisis regresi linear berganda yang digunakan untuk mengetahui ada tidaknya pengaruh signifikan dua atau lebih variabel prediktor $(X_1, X_2, ..., X_k)$ terhadap variabel respon (Y). Model regresi linear berganda untuk populasi ditunjukkan sebagai berikut:

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n + \varepsilon$$

Dengan :

$\hat{Y}$: nilai penduga variabel Y

β_0 : parameter konstanta

$\beta_1, \beta_2, ..., \beta_k$: parameter koefisien variabel prediktor (X)

Sebelum model regresi digunakan untuk menguji hipotesis, terlebih dahulu dilakukan:

1. Pengujian asumsi klasik
   - Uji multikolinearitas

     Uji multikolinearitas bertujuan untuk menguji apakah model yang dibangun terdapat korelasi antar variabel independen atau tidak. Model regresi yang baik seharusnya tidak terjadi korelasi antara variabel bebas. Jika antar variabel bebas saling berkorelasi maka variabel-variabel ini jadi tidak ortogonal. Variabel ortogonal adalah variabel independen yang nilai korelasi antar sesama variabel independen sama dengan nol. Adanya multikolinearitas dapat diketahui dengan menganlisis nilai *tolerance* dan *variance inflation factor* (VIF). Nilai *tolerance* yang rendah dan nilai VIF yang tinggi menunjukka adanya multikolinearitas. Nilai *tolerance* lebih besar dari 0,10 atau nilai VIF lebih kecil dari 10 menunjukkan tidak adanya multikolinearitas.

- Uji heteroskedastisitas

  Uji ini bertujuan untuk menguji apakah dalam model regresi terjadi ketidaksamaan varian dari residual satu pegamatan ke pengamatan yang lain. model regresi yang baik adalah yang homoskedastisitas atau tidak terjadi heterokedastisitas. Jika variabel independen secara statistik signifikan mempengaruhi variabel dependen, maka terdapat indikasi terjadi heteroskedastisitas pada model regresi.

- Uji normalitas

  Uji asumsi normalitas adalah untuk mengetahui apakah variabel-variabel dalam model regresi memiliki distribusi normal atau tidak. Model regresi yang baik adalah model regresi yang memiliki distribusi data yang normal atau mendekati normal. Normalitas data dapat dideteksi dengan melihat penyebaran data (titik) pada sumbu diagonal. Dari grafik atau dengan melihat hostogram dari residualnya. Dasar pengambilan keputusannya adalah:

  - Jika data menyebar disekitar garis diagonal dan mengikuti arah garis diagonal atau grafik histogram menunjukkan pola distribusi normal maka model regresi memenuhi asumsi normalitas.
  - Jika data menyebar jauh dari garis diagonal dan atau tidak mengikuti arah garis diagonal atau grafik histogramnya, berarti tidak menunjukkan pola distribusi normal maka model regresi tidak memenuhi asumsi normalitas.

2. Pengujian Hipotesis

- Uji *F*

  Uji simultan menunjukkan apakah semua variabel independen secara bersama-sama mempunyai pengaruh terhadap variabel dependen.

  Hipotesis

  $H_0 = 0$ Variabel independen bukan merupakan penjelas yang signifikan terhadap variabel dependen

  $H_1 \neq 0$ Variabel independen merupakan penjelas yang signifikan terhadap variabel dependen

  Uji Statistik :

  H_0 di tolak Jika nilai sig lebih kecil dari 5% atau nilai *F* hitung lebih lebih besar dari nilai *F* tabel.

- Uji *T*

  Uji *T* merupakan alat uji untuk mengetahui pengaruh masing-masing variabel independen terhadap variabel dependen.

  Uji hipotesis:

  H_0 : Variabel independen bukan merupakan variabel penjelas variabel dependen

  H_1 : Variabel independen merupakan variabel penjelas variabel dependen

  Uji statistik :

  H_0 ditolak jika nilai sig lebih kecil dari 5% atau nilai *t* hitung lebih besar dari nilai *t* tabel.

2.9 Regresi Poisson Sederhana

2.9.1 Pengertian

Analisis regresi Poisson sederhana adalah sebuah metode statistika yang digunakan untuk menganalisis hubungan antara sebuah variabel respon (Y) yang menyatakan data diskrit dan sebuah variabel prediktor (X). Variabel respon (Y) diberikan $X = x$ diasumsikan berdistribusi Poisson, sedangkan variabel prediktor X dapat berjenis diskrit, kontinu, atau berjenis kategorik (Hertriyanti, 2006).

2.9.2 Model Regresi Poisson Sederhana

Misalkan ingin diketahui hubungan antara sebuah variabel respon (Y) yang menyatakan data diskrit dan sebuah variabel prediktor yang berjenis kontinu atau kategorik. Variabel random respon Y diberikan $X = x$ diasumsikan berdistribusi Poisson. Jika diberikan sebuah sampel berisi n buah pasangan pengamatan yang saling bebas, yaitu $\{(X_i, Y_i); i = 1, 2, \ldots, n\}$ dengan X_i dan Y_i berturut-turut adalah pengamatan ke-i dari variabel X dan, Y maka hubungan antara Y dan X tidak dapat dijelaskan oleh model regresi linear sederhana.

$$f(\mathrm{x}; \mu) = \frac{e^{-\mu}\mu^{x}}{\mathrm{x}!}$$

$$E(Y) = \mu$$

$$\mu = \beta_0 + \beta_1 X_i;\ i = 1, 2, \ldots, n$$

Maka model regresi Poisson sederhana yaitu

$$\ln(\mu_i(x_i)) = \beta_0 + \beta_1 X_i;\ i = 1, 2, \ldots, n \qquad \text{(Fungsi log link)}$$

atau equivalen dengan

$$(\mu_i(x_i) = e^{\beta_0 + \beta_1 X_i};\ i = 1, 2, \ldots, n \qquad (2.11)$$

dengan β_0 , β_1 adalah parameter yang tidak diketahui.

2.10 Regresi Poisson Berganda

2.10.1 Pengertian

Regresi Poisson berganda adalah regresi yang menganalisis hubungan antara sebuah variabel respon (Y) yang merupakan data berjenis diskrit, diasumsikan berdistribusi Poisson dan p buah variabel prediktor (X) $x_1, x_2, \ldots, x_p$ yang berjenis diskrit, kontinu atau kategorik (Hertriyanti, 2006).

2.10.2 Model Regresi Poisson Berganda

Model regresi Poisson berganda merupakan perluasan dari model regresi Poisson sederhana, dimana regresi Poisson berganda akan diketahui hubungan antara sebuah variabel respon (Y) yang berjenis diskrit dan p buah variabel prediktor $x_1, x_2, ..., x_p$ yang berjenis diskrit, kontinu atau kategorik. Variabel random dari respon Y diberikan $X_1 = x_1, X_2 = x_2, ..., X_p = x_p$ yang diasumsikan berdistribusi Poisson.

Diberikan sebuah sampel yang berisi n buah pasangan pengamatan yang saling bebas $\{(X_{1i}, \mathrm{X}_{2i}, ..., \mathrm{X}_{pi}, \mathrm{Y}_i); i = 1, 2, ..., n\}$ dengan $X_{1i}, \mathrm{X}_{2i}, ..., \mathrm{X}_{pi}$ berturut-turut adalah pengamatan ke-i dari variabel $x_1, x_2, ..., x_p$ dan Y_i adalah pengamatan ke-i dari variabel Y . jika rata-rata bersyarat dari Y_i diberikan nilai dinyatakan oleh

$$E(Y_i | X_{1i} = x_{1i}, X_{2i} = x_{2i}, X_{pi} = x_{pi}) = \mu_i(x_{1i}, x_{2i}, ..., x_{pi}), \qquad i = 1, 2, ..., n$$

$$\mu_i(x_{1i}, x_{2i}, ..., x_{pi}) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_p X_{pi} \qquad i = 1, 2, ..., n$$

Maka model regresi Poisson bergandanya adalah

$$\ln(\mu_i(x_{1i}, x_{2i}, ..., x_{pi})) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_p X_{pi}$$

Dengan $\beta_0, \beta_1, \beta_2, ..., \beta_p$ menyatakan koefisien parameter-parameter yang tidak diketahui. Ekuivalen dengan:

$$\mu_i(x_{1i}, x_{2i}, ..., x_{pi}) = e^{(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_p X_{pi})}$$

$$\mu_i(x_{1i}, x_{2i}, ..., x_{pi}) = e^{(\beta_0 + \sum_{j=1}^{p} \beta_j X_{ji})} \tag{2.12}$$

2.11 Parameter Model Regresi Poisson

2.11.1 Penaksiran parameter Model Regresi Poisson

Salah satu metode penaksir parameter pada model regresi Poisson yang diketahui distribusinya adalah metode *Maximum Likelihood Estimation* (MLE). Penaksiran parameter $(\beta_j), j = 1, 2, 3, ..., p$ dapat dilakukan dengan langkah-langkah sebagai berikut:

1. Mengambil *n* buah sampel acak
2. Membentuk fungsi likelihood

$$L\left(y_i,\beta\right)=\prod_{i=1}^{n} f\left(y_i,\beta\right)$$
$$=\prod_{i=1}^{n}\frac{\mu_i^{y_i}e^{-\mu_i}}{y_i!}$$
$$=\frac{\left\{\prod_{i=1}^{n}\mu_i^{y_i}\right\}\prod_{i=1}^{n}e^{-\mu_i}}{\prod_{i=1}^{n}y_i!}$$
$$=\frac{\left\{\prod_{i=1}^{n}\mu_i^{y_i}\right\}e^{-\sum_{i=1}^{n}\mu_i}}{\prod_{i=1}^{n}y_i!}$$

3. Mengambil bentuk log dari fungsi *likelihood* yang diperoleh. Fungsi log-likelihood yang terbentuk ditunjukkan oleh persamaan

$$\ln L\left(y_i,\beta\right)=\ln\left\{\frac{\left\{\prod_{i=1}^{n}\mu_i^{y_i}\right\}e^{-\sum_{i=1}^{n}\mu_i}}{\prod_{i=1}^{n}y_i!}\right\}$$
$$=\ln\left\{\prod_{i=1}^{n}\left(\frac{\mu_i^{y_i}e^{-\mu_i}}{y_i!}\right)\right\}$$
$$=\ln\left\{\prod_{i=1}^{n}\left\{\frac{\left[e^{x_i^T\beta}\right]^{y_i}e^{-\left[e^{x_i^T\beta}\right]}}{y_i!}\right\}\right\}$$

$$= \ln \left\{ \prod_{i=1}^{n} \left\{ \frac{\left[e^{\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}} \right]^{y_i} e^{-\left[e^{\beta_0 + \Sigma_{j=1}^{p} \beta_j x_{ij}} \right]}}{y_i\,!} \right\} \right\}$$

$$= \sum_{i=1}^{n} \ln \left\{ \frac{\left[e^{\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}} \right]^{y_i} e^{-\left[e^{\beta_0 + \Sigma_{j=1}^{p} \beta_j x_{ij}} \right]}}{y_i\,!} \right\}$$

$$= \sum_{i=1}^{n} \left\{ \ln \left[e^{(\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij})^{y_i}} \right] + \ln \left(e^{-\left[e^{\beta_0 + \Sigma_{j=1}^{p} \beta_j x_{ij}} \right]} \right) - \ln(y_i\,!) \right\}$$

$$= \sum_{i=1}^{n} \left\{ y_i \ln \left(e^{(\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij})} \right) - \left[e^{\beta_0 + \Sigma_{j=1}^{p} \beta_j x_{ij}} \right] - \ln(\mathrm{y}_i!) \right\}$$

$$= \sum_{i=1}^{n} \left\{ y_i (\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}) - e^{(\beta_0 + \Sigma_{j=1}^{p} \beta_j x_{ij})} - \ln(\mathrm{y}_i!) \right\}$$

4. Mendiferensialkan fungsi *log-likelihood* terhadap masing-masing parameter β dan disamakan dengan nol sehingga diperoleh taksiran parameter yang maksimum.

$$R(\beta)=\begin{bmatrix} R_0(\beta) \\ R_1(\beta) \\ \vdots \\ R_p(\beta) \end{bmatrix} = \begin{bmatrix} \dfrac{\partial \ln L(y_i,\beta)}{\partial \beta_0} \\ \dfrac{\partial \ln L(y_i,\beta)}{\partial \beta_1} \\ \vdots \\ \dfrac{\partial \ln L(y_i,\beta)}{\partial \beta_p} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{n}\left\{ y_i - \exp\left(\beta_0 + \sum_{j=1}^{p}\beta_j x_{ij}\right)\right\} \\ \sum_{i=1}^{n}\left\{ x_{i1}\left(y_i - \exp\left(\beta_0 + \sum_{j=1}^{p}\beta_j x_{ij}\right)\right)\right\} \\ \vdots \\ \sum_{i=1}^{n}\left\{ x_{ip}\left(y_i - \exp\left(\beta_0 + \sum_{j=1}^{p}\beta_j x_{ij}\right)\right)\right\} \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$= 0$$

Penaksir titik β secara manual akan mengalami kesulitan, sehingga dalam penentuannya akan dilakukan dengan menggunakan bantuan software statistik yaitu R versi 4.1.1 untuk membantu proses perhitungan taksiran nilai β .

2.11.2 Pengujian Parameter Model Regresi Poisson

Sebelum menentukan statistik uji perlu dilakukan pengujian parameter pada model regresi poisson untuk mengetahui ada atau tidaknya pengaruh variabel prediktor terhadap variabel respon. Pengujian parameter dilakukan dengan menggunakan Uji Serentak dan Uji Parsial.

a. Uji Serentak Parameter Model Regresi Poisson

   Uji Serentak Parameter dilakukan dengan menggunakan metode *Maximum Likelihood Ratio Test* (MLRT). Hal pertama yang dilakukan adalah menentukan dua fungsi likelihood yang berhubungan dengan model regresi yang diperoleh, yaitu $L(\Omega)$ dan $L(\omega)$.

$L(\Omega)$ adalah nilai maksimum untuk model lengkap dengan melibatkan variabel prediktor dan $L(\omega)$ adalah nilai maksimum likelihood untuk model sederhana tanpa melibatkan variabel prediktor. Setelah memaksimumkan kedua fungsi likelihood maka akan didapatkan nilai likelihood-nya.

Hipotesis :

$H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$

$H_1 : paling\ sedikit\ terdapat\ satu\ \beta_j \neq 0 ;\ j = 1, 2, 3, ..., p$

Taraf signifikan :

$a = 0,05$

Statistik Uji :

$$\begin{aligned} G &= -2 \ln \left(\frac{L(\omega)}{L(\Omega)} \right) \\ &= 2 \left[\ln L(\Omega) - \ln L(\omega) \right] \end{aligned} \tag{2.18}$$

G merupakan deviansi model regresi Poisson atau devians yang dihitung pada seluruh parameter dalam model. Nilai G yang semakin kecil menunjukkan semakin kecil pula tingkat kesalahan yang dihasilkan model sehingga model menjadi semakin tepat.

Statistik uji rasio Likelihood mengikuti sebaran χ^2 (Chi-Square) dengan derajat bebasnya adalah banyaknya parameter dibawah populasi dikurangi dengan banyak parameter di bawah *H*$_0$.

Kriteria Uji :

H$_0$ ditolak jika $G > \chi^2_{(k;a)}$

b. Uji Parsial Parameter Model Regresi Poisson

Uji Parsial digunakan untuk mengetahui parameter mana yang memberikan pengaruh signifikan terhadap model. Statistik uji yang digunakan adalah uji *Wald.*

Hipotesis :

$H_0 : \beta_j = 0$ *untuk suatu* $j = 1, 2, ..., p$

$H_1 : \beta_j \neq 0$ *untuk suatu* $j = 1, 2, ..., p$

Taraf signifikan :

$a = 0,05$

Statistik Uji :

$$W = \left(\frac{\beta_j}{SE\left(\beta_j\right)} \right)^2 \tag{2.19}$$

Dengan

β = taksiran parameter β_j

$SE\left(\beta_j\right)$ = taksiran standar error dari β_j

Kriteria Uji :

H_0 ditolak jika $W > \chi^2_{(a;1)}$

2.12 *Generalized Poisson Regression*

Model Regresi Poisson yang mengalami *overdispersi* atau *underdispersi* diperlukan pendekatan dengan model regresi yang lebih sesuai, salah satu model regresi yang dapat digunakan adalah model *Generalized Poisson Regression* (GPR).

Fungsi Peluang *Generalized Poisson Regression* :

$$f\left(y_i, \mu_i, k\right) = \left(\frac{\mu_i}{1 + k\mu_i} \right)^{y_i} \frac{\left(1 + ky_i\right)^{y_i - 1}}{y_i\,!} \exp\left(-\frac{\mu_i\left(1 + ky_i\right)}{\left(1 + k\mu_i\right)} \right) \tag{2.21}$$

Dengan rata-rata dan variansinya masing-masing μ_i dan $\mu_i\left(1 + ky_i\right)^2$.

2.13 Parameter Model *Generalized Poisson Regression*

2.13.1 Penaksiran Parameter Model *Generalized Poisson Regression*

Jika diketahui diketahui $\mu_i = \exp\left(\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}\right) = \exp\left(x_i^T \beta\right)$. Maka fungsi peluang model *Generalized Poisson Regression* dapat dinyatakan

$$f\left(y_i, \mu_i, k\right) = \left(\frac{\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}}{1 + k\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}}\right)^{y_i} \frac{\left(1 + ky_i\right)^{y_i - 1}}{y_i\,!} \exp\left(-\frac{\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}\left(1 + ky_i\right)}{\left(1 + k\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}\right)}\right)$$

Parameter β dan *k* dapat ditaksir dengan metode *Maximum Likelihood Estimation* (MLE). Berikut langkah-langkahnya:

a. Membentuk fungsi Likelihood

Fungsi likelihood model *Generalized Poisson Regression* sebagai berikut:

$$L\left(\beta, k, y\right) = \prod_{i=1}^{n} f\left(y_i, \beta, k\right)$$

b. Membentuk fungsi log-likelihood dari fungsi likelihood di atas

$$\ln L\left(\beta, k\right) = \ln \prod_{i=1}^{n} f\left(y_i, \beta, k\right)$$

c. Mendiferensialkan persamaan log-likelihood yang diperoleh terhadap β_0, $\beta_j;\ j = 1, 2, ..., p$, dan terhadap k kemudian disamakan dengan nol untuk menaksir parameter model *Generalized Poisson Regression*. Setelah penaksiran parameter selesai maka diperoleh taksiran model *Generalized Poisson Regression* seperti pada persamaan :

$$y_i = \exp\left(x_i^T \beta\right) + \varepsilon_i$$

2.13.2 Pengujian Parameter Model *Generalized Poisson Regression*

Penaksiran parameter menggunakan metode Maximum Likelihood Ratio Test (MLRT) dan menggunakan statistik uji *Wald* untuk pengujian parameter secara parsial.

a. Pengujian menggunakan uji *Wald*

Hipotesis :

$H_0 : \beta_j = 0$

$H_1 : \beta_j \neq 0$

Taraf signifikan :

$a = 0,05$

Statistik uji :

$$W = \left(\frac{\beta_j}{SE\left(\beta_j\right)} \right)^2$$

Kriteria Uji:

Tolak H_0 jika nilai $W > \chi^2_{(a,1)}$ atau tolak H_0 jika nilai p-value > a

b. Maximum Likelihood Ratio Test (MLRT)

Merupakan salah satu metode yang digunakan untuk menguji signifikansi parameter disperse. Pengujian signifikansi parameter dispersi dilakukan untuk menentukan apakah kondisi *equidispersi* terpenuhi atau tidak. Kondisi *equidispersi* dikatakan terpenuhi jika nilai parameter dispersi bernilai 0, dalam hal ini $k = 0$.

Hipotesis :

$H_0 : k = 0$

$H_1 : k \neq 0$

Taraf sigifikansi :

$a = 0,05$

Statistik uji :

$$T = -2\left(\ell_p - \ell_{GP}\right) = 2\left(\ell_{GP} - \ell_p\right) \quad (2.22)$$

ℓ_{GP} adalah nilai log-likelihood pada model *Generalized Poisson Regression* dan ℓ_p adalah nilai log-likelihood pada model regresi Poisson.

Pada kondisi H_0, statistic uji T mendekati distribusi Chi-Square (χ^2) dengan derajat bebas 1.

Kriteria Uji :

Tolak H_0 jika $T > \chi^2_{(a,1)}$. Penolakan H_0 menunjukkan bahwa *Generalized Poisson Regression* lebih cepat digunakan dibandingkan model regresi Poisson.

2.14 Parameter Dispersi

Menurut Darnah (2011) parameter dispersi (ϕ) diperoleh dari

$$\phi = \frac{Deviance}{df}$$

Dengan

$$Deviance = D = 2\sum_{i=1}^{n}\left\{ y_i \log \frac{y_i}{\mu_i} \right\}$$

$$df = derajat\, bebas$$

Apabila nilai $\phi > 0$ maka terjadi overdispersi dan apabila $\phi < 0$ maka terjadi underdispersi.

2.15 Overdispersi

Terdapat beberapa asumsi yang harus dipenuhi dalam model regresi poisson. Salah satu asumsi adalah kesamaan antara rata-rata (*mean)* dan variansi (*variance*) yang disebut ekuidispersi (Darnah, 2011). Namun sering dijumpai data yang variansinya lebih kecil atau lebih besar dari rataannya. Keadaan ini disebut dengan underdispersi (*underdispersion*) atau overdispersi (*overdisperslion*). Beberapa hal yang menyebabkan overdispersi, diantaranya:

1) Terdapat korelasi antar pengamatan
2) Terdapat pelanggaran asumsi distribusi poisson, yaitu

$$Var(Y) > E(Y)$$

3) Terdapat excess zero
4) Terdapat outlier dalam data

Hubungan nilai varians dan nilai rata-rata (*mean*) dalam *Generalized Poisson Regression* dapat dikondisikan sebagai berikut (Putra, 2013):

1) Jika nilai varians sama dengan nilai rata-rata $Var\left(y_i \middle| x_i\right) = E\left(y_i \middle| x_i\right)$, maka nilai parameter dispersi $k = 0$, sehingga fungsi densitas peluang *Generalized Poisson*, akan diturunkan ke regresi Poisson
2) Jika nilai varians lebih besar dari nilai rata-rata $Var\left(y_i \middle| x_i\right) > E\left(y_i \middle| x_i\right)$, maka nilai parameter dispersi $k > 0$, sehingga dapat dikatakan pada data terjadi *overdispersi*
3) Jika nilai varians lebih kecil dari nilai rata-rata $Var\left(y_i \middle| x_i\right) < E\left(y_i \middle| x_i\right)$, maka nilai parameter dispersi $k < 0$, sehingga dapat dikatakan pada data terjadi *underdispersi*.

Overdispersi atau underdispersi dapat menyebabkan taksiran parameter yang diperoleh tidak efisien mengakibatkan model yang dihasilkan menjadi kurang tepat dan berakibat fatal pada intepretasi model, terutama pada estimasi parameter model karena dapat menaksir standart error yang terlalu rendah dan dapat memberikan kesimpulan yang keliru tentang signifikan atau tidaknya parameter regresi yang terlibat (Darnah, 2011).

Rumus untuk mendeteksi overdispersi yaitu:

- Nilai Deviance

$$\theta_1 = \frac{D}{db} > 1; D = 2\sum_{i=1}^{n} \left\{ y_i \log \frac{y_i}{\mu_i} \right\}$$

- Nilai Pearson Chi-Square

$$\theta_2 = \frac{x^2}{db} > 1; x^2 = \sum_{i=1}^{n} \frac{(y_i - \mu_i)^2}{\sigma_i}$$

Jika hasil bagi antara nilai statistik D terhadap derajat bebasnya atau statistik χ^2 terhadap derajat bebasnya lebih dari 1 maka data mengalami overdispersi, sedangkan jika nilai hasil bagi lebih kecil atau kurang dari 1, maka dinyatakan kasus data sebagai kasus underdispersi.

Uji statistik yang bisa juga digunakan untuk mendeteksi overdispersi pada suatu data adalah uji overdispersi yang dapat menggunakan package AER dari software R studio (Herindrawati, Latra, Purhadi, 2017).

Keputusan hipotesis:

H_0 : Tidak terjadi overdispersi

H_1 : Terjadi overdispersi

Keputusan yang diambil menggunakan software R yaitu jika nilai p-value < α maka H_0 ditolak yang berarti terjadi overdispersi. Tetapi jika p-value > α maka H_0 diterima yang berarti tidak terjadi overdispersi.

2.16 Multikolinearitas

Multikolinearitas artinya terdapat hubungan antara variabel prediktor yang satu dengan variabel prediktor yang lain (Aulele, 2012). Syarat yang harus dipenuhi dalam model regresi adalah tidak terjadi kasus multikolinearitas atau tidak adanya korelasi antara satu variabel prediktor dengan variabel prediktor lainnya. Apabila kasus ini terjadi maka akan mengakibatkan nilai standart error yang sangat besar.

Untuk mendeteksi ada tidaknya multikolinearitas dapat dilakukan dengan melihat nilai Tolerence dan VIF (Variance Inflation Factor). Jika nilai *Tolerance* lebih dari 0,1 dan VIF kurang dari 10 maka tidak terjadi multikolinearitas.

Hipotesis:

H_0 : Model regresi memiliki masalah multikolinearitas

H_1 : Model regresi tidak memiliki masalah multikolinearitas

Taraf signifikansi:

$\alpha = 0,05$

Statistik Uji:

$$VIF = \frac{1}{(1 - r_{i,j}{}^2)} \tag{2.20}$$

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

Kriteria Uji:

Tolak H_0 jika seluruh variabel predictor memiliki nilai VIF kurang dari 10 dan nilai *Tolerance* lebih dari 0,1. Sebaliknya, jika seluruh variabel prediktor memiliki nilai VIF lebih dari 10 dan nilai *Tolerance* kurang dari 0,1 maka terima H_0.

2.17 Uji Kesesuaian Model (*Goodness of fit*)

Pengujian Kesesuaian Model untuk mengetahui model yang digunakan sesuai atau tidak dengan data yang diamati. Pengujian kesesuaian model dilakukan dengan menggunakan uji *goodness of fit* yang bertujuan untuk mengambil kesimpulan mengenai sebaran populasi. Suatu contoh acak dipilih dari populasi yang bersangkutan, kemudian informasi contoh tersebut digunakan untuk menguji kebenaran sebaran populasi tersebut. Uji ini didasarkan pada seberapa baik kesesuaian atau kecocokan (*goodness of fit*) antara frekuensi pengamatan data yang diperoleh. Data sampel dengan frekuensi harapan dari distribusi yang dihipotesiskan. Dalam hal ini akan ditunjukkan apakah data sudah sesuai atau tidak dengan model yang didapatkan dengan menggunakan software R Studio.

Pengujian model dengan melihat nilai uji *goodness of fit* menggunakan hipotesis sebagai berikut:

H_0 : Model yang didapat sesuai dengan data yang diamati

H_1 : Model yang didapat tidak sesuai dengan data yang diamati

Dengan taraf signifikan $\alpha = 0{,}05$

Kriteria uji:

H_0 ditolak jika nilai Pearson Chi Square lebih kecil dari $\alpha = 0{,}05$

2.18 AIC (Akaike Information Criterion)

Terdapat beberapa metode dalam menentukan model terbaik, salah satunya adalah Akaike Information Criterion (AIC). AIC (Akaike Information

Criterion) atau kriteria informasi adalah kriteria untuk memilih model dalam ekonometrika (Ruliana, 2015). AIC digunakan untuk memperoleh model terbaik dari beberapa metode dengan asumsi Poisson. AIC didefinisikan oleh persamaan sebagai berikut:

$$AIC = -2\ln \mathrm{L}(\hat{\theta}) + 2p \qquad (2.23)$$

Dengan

$\mathrm{L}(\hat{\theta})$ adalah nilai likelihood dan p adalah jumlah parameter

Model terbaik adalah model yang memiliki nilai AIC terkecil

2.19 Mortalitas

Kematian atau mortalitas merupakan salah satu komponen proses demografi yang berpengaruh terhadap struktur penduduk, dua komponen yang lainnya adalah kelahiran (fertilitas) dan mobilitas penduduk (Mantra, 2000). Menurut PBB dan WHO Kematian adalah peristiwa hilangnya semua tanda-tanda kehidupan secara permanen yang biasa terjadi setiap saat setelah kelahiran hidup. Keguguran dan lahir mati (Still birth) tidak termasukdalam pengertian kematian (Mantra, 2000).

Terdapat 2 faktor yang mempengaruhi kematian yaitu faktor langsung (internal) dan tidak langsung (eksternal). Faktor langsung dapat dipengaruhi dengan beberapa variabel seperti *umur*, manusia memiliki kapasitas atau batas untuk hidup di dunia. Semakin tua seseorang maka kemampuannya semakin berkurang dan melemah dan berakhir pada kematian. Kemudian faktor *penyakit* WHO pada tahun 2014 menunjukkan bahwa penyakit *kardiovaskular* (Penyakit jantung dan pembuluh darah) merupakan penyebab kematian tertinggi di Asia tenggara termasuk Indonesia sebesar 37%. Faktor lainnya yaitu *kecelakaan, kekerasan, dan bunuh diri.*

Faktor tidak langsung dipengaruhi oleh beberapa variabel seperti *tekanan* (baik psikis maupun fisik), terdapat kasus bullying di Indonesia yang biasa terjadi dikalangan anak dan remaja, dimulai dari hal-hal sepele seperti mengejek, menghina, mengambil uang jajan, mengancam, menendang dan lain sebagainya semakin lama akan membuat seseorang merasa tertekan dan keadaan terburuknya akan depresi lalu bunuh diri. *Kedudukan sosial ekonomi* yang belum dapat memenuhi kebutuhan pokok maka akan menimbulkan beberapa masalah dan berujung kematian. Faktor dalam pernikahan seperti *kekerasan dalam rumah tangga* (KDRT) dapat berujung pada kematian, *tingkat pendidikan* seperti kesadaran terhadap menjaga kesehatan yang rendah dapat berperan dalam meningkatkan tingkat kematian.

2.20 Covid-19

Berawal dari salah satu kota di China yaitu Wuhan yang dilaporkan terdapat 27 orang menderita penyakit mirip pneumonia, demam, kesulitan bernafas, dan paru-paru yang tidak normal. (Bramasta, Dandy Bayu; 2020). 5 Januari 2020 China melaporkan kasus ini kepada WHO (World Health Organization) bahwa telah terdapat 41 orang dan satu orang diantaranya meninggal dunia. WHO menyatakan bahwa virus misterius ini adalah virus baru yang bernama Corona virus

Corona virus merupakan virus jenis baru yang kini telah menggemparkan masyarakat dunia. Virus ini berhasil menginfeksi ribuan juta masyarakat global dalam waktu yang sangat singkat. Bahkan manusia tanpa menunjukkan gejala terinfeksi Covid-19 dapat pula menyebarkan kepada manusia lainnya. Mengantisipasi peningkatan penyebaran dan jumlah infeksi masyarakat dihimbau untuk melakukan protokol kesehatan dan pola hidup sehat. Yaitu dengan menerapkan Physical distancing (menjaga jarak interaksi langsung) dan Work From Home (WFH) juga protokol kesehatan diantaranya menjaga kebersihan, menggunakan masker pelindung wajah, mencuci tangan dengan sabun, menggunakan handsanitizier, menutup mulut ketika bersin/batuk.

Terdapat kurang lebih 200 negara yang telah terkena virus ini dari yang terbanyak yaitu Amerika Serikat dengan kurang lebih 186.046 positif corona, lalu Italia dengan 105.792 kasus, dan Spanyol sebanyak 95.923 kasus, dan negara lainnya. Di Indonesia sendiri pada 10 April 2020 terdapat 3.512 kasus positif, sembuh 282 orang dan meninggal sebanyak 306 orang dengan *fatality rate* atau tingkat kematian sebesar 9,1% (Kementrian Kesehatan Republik Indonesia; 2020).

Tingginya tingkat kematian akibat virus corona ini disebabkan oleh dua faktor yaitu faktor dari dalam individu seperti penyakit bawaan yang dialami, kurangnya awareness masing-masing individu terhadap virus ini dan faktor eksternal seperti fasilitas rumah sakit yang kurang memadai, peraturan pemerintah yang belum efektif, dan sebagainya.

BAB III
METODOLOGI PENELITIAN

3.1 Jenis dan Sumber Data

Penelitian ini menggunakan tipe data kuantitatif yaitu data yang dinyatakan dalam bentuk angka cacahan (diskrit). Penelitian ini merupakan penerapan model regresi nonlinear, yaitu model *Generalized Poisson Regression* pada kasus jumlah kematian akibat Covid-19 di Indonesia.

Data yang di kumpulkan dalam penelitian ini adalah jenis data sekunder. Data sekunder yaitu data yang diperoleh dan dikumpulkan oleh Lembaga pengumpul data serta di publikasikan kepada pengguna data. Data yang diperoleh dari situs website *http://Covid19.go.id/* yaitu jumlah kematian akibat Covid-19 di Indonesia per Maret 2021.

3.2 Metode Pengumpulan Data

Metode pengumpulan data yang digunakan dalam penelitian ini adalah *non-participant observer*, dimana peneliti hanya mengamati data yang sudah tersedia tanpa perlu melakukan survei di lapangan.

Data jumlah kematian akibat Covid-19 di Indonesia bersumber dari *http://Covid19.go.id/*. Data yang di ambil adalah data per provinsi di Indonesia.

3.3 Populasi, Sampel, dan Variabel Penelitian

1. Populasi

   Populasi objek penelitian dari data sekunder yang diperoleh yaitu Masyarakat yang terjangkit Covid-19 di Indonesia.

2. Sampel

   Sampelnya adalah jumlah kematian akibat Covid-19 di Indonesia

3. Variabel penelitian

   a. Variabel Respon

      Pada penelitian ini variabel respon yang digunakan adalah jumlah kematian akibat Covid-19 di Indonesia (Y)

   b. Variabel Prediktor

1) Persentase kematian berjenis kelamin laki laki (X_1)

2) Persentase kematian berjenis kelamin perempuan (X_2)

3) Persentase umur 35 – 45 tahun (X_3)

4) Persentase umur 46 – 59 tahun (X_4)

5) Persentase umur lebih dari 60 tahun (X_5)

6) Persentase riwayat penyakit hipertensi (X_6)

7) Persentase riwayat penyakit diabetes melitus (X_7)

8) Persentase riwayat penyakit jantung (X_8)

9) Persentase riwayat penyakit ginjal (X_9)

10) Persentase riwayat penyakit paru-paru (X_{10})

11) Jumlah rumah sakit rujukan (X_{11})

12) Jumlah tenaga kesehatan (X_{12})

3.4 Analisis Data

Berikut ini merupakan prosedur analisis data untuk mencapai tujuan penelitian:

1. Mengumpulkan data dan asumsikan data berdistribusi poisson
2. Mendeteksi adanya multikolinearitas antara variabel prediktor

   Ada tidaknya multikolinearitas dapat diketahui dengan melihat nilai VIF (Variance Inflation Factor) dan *Tolerance*.
3. Menaksir parameter model regresi Poisson dengan menggunakan model *Maximum Likelihood Estimation.* Yang dapat dilakukan dengan langkah-langkah:
   a. Mengambil n buah sampel
   b. Membentuk fungsi likelihood sehingga diperoleh fungsi *likelihood* untuk model regresi Poisson.
   c. Mengambil bentuk log dari fungsi *likelihood* yang diperoleh bentuk dari fungsi log-likelihoodnya.
   d. Mendiferensialkan fungsi log-likelihoodnya terhadap masing-masing parameter dan disamakan dengan nol. Sehingga diperoleh taksiran parameter.
4. Melakukan pengujian parameter model regresi Poisson

Uji parameter pada model regresi Poisson dilakukan menggunakan Uji Serentak dan Uji Parsial.

5. Mendapatkan model regresi Poisson
6. Uji asumsi *equidispersion*

   Uji asumsi *equidispersion* dapat dilakukan dengan melihat nilai *mean* dan nilai varians dari variabel respon Y. dikatakan memenuhi asumsi jika nilai *mean* dan nilai variansnya sama. Jika nilai varians lebih besar dari nilai *mean* maka data mengalami overdispersi, dan jika nilai varians lebih kecil dari nilai *mean* maka data mengalami underdipersi, sehingga pemodelan sebaiknya dilakukan menggunakan metode *Generalized Poisson Regression*.

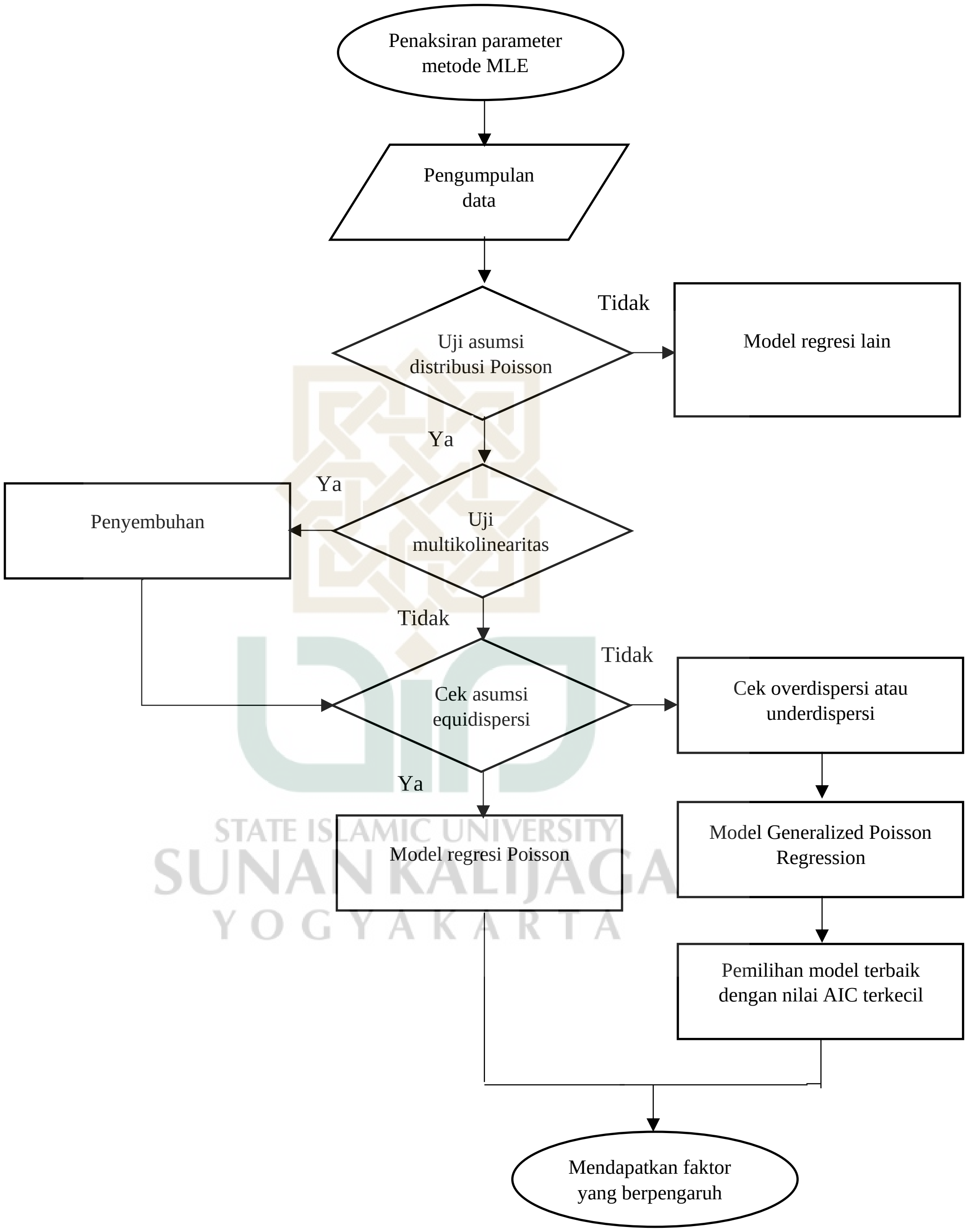
   Jika uji asumsi equidispersi terpenuhi maka langkah kerja akan dilanjutkan pada langkah kerja ke-11. Jika asumsi equidispersi tidak terpenuhui maka langkah kerja akan dilanjutkan pada langkah ke-7.
7. Menaksir Parameter model GPR dengan metode *Maximum Likelihood Estimation*

   Penaksiran parameter model GPR dilakukan dengan metode *Maximum Likelihood Estimation* (MLE). Pada umumnya penaksiran parameter model *Generalized Poisson Regression* memiliki langkah yang sama dengan penaksiran parameter pada model regresi Poisson.
8. Uji signifikansi model GPR

   Pengujian parameter model GPR dilakukan sama dengan pengujian parameter regresi Poisson, yaitu dengan menggunakan metode *Maximum Likelihood Ratio Test* (MLRT) dan menggunakan statistik uji Wald untuk pengujian parameter secara parsial.
9. Mendapatkan model GPR
10. Memilih model terbaik

    Pemilihan model terbaik dilakukan dengan melihat nilai AIC. Model yang memiliki AIC terkecil merupakan model terbaik.
11. Mendapatkan faktor yang mempengaruhi jumlah kematian akobat Covid-19 di Indonesia.

3.5 Bagan Prosedur Kerja



Gambar 3.1 Bagan

BAB IV
PEMBAHASAN

4.1 Regresi Poisson

Regresi Poisson memiliki fungsi penghubung g yang linier yang menghubungkan *mean* dari variabel respon dengan variabel prediktor yaitu:

$$g(\mu_i) = \eta_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_p x_{pi} \tag{4.1}$$

Sehingga hubungan antara *mean* dan prediktor adalah sebagai berikut:

$$\mu_i = g^{-1}(\eta_i) = g^{-1}(x_i^T \beta) \tag{4.2}$$

Fungsi penghubung pada regresi Poisson yaitu fungsi penghubung log, karena *mean* pada variabel respon berbentuk fungsi eksponensial dan menjamin bahwa nilai variabel yang ditaksir dari variabel respon akan bernilai non-negatif.
Bentuk dari fungsi penghubung log yaitu:

$$g(\mu_i) = \log(\mu_i) = x_i^T \beta \tag{4.3}$$

Sehingga dapat ditulis:

$$\log(\mu_i) = x_i^T \beta$$

Kemudian kedua ruas diambil fungsi eksponensialnya, sehingga diperoleh:

$$\begin{aligned} \log(\mu_i) &= x_i^T \beta \\ e^{\log(\mu_i)} &= e^{x_i^T \beta} \\ \mu_i &= e^{x_i^T \beta} \end{aligned}$$

Sehingga diperoleh fungsi model regresi Poisson sebagai berikut

$$\mu_i = \exp(x_i^T \beta) = \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_p x_{pi}) \tag{4.4}$$

Fungsi peluang distribusi Poisson adalah

$$P(y_i;\beta) = \frac{\left[\mu(x_i;\beta)\right]^{y_i} e^{-\left[\mu(x_i;\beta)\right]}}{y_i!} \tag{4.5}$$

μ	: *mean* Poisson
Vektor β	: parameter model

Maka diperoleh

Mean : $\mu_i = \exp\left(x_i;\beta\right) = \exp(x_i^T\beta) = \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_p x_{pi})$

Var : $\mathrm{var}(\mathrm{y}_i) = \exp\left(x_i;\beta\right) = \exp(x_i^T\beta) = \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_p x_{pi})$

Sehingga diperoleh model regresi Poisson dengan penghubung log yaitu:

$$\begin{aligned} y_i &= \mu_i + \varepsilon_i \\ &= \exp\left(x_i;\beta\right) + \varepsilon_i \\ &= \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_p x_{pi}) + \varepsilon_i \end{aligned} \quad 4.6$$

Analisis regresi merupakan metode statistika yang paling sering digunakan dalam segala bidang ilmu pengetahuan. Analisis regresi bertujuan untuk memodelkan hubungan antara dua variabel, yang terdiri dari variabel prediktor dan variabel respon. Apabila terdapat variabel respon berdistribusi Poisson dan terdapat satu atau lebih variabel prediktor maka model regresi yang menggambarkan hubungan kedua variabel adalah regresi Poisson. Jika Y adalah data diskrit yang berdistribusi Poisson dengan parameter μ maka fungsi probabilitasnya adalah

$$f\,(\mathrm{x};\mu) = \frac{e^{-\mu}\mu^x}{\mathrm{x}!} \quad 4.7$$

Dimana:

$x = 0,1,2,...$

$\mu > 0$, menyatakan banyaknya sukses yang terjadi dalam selang waktu atau daerah tertentu tersebut

$e = 2,7183$

Dengan $E(Y) = Var(Y) = \mu$

$$Y_i \sim poisson(\mu_i)\ i = 1,2,...,n$$

Dimana model regresi Poisson

$$\mu_i = e^{x_i^{\mathrm{T}}\beta} \quad 4.8$$

μ_i adalah rata-rata jumlah kejadian yang terjadi dalam interval waktu tertentu x adalah variabel prediktor yang dinotasikan:

$$\mathrm{x}_i = \begin{bmatrix} 1 & \mathrm{x}_{1i} & \ldots & \mathrm{x}_{ki} \end{bmatrix}^T$$

β adalah parameter regresi Poisson yang di notasikan:

$$\beta = \begin{bmatrix} \beta_0 & \beta_1 & \ldots & \beta_k \end{bmatrix}^T$$

4.1.1 Penduga Parameter Regresi Poisson

Pendugaan parameter regresi Poisson dilakukan dengan menggunakan metode *Maximum Likelihood Estimation* (MLE) sebagai berikut:

$$\begin{aligned}
\ln L(y_i,\beta) &= \ln\left(\prod_{i=1}^{n} \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!}\right) \\
\ln L(y_i,\beta) &= \sum_{i=1}^{n} \ln\left(\frac{\exp(-\mu_i)\,\mu_i^{y_i}}{y_i!}\right) \\
\ln L(y_i,\beta) &= \sum_{i=1}^{n}\left(\ln\left(e^{-\mu_i}\right) + \ln\left(\mu_i^{y_i}\right) - \ln\left(y_i!\right)\right) \\
\ln L(y_i,\beta) &= \left(-\sum_{i=1}^{n} e^{\mathrm{x}_i^T\beta} + \sum_{i=1}^{n} y_i \mathrm{x}_i^T\beta - \sum_{i=1}^{n}\ln\left(y_i!\right)\right)
\end{aligned} \qquad 4.9$$

Menggunakan metode MLE maka penduga parameter regresi Poisson yang dilambangkan β didapatkan dari turunan pertama fungsi ln likelihood.

Turunan pertama fungsi likelihood terhadap β^T :

$$\frac{\partial \ln L(\beta)}{\partial \beta^T} = -\sum_{i=1}^{n} \mathrm{x}_i e^{\mathrm{x}_i^T\beta} + \sum_{i=1}^{n} \mathrm{x}_i y_i$$

Turunan kedua fungsi ln likelihood terhadap β :

$$\frac{\partial^2 \ln L(\beta)}{\partial \beta^T \partial \beta} = \sum_{i=1}^{n} \mathrm{x}_i \mathrm{x}_i^T e^{\mathrm{x}_i^T\beta}$$

Persamaan di atas kemudian disamakan dengan nol, selanjutnya dapat diselesaikan dengan metode literasi numerik yaitu *Newton Rapshon.* Tujuan dari metode iterasi numerik adalah memaksimumkan fungsi ln likelihood. Langkah-langkahnya adalah:

1. Menghitung nilai penduga awal parameter $\beta_{(o)}$ yang diperoleh dari metode OLS atau MKT yaitu

$$\beta_{(o)} = \left(\mathrm{x}^T \mathrm{x}\right)^{-1} \mathrm{x}^T y$$

Di mana $\mathrm{x} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1p} \\ 1 & X_{21} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \cdots & X_{np} \end{bmatrix}$ dan $Y_{n\times 1} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$

2. Membentuk vektor gradien g dengan k adalah banyak parameter yang diduga

$$g^T\left(\beta_{(m)}\right)_{(k+1)} = \left(\frac{\partial \ln L(\beta)}{\partial \beta_0} \quad \frac{\partial \ln L(\beta)}{\partial \beta_1} \quad \cdots \quad \frac{\partial \ln L(\beta)}{\partial \beta_k} \right)_{\beta=\beta_{(m)}}$$

3. Membentuk matriks Hessian H:

$$H\left(\beta_{(m)}\right)_{(k+1)(k+1)} = \begin{bmatrix} \frac{\partial^2 \ln L(\beta)}{\partial \beta_0^2} & \frac{\partial^2 \ln L(\beta)}{\partial \beta_0 \partial \beta_1} & \cdots & \frac{\partial^2 \ln L(\beta)}{\partial \beta_0 \partial \beta_k} \\ & \frac{\partial^2 \ln L(\beta)}{\partial \beta_1^2} & \cdots & \frac{\partial^2 \ln L(\beta)}{\partial \beta_1 \partial \beta_k} \\ & & \ddots & \vdots \\ \textit{simetris} & & & \frac{\partial^2 \ln L(\beta)}{\partial \beta_k^2} \end{bmatrix}$$

4. Memasukkan nilai $\beta_{(o)}$ ke dalam elemen-elemen vector g dan matriks H sehingga diperoleh vektor $g\left(\beta_{(o)}\right)$ dan matriks $H\left(\beta_{(o)}\right)$

5. Mulai dari $m = 0$ dilakukan iterasi pada persamaan $\beta_{(m+1)} = \beta_{(m)} - H^{-1}\left(\beta_{(m)}\right) g\left(\beta_{(m)}\right)$. Nilai $\beta_{(m)}$ merupakan sekumpulan penduga parameter yang konvergen pada iterasi ke-*m*.

6. Penduga parameter yang konvergen diperoleh jika $\left\|\beta_{(m+1)} - \beta_{(m)}\right\| \le \varepsilon$, jika belum diperoleh pendugaan yang konvergen maka dilanjutkan kembali langkah 5 hingga iterasi ke $m = m+1$

4.1.2 Pengujian Parameter Regresi Poisson

Pengujian parameter regresi Poisson dilakukan dengan menggunakan metode *Maximum Likelihood Ratio Test* (MLRT).

Hipotesis:

$H_0 : \beta_1 = \beta_2 = ... = \beta_k = 0$

$H_1 = paling\ sedikit\ ada\ satu\ \beta_i \neq 0\ dengan\ i = 1, 2, ..., k$

Himpunan parameter dibawah populasi:

$$\Omega = \{\beta_0, \beta_1, \beta_2, ..., \beta_k | -\infty < \beta_i < \infty\}$$

Himpunan parameter dibawah H_0 :

$$\omega = \{\beta_0 | -\infty < \beta_0 < \infty\}$$

$L(\hat{\Omega})$ adalah nilai *Maximum Likelihood* untuk model lengkap di mana melibatkan variabel prediktor. $L(\hat{\omega})$ adalah nilai *Maximum Likelihood* untuk model sederhana tanpa melibatkan variabel prediktor. Menurut Agresti (2002) *Likelihood Ratio Test* dapat ditulis dalam persamaan :

$$D(\hat{\beta}) = -2 \ln \left(\frac{L(\hat{\omega})}{L(\hat{\Omega})} \right) = -2\left(\ln L(\hat{\Omega}) - \ln L(\hat{\omega}) \right) \quad 4.10$$

Dimana

$$L(\hat{\omega}) = \frac{\exp\left[-\sum_{i=1}^{n} \exp(\hat{\beta}_0) \right] \left[\exp(\hat{\beta}_0) \sum_{i=1}^{n} y_i \right]}{\prod_{i=1}^{n} y_i\,!} \quad 4.11$$

$$L(\hat{\Omega}) = \frac{\exp\left[-\sum_{i=1}^{n} \exp(x_i^T \hat{\beta}) \right] \left[\prod_{i=1}^{n} \exp(x_i^T \hat{\beta})^{y_i} \right]}{\prod_{i=1}^{n} y_i\,!} \quad 4.12$$

Sehingga didapat

$$D(\hat{\beta}) = 2\sum_{i=1}^{n} \left(\left(y_i x_i^T \hat{\beta} - \exp(x_i^T \hat{\beta}) - \ln y_i\,! \right) - \left(y_i \hat{\beta}_0 - \exp(\hat{\beta}) - \ln y_i\,! \right) \right)$$

$$D(\hat{\beta}) = 2\sum_{i=1}^{n} \left(y_i x_i^T \hat{\beta} - \exp(x_i^T \hat{\beta}) - \left(y_i \hat{\beta}_0 - \exp(\hat{\beta}_0) \right) \right) \quad 4.13$$

$D(\hat{\beta})$ merupakan pendekatan dari subsitusi χ^2 dengan derajat bebas ν, dimana ν adalah jumlah parameter dibawah populasi dikurangi jumlah parameter dibawah H_0. Sehingga kriteria pengujian adalah tolak H_0 jika $D\left(\hat{\beta}\right) > \chi^2_{a,v}$.

Menurut Agresti (2002), apabila H_0 ditolak maka langkah selanjutnya adlah melakukan pengujian parameter secara parsial untuk mengidentifikasi parameter yang memberikan pengaruh signifikan terhadap model. Hipotesis yang digunakan adalah

$H_0 : \beta_i = 0$

$H_1 : \beta_i \neq 0 \text{ dengan } i = 1, 2, ..., k$

Statistik uji

$$Z = \frac{\hat{\beta}_i}{SE\left(\hat{\beta}_i\right)}$$

H_0 ditolak jika $\left|Z_{hitung}\right| > Z_{a/2}$ di mana a adalah tingkat signifikansi yang digunakan. $SE\left(\hat{\beta}_i\right)$ merupakan elemen diagonal ke (i + 1) pada matriks $\operatorname{var}\left(\hat{\beta}\right)$. Nilai $\operatorname{var}\left(\hat{\beta}\right) = -E\left(H^{-1}\left(\hat{\beta}\right)\right)$.

4.2 *Generalized Poisson Regression*

Model *Generalized Poisson Regression* merupakan suatu model yang sesuai diterapkan pada data count yang terjadi pelanggaran asumsi rata-rata sampel sama dengan ragam sampel pada dsitribusi Poisson degan kata lain jika terjadi overdispersi atau underdispersi. sehingga selain μ dalam Genralized Poisson terdapat juga k sebagai parameter dispersi.

Model *Generalized Poisson Regression* mirip dengan model Poisson Regression merupakan suatu model GLM. Model *Generalized Poisson Regression* mengasumsikan bahwa komponen acak berdistribusi Generalized Poisson. Fungsi probabilitas *Generalized Poisson Regression* sebagai berikut:

$$f\left(y_i, \mu_i, k\right) = \left(\frac{\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}}{1 + k\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}}\right)^{y_i} \frac{\left(1 + ky_i\right)^{y_i - 1}}{y_i !} \exp\left(-\frac{\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}\left(1 + ky_i\right)}{\left(1 + k\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}\right)}\right)$$

Dengan $i = 1, 2, \ldots$ dan $\mu_i = \mu_i(x_i) = \exp\left(x_i^T \beta\right)$

Dimana

x_i : vector dari variabel prediktor

β : vector dari parameter disperse

Rata-rata dan ragam dari Y adalah

$E(Y_i) = \mu_i$ dan $Var(Y_i) = \mu_i (1 + k\mu_i)^2$

Pada model *Generalized Poisson Regression* apabila k =1 maka sama dengan model regresi Poisson. Kondisi overdispersi ditunjukkan jika nilai k >1, sedangkan kondisi underdispersi ditunjukkan jika nilai k <1.

Berikut Model *Generalized Poisson Regression*:

$$\begin{aligned} \ln(\mu_i) &= x_i^T \beta = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} \\ \mu_i &= \exp\left(x_i^T \beta\right) = \exp\left(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki}\right) \end{aligned} \tag{4.14}$$

Penduga Parameter *Generalized Poisson Regression*

Metode *Maximum Likelihood Estimation* digunakan untuk menduga parameter dalam *Generalized Poisson Regression* yaitu dengan cara memaksimalkan fungsi likelihood dari parameter (k, β). Menurut Ismail dan Jemain (2015), fungsi likelihood dari *Generalized Poisson Regression* adalah:

$$f(y_i, \mu_i, k) = \left(\frac{\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}}{1 + k\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}} \right)^{y_i} \frac{(1 + ky_i)^{y_i - 1}}{y_i!} \exp\left(- \frac{\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} (1 + ky_i)}{\left(1 + k\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}\right)} \right)$$

Fungsi ln likelihood dari *Generalized Poisson Regression* adalah

$$\ln L(k, \mu) = \sum_{i=1}^{n} \left\{ y_i \ln\left[\frac{\mu_i}{1 + k\mu_i} \right] + (y_i - 1)\ln(1 + ky_i) - \ln(y_i!) - \frac{\mu_i (1 + ky_i)}{1 + k\mu_i} \right\}$$

$$\ln L(k, \beta) = \left(y_i \ln\left(\exp\left(x_i^T \beta\right)\right) - y_i \ln\left(1 + k\exp\left(x_i^T \beta\right)\right) + (y_i - 1)\ln(1 + ky_i) \right) - \sum_{i=1}^{n} \left(\ln(y_i!) - \frac{\exp\left(x_i^T \beta\right)(1 + ky_i)}{1 + k\exp\left(x_i^T \beta\right)} \right)$$

Metode iterasi *Newton Raphson* merupakan metode yang digunakan untuk memaksimalkan fungsi likelihood. Menurut Agesti (2002), algoritma dengan metode *Newton Raphson* adalah:

1. Menghitung nilai duga awal parameter. Nilai dugaan awal dengan metode *Ordinary Least Square* (OLS) :

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

2. Membentuk gradien g dimana n adalah banyak parameter yang diduga

   $$g^T\left(\beta_{(m)}\right)_{(n+1)} = \left(\frac{\partial \ln L(\beta)}{\partial \beta_0} \quad \frac{\partial \ln L(\beta)}{\partial \beta_1} \quad \ldots \quad \frac{\partial \ln L(\beta)}{\partial \beta_n}\right)_{\beta=\beta_{(m)}}$$

3. Membentuk matriks Hessian H

Matriks Hessian dari Generalized Poisson adalah sebagai berikut :

$$H\left(\beta_{(m)}\right)_{(n+1)(n+1)} = \begin{bmatrix} \frac{\partial^2 \ln L(\beta)}{\partial \beta_0{}^2} & \frac{\partial^2 \ln L(\beta)}{\partial \beta_0 \partial \beta_1} & \cdots & \frac{\partial^2 \ln L(\beta)}{\partial \beta_0 \partial \beta_n} & \frac{\partial^2 \ln L(\beta)}{\partial \beta_0 \partial k} \\ & \frac{\partial^2 \ln L(\beta)}{\partial \beta_1{}^2} & \cdots & \frac{\partial^2 \ln L(\beta)}{\partial \beta_1 \partial \beta_n} & \frac{\partial^2 \ln L(\beta)}{\partial \beta_1 \partial k} \\ & & \ddots & \vdots & \vdots \\ & & & \frac{\partial^2 \ln L(\beta)}{\partial \beta_n{}^2} & \frac{\partial^2 \ln L(\beta)}{\partial \beta_n \partial k} \\ simetris & & & & \frac{\partial^2 \ln L(\beta)}{\partial k \partial k^T} \end{bmatrix}$$

4. Memasukkan nilai $\hat{\beta}_{(0)}$ kedalam elemen-elemen vector g dan matriks H sehingga diperoleh vector $\hat{g}_{(0)}$ dan matriks $H_{(0)}$

5. Melakukan iterasi dari m = 0 yaitu nilai dugaan awal dilakukan iterasi pada persamaan : $\hat{\beta}_{m+1} = \hat{\beta}_m - H\left(\hat{\beta}_m\right) g\left(\hat{\beta}_m\right)$

6. Nilai $\hat{\beta}_m$ merupakan kumpulan dari pendugaan parameter pada iterasi ke-m. Proses iterasi pada langkah 5 dilakukan berulang-ulang hingga diperoleh pendugaan parameter yang konvergen, yaitu $\left\|\hat{\beta}_{m+1} - \hat{\beta}_m\right\| \leq \varepsilon$

4.3 Multikolinearitas

Multikolinearitas adalah hubungan antara beberapa atau semua variabel yang menjelaskan model regresi. Terdapat 2 jenis multikolinearitas yaitu multikolinearitas sempurna dan multikolinearitas tidak sempurna. Pada multikolinearitas sempurna terdapat hubungan linear antara variabel prediktor di mana satu variabel prediktor adalah fungsi linear dari variabel prediktor yang lain. Sedangkan multikolinearitas tidak sempurna terjadi apabila terdapat hubungan linear yang tidak sempurna antar variabel prediktor (Gujarati, 1991).

Menurut Li (2000), pendeteksian multikolinearitas dapat dilakukan menggunakan nilai *Variance Inflation Factor* (VIF). Untuk regresi dengan lebih dari 2 variabel :

$$VIF_j = \frac{1}{1 - R_j^{\,2}}$$

Di mana:

$R_j^{\,2}$: koefisien determinasi dari *auxiliary regression*

Auxiliary regression adalah regresi X_j sebagai variabel respon, dan X lainnya sebagai variabel prediktor. Nilai $R_j^{\,2}$ berkisar antara 0 sampai 1 sehingga nilai VIF akan naik seiring dengan kenaikan koefisien determinasi dari *auxiliary regression.* Nilai VIF yang lebih dari 10 merupakan bukti cukup untuk mendeteksi adanya multikolinearitas (Li, 2000)

Multikolinearitas sempurna yang terjadi dalam data menyebabkan koefisien regresi menjadi *underminerd* (tidak dapat diduga), sedangkan pada multikolinearitas tidak sempurna dapat menyebabkan ragam dari penduga kuadrat terkecil menjadi relative besar walaupun penduga tersebut masih tetap dapat diduga dan bersifat BLUE (*Best Linear Unbiased Estimators)* serta tetap efisien (ragam dari penduga paling kecil dari semua penduga yang mungkin). Selain itu multikolinearitas tidak sempurna juga dapat menyebabkan selang kepercayaan menjadi lebih lebar sehingga koefisien regresi menjadi tidak nyata (Gujarati, 1991).

4.4 Overdispersi

Terdapat beberapa asumsi yang harus dipenuhi dalam model regresi poisson. Salah satu asumsi adalah kesamaan antara rata-rata (*mean)* dan variansi (*variance*) yang disebut ekuidispersi (Darnah, 2011). Namun sering dijumpai data yang variansinya lebih kecil atau lebih besar dari rataannya. Keadaan ini disebut dengan underdispersi (*underdispersionin*) atau overdispersi (*overdisperslion*). Beberapa hal yang menyebabkan overdispersi, diantaranya:

1) Terdapat korelasi antar pengamatan
2) Terdapat pelanggaran asumsi distribusi poisson, yaitu

$$Var(Y) > E(Y)$$

3) Terdapat excess zero
4) Terdapat outlier dalam data

Overdispersi atau underdispersi dapat menyebabkan taksiran parameter yang diperoleh tidak efisien mengakibatkan model yang dihasilkan menjadi kurang tepat dan berakibat fatal pada intepretasi model, terutama pada estimasi parameter model karena dapat menaksir standart error yang terlalu rendah dan dapat memberikan kesimpulan yang keliru tentang signifikan atau tidaknya parameter regresi yang terlibat (Darnah, 2011).

Rumus untuk mendeteksi overdispersi yaitu:

- Nilai Deviance

$$\theta_1 = \frac{D}{db} > 1; D = 2\sum_{i=1}^{n} \left\{ y_i \log \frac{y_i}{\mu_i} \right\}$$

- Nilai Pearson Chi-Square

$$\theta_2 = \frac{x^2}{db} > 1; x^2 = \sum_{i=1}^{n} \frac{(y_i - \mu_i)^2}{\sigma_i}$$

Jika hasil bagi antara nilai statistik D terhadap derajat bebasnya atau statistik χ^2 terhadap derajat bebasnya lebih dari 1 maka data mengalami overdispersi, sedangkan jika nilai hasil bagi lebih kecil atau kurang dari 1, maka dinyatakan kasus data sebagai kasus underdispersi.

Uji statistik yang bisa juga digunakan untuk mendeteksi overdispersi pada suatu data adalah uji overdispersi yang dapat menggunakan package AER dari software R studio (Herindrawati, Latra, Purhadi, 2017).

Keputusan hipotesis:

H_0 : Tidak terjadi overdispersi

H_1 : Terjadi overdispersi

Keputusan yang diambil menggunakan software R yaitu jika nilai p-value < α maka H_0 ditolak yang berarti terjadi overdispersi. Tetapi jika p-value > α maka H_0 diterima yang berarti tidak terjadi overdispersi.

4.5 *Goodness of Fit*

Pengujian Kesesuaian Model untuk mengetahui model yang digunakan sesuai atau tidak dengan data yang diamati. Pengujian kesesuaian model dilakukan dengan menggunakan uji *goodness of fit* yang bertujuan untuk mengambil kesimpulan mengenai sebaran populasi. Suatu contoh acak dipilih dari populasi yang bersangkutan, kemudian informasi contoh tersebut digunakan untuk menguji kebenaran sebaran populasi tersebut. Uji ini didasarkan pada seberapa baik kesesuaian atau kecocokan (*goodness of fit*) antara frekuensi pengamatan data yang diperoleh. Data sampel dengan frekuensi harapan dari distribusi yang dihipotesiskan. Dalam hal ini akan ditunjukkan apakah data sudah sesuai atau tidak dengan model yang didapatkan dengan menggunakan software R Studio.

Pengujian model dengan melihat nilai uji *goodness of fit* menggunakan hipotesis sebagai berikut:

H_0 : Model yang didapat sesuai dengan data yang diamati

H_1 : Model yang didapat tidak sesuai dengan data yang diamati

Dengan taraf signifikan $\alpha = 0,05$

Kriteria uji:

H_0 ditolak jika nilai Pearson Chi Square lebih kecil dari $\alpha = 0,05$

4.6 Pemilihan Model Terbaik Metode AIC

Akaike Information Criterion (AIC) adalah kriteria kesesuaian model dalam menduga model secara statistic. Kriteria AIC digunakan apabila pemodelan regresi bertujuan untuk mengidentifikasi faktor-faktor yang berpengaruh terhadap model. Perhitungan nilai AIC menggunakan persamaan:

$$AIC = -2\ln \mathrm{L}(\hat{\theta}) + 2p \quad 4.15$$

$$AIC = -2\ln\left(\max imum\, likelihood\right) + 2\left(number\ of\ \ parameters\right)$$

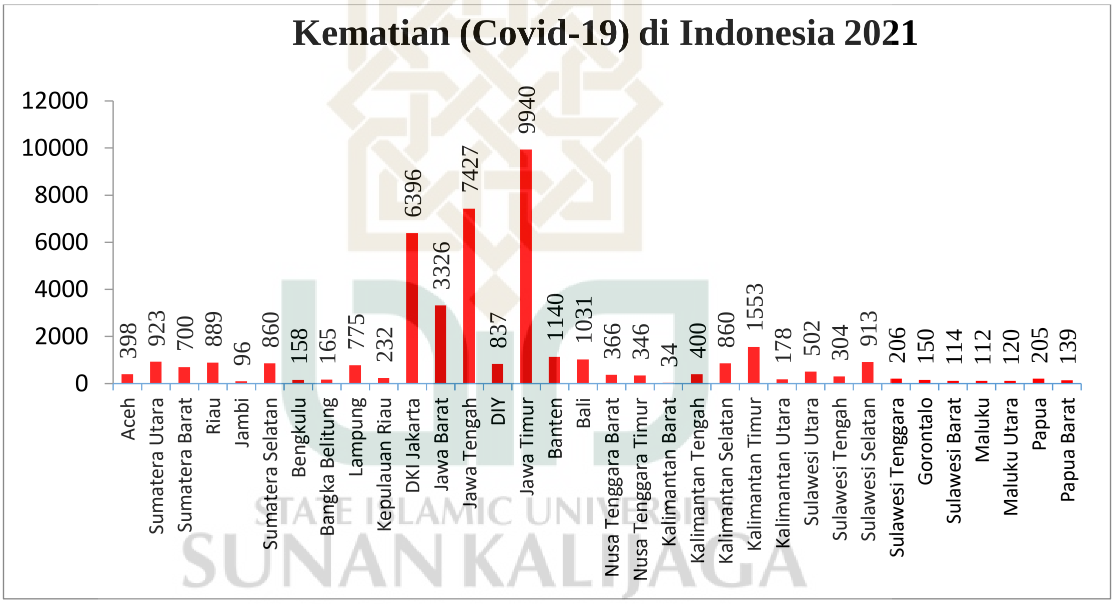
Model regresi terbaik adalah model regresi yang menghasilkan nilai AIC terkecil (Akaike, 1978)

BAB V

STUDI KASUS

5.1 Deskripsi Data

Penelitian ini memodelkan Tingkat Kematian Akibat Covid-19 Di Indonesia tahun 2021 menggunakan model regresi Poisson. Data yang akan digunakan dalam penelitian ini adalah data covid 19 di Indonesia per 4 Maret 2021 yang diperoleh dari situs website *http://Covid19.go.id/*. Berikut data kematian akibat Covid-19 di Indonesia:



Gambar 5.1 Data kematian akibat Covid-19 di Indonesia 2021

Gambar diatas memperlihatkan grafik angka kematian akibat Covid-19 di Indonesia pada tahun 2020. Pada gambar dapat dilihat angka kematian tertinggi terdapat di provinsi Jawa Timur, sebesar 9940 jiwa dan angka kematian terendah di provinsi Kalimantan Barat, sebesar 34 jiwa. Dalam penelitian ini ingin mengetahui seberapa besar pengaruh faktor-faktor yang diteliti terhadap pertumbuhan angka kematian akibat Covid-19 yang terjadi di Indonesia pada tahun 2020.

Sebelum melakukan pemodelan pada faktor-faktor yang ada, perlu dilakukan analisis secara statistik deskriptif untuk mengetahui karakteristik setiap variabel. Analisis statistik deskriptif tingkat kematian akibat Covid-19 di Indonesia tahun 2020 beserta variabel yang mempengaruhi disajikan dalam tabel berikut:

Tabel 5.1 Deskriptif data kematian akibat Covid-19 di Indonesia

Descriptive Statistics

	N	Minimum	Maximum	*Mean*	Std. Deviation	Variance
Y	34	34	9940	1229,2647	2244,92203	5039674,928
X_1	34	48,2	79,4	58,2265	5,50137	30,265
X_2	34	20,6	51,8	41,7735	5,50137	30,265
X_3	34	5,8	27,3	13,2824	3,97892	15,832
X_4	34	20,4	50	36,5441	5,95204	35,427
X_5	34	29,3	71,6	44,7618	8,54113	72,951
X_6	34	0	28,6	8,3382	7,15821	51,24
X_7	34	0	50	10,9412	12,45134	155,036
X_8	34	0	50	6,9	10,11704	102,355
X_9	33	0	10,6	1,9152	2,91259	8,483
X_{10}	34	0	10	1,5676	2,29237	5,255
X_{11}	34	1	13	3,8824	2,77177	7,683
X_{12}	34	5392	165158	36593	40562,15498	1645288416

Berdasarkan tabel diketahui bahwa rata-rata tingkat kematian akibat Covid-19 di Indonesia tahun 2020 sebesar 1229 jiwa. Menurut jenis kelamin rata-rata kematian berjenis kelamin laki-laki (X_1) adalah 58,2265%, rata-rata kematian berjenis kelamin perempuan (X_2) adalah 41,7735%. Menurut usia rata-rata kematian umur 35-45 tahun (X_3) adalah 13,2824%, rata-rata kematian umur 46-60 tahun (X_4) adalah 36,5441%, rata-rata kematian lebih dari 60 tahun (X_5) adalah 44,7618%. Menurut riwayat penyakit rata-rata kematian dengan hipertensi (X_6) adalah 8,3382%, rata-rata kematian dengan diabetes melitus (X_7) adalah 10,9412%, rata-rata kematian dengan jantung (X_8) adalah 6,9%, rata-rata kematian dengan ginjal (X_9) adalah 1,9152%, rata-rata kematian dengan paru-paru (X_{10}) adalah 1,5676%. Faktor tersebut merupakan faktor internal yang mempengaruhi angka kematian akibat Covid-19 di Indonesia tahun 2020. Sedangkan faktor eksternal yang mempengaruhi yaitu, menurut fasilitas kesehatan rata-rata rumah sakit rujukan (X_{11}) sebanyak 4 rumah sakit, rata-rata tenaga kesehatan (X_{12}) sebanyak 36593 orang.

5.2 Uji Kesesuaian Distribusi Poisson

Uji kesesuaian distribusi Poisson di uji pada data dependen yaitu variabel jumlah kematian akibat Covid-19 di Indonesia tahun 2020. Pengujian kesesuaian distribusi Poisson menggunakan Uji Kolmogrov Smirnov. Hipotesis dari pengujian ini adalah

H_0 : Data jumlah kematian akibat Covid-19 berdistribusi Poisson

H_1 : Data jumlah kematian akibat Covid-19 tidak berdistribusi Poisson

Tabel 5.2 Uji *Kolmogrov-Smirnov*

One-Sample Kolmogorov-Smirnov Test

		Y
N		34
Poisson Parameter[a,b]	Mean	1229,2647
Most Extreme Differences	Absolute	0,848
	Positive	0,848
	Negative	-0,147
Kolmogorov-Smirnov Z		4,943
Asymp, Sig, (2-tailed)		0,000

Berdasarkan hasil uji kolmogrov Smirnov yang ditunjukkan oleh tabel, diperoleh nilai signifikansi sebesar 0,000 dimana nilai signifikansi tersebut lebih kecil dari $\alpha = 0,05$. Sehingga H_0 ditolak yang artinya data jumlah kematian akibat Covid-19 tidak berdistribusi Poisson.

5.3 Uji Multikolinearitas

Uji multikolinearitas bertujuan untuk mengetahui ada atau tidaknya penyimpangan asumsi klasik multikolinearitas yaitu adanya hubungan linier antara variabel independent dalam model regresi. Pengujian ada atau tidaknya gejala multikolinearitas dilakukan dengan melihat nilai VIF (*Variane Inflation Factor*) dan *Tolerance*. Rumus yang digunakan untuk mendapatkan nilai VIF adalah

$$VIF = \frac{1}{0,529} = 1,891$$

Tabel 5.3 Multikolinearitas

Coefficients[a]

Model		Collinearity Statistics	
		Tolerance	VIF
1	X_2	0,529	1,891
	X_3	0,098	10,200
	X_4	0,131	7,616
	X_5	0,059	17,093
	X_6	0,741	1,349
	X_7	0,300	3,330
	X_8	0,365	2,737
	X_9	0,485	2,062
	X_{10}	0,710	1,408
	X_{11}	0,176	5,668
	X_{12}	0,178	5,624

Model regresi dikatakan tidak mengandung multikolinearitas apabila nilai VIF berada di bawah 10 dan nilai *Tolerance* lebih dari 0,1. Berdasarkan uji multikolinearitas pada tabel dapat dilihat variabel X_1 diperoleh nilai *Tolerance*

0,000 lebih kecil dari 0,1 dan VIF 0,000. Variabel X_3 memiliki Nilai *Tolerance* 0,098 lebih kecil dari 0,1 dan VIF 10,200 lebih besar dari 10. Variabel X_5 memiliki nilai *Tolerance* 0,059 lebih kecil dari 0,1 dan VIF 17,093 lebih besar dari 10. Sehingga dapat disimpulkan bahwa variabel X_1, X_3, dan X_5 mengalami multikolinearitas. Oleh karena itu perlu dilakukan penyembuhan data sebelum dilakukan analisis regresi. Ada beberapa cara untuk menyembukan data dengan multikolinearitas yaitu dengan transformasi data, matriks korelasi, dan eliminasi variabel prediktor yang terjadi multikolinearitas. Karena transformasi data dan matriks korelasi menimbulkan banyak data yang harus di eliminasi dan masih mengandung multikolinearitas. Maka penyembuhan dilakukan dengan mengeliminasi variabel prediktor yang mengandung multikolinearitas sehingga diperoleh:

Tabel 5.4 Uji multikolinearitas setelah penyembuhan

Coefficients

Model		Collinearity Statistics	
		Tolerance	VIF
1	X_3	0,788	1,269
	X_4	0,821	1,218
	X_6	0,745	1,342
	X_7	0,302	3,315
	X_8	0,368	2,719
	X_9	0,490	2,042
	X_{10}	0,744	1,344
	X_{11}	0,183	5,472
	X_{12}	0,180	5,555

Nilai VIF pada masing-masing variabel kurang dari 10 dan memiliki nilai *Tolerance* lebih besar dari 0,1. Dapat disimpulkan bahwa sudah tidak terjadi korelasi yang cukup tinggi pada variabel prediktor tersebut. Selanjutnya analisis data menggunakan model regresi dapat dilanjutkan.

5.4 Pemodelan Regresi Poisson

5.4.1 Model Regresi Poisson

Bentuk umum dari model regresi Poisson adalah

$$y_i = \mu_i + \varepsilon_i$$

$$y_i = \exp\left(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_p x_{pi}\right) + \varepsilon_i$$

Selanjutnya akan dilakukan pemodelan faktor-faktor yang mempengaruhi jumlah kematian akibat Covid-19. Yaitu dengan mencari nilai taksiran parameter β terlebih dahulu. Dengan menggunakan software SPSS atau R studio.

Tabel 5.5 Hasil penaksiran parameter β model regresi Poisson

Parameter	B	Std. Error
(Intercept)	6,878655	,039470
X_3	-0,090950	,002370
X_4	-0,007349	,001159
X_6	0,012585	,001102
X_7	0,011787	,000998
X_8	-0,007395	,001179
X_9	0,038577	,003122
X_{10}	-0,062727	,004436
X_{11}	-0,001875	,004657
X_{12}	0,000022	,0000003

5.4.2 Uji Signifikansi Parameter

Uji Serentak Parameter Model Regresi Poisson

Tabel 5.6 Uji *Goodness of fit*

***Goodness of fit*[a]**

	Value	df	Value/df
Deviance	4620,707	24	192,529
Scaled Deviance	4620,707	24	
Pearson Chi-Square	4982,822	24	207,618
Scaled Pearson Chi-Square	4982,822	24	
Log Likelihood[b]	-2446,402		

Hasil Pengujian menunjukkan nilai statistik uji G adalah 4982,822 dengan derajat bebas (df) 24 dan nilai $\chi^2_{(0.05;24)} = 36,415$. pada taraf signifikansi $a = 0,05$ diperoleh $G > \chi^2_{(0.05;24)}$ sehingga dapat diambil keputusan bahwa H_0 di tolak. Dapat disimpulkan bahwa sembilan variabel prediktor secara bersama-sama memberikan pengaruh terhadap jumlah kematian akibat Covid-19 di Indonesia.

Uji Parsial Parameter Model Regresi Poisson

Setelah mendapatkan hasil taksiran parameter maka selanjutnya akan dilakukan pengujian parameter untuk mendapatkan variabel prediktor yang berpengaruh secara nyata terhadap variabel respon.

Tabel 5.7 Uji Wald

Parameter	Wald Chi-Square	df	P-value
(Intercept)	30372,46	1	0,000
X_3	1473,171	1	0,000
X_4	40,2	1	0,000
X_6	130,414	1	0,000
X_7	139,555	1	0,000
X_8	39,353	1	0,000
X_9	152,716	1	0,000
X_{10}	199,931	1	0,000
X_{11}	0,162	1	0,687
X_{12}	4768,146	1	0,000

Hipotesis:

$H_0 : \beta_j = 0, \textit{minimal terdapat satu variabel prediktor yang tidak signifikan}$

$H_1 : \beta_j \neq 0, \textit{minimal terdapat satu variabel prediktor signifikan}$

Kriteria pengujian Uji Wald adalah tolak H_0 jika nilai dari Uji Wald lebih besar dari nilai $\chi^2_{(a,1)}$ atau nilai sig < $a = 0,05$. Berdasarkan hasil pengujian parameter yang telah dilakukan terdapat sembilan variabel yang memiliki pengaruh signifikan, di antaranya persentase usia 31 – 45 (X_3), persentase usia 45 – 59 (X_4), persentase riwayat penyakit hipertensi (X_6), persentase riwayat penyakit diabetes melitus (X_7), persentase riwayat penyakit jantung (X_8), persentase riwayat penyakit ginjal (X_9), persentase riwayat penyakit Paru-paru (X_{10}), jumlah tenaga kesehatan (X_{12}). Dan terdapat satu variabel yang tidak memiliki pengaruh signifikan yaitu jumlah rumah sakit rujukan (X_{11}). Dengan demikian model angka kematian akibat Covid-19 menggunakan regresi Poisson dapat dituliskan sebagai berikut:

$$y_i = \exp\left(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi}\right) + \varepsilon_i$$

$$\begin{aligned}\hat{y} = \exp(&6{,}878655 - 0{,}090950x_3 - 0{,}007349x_4 + 0{,}012585x_6 + 0{,}011787x_7 \\ &-0{,}007395x_8 + 0{,}038577x_9 - 0{,}062727x_{10} - 0{,}001875x_{11} + 0{,}000022x_{12})\end{aligned}$$

5.4.3 Uji Asumsi Equidispersi

Asumsi yang harus dipenuhi dalam regresi Poisson yaitu nilai *mean* dan nilai varians variabel respon harus memiliki nilai yang sama.

Tabel 5.8 Deskriptif data variabel respon

Descriptive Statistics				
	N	*Mean*	Std. Deviation	Variance
Y	34	1229,2647	2244,92203	5039674,928

Berdasarkan nilai *mean* dan varians variabel respon diatas dapat disimpulkan bahwa model regresi Poisson biasa tidak dapat digunakan untuk melanjutkan pemodelan angka kematian akibat Covid-19 karena tidak memenuhi asumsi *equidispersi*. Dapat dilihat juga bahwa nilai varians variabel respon lebih besar dari pada nilai *mean*-nya yang artinya data yang diamati mengalami *overdispersi*.

Berdasarkan tabel 5.6 diatas nilai parameter dispersi dari pembagian nilai deviance dan derajat bebasnya yaitu 192,529 yang lebih besar dari 1. Dapat disimpulkan bahwa variabel respon mengalami kasus *overdispersi*.

5.5 Model *Generalized Poisson Regression*

Model *Generalized Poisson Regression* pada dasarnya sama dengan model Regresi Poisson biasa namun pada model *Generalized Poisson Regression* dilakukan dengan melibatkan satu atau beberapa variabel prediktor. Dengan demikian akan didapatkan beberapa model dan akan di pilih berdasarkan nilai AIC-nya.

Data yang digunakan berupa variabel respon yaitu angka kematian akibat Covid-19 (*Y*) dengan sembilan variabel prediktor yaitu persentase kematian umur 35-45 tahun (X_3) persentase kematian umur 46-60 tahun (X_4) persentase kematian dengan hipertensi (X_6) persentase kematian dengan diabetes melitus (X_7) persentase kematian dengan jantung (X_8) persentase kematian dengan ginjal (X_9) persentase kematian dengan paru-paru (X_{10}) jumlah rumah sakit rujukan (X_{11}) jumlah tenaga kesehatan (X_{12}).

Tabel 5.9 Penaksiran parameter β model *Generalized Poisson Regression*

Parameter	B	Std. Error
(Intercept)	6.716000	0,441600
X_3	-0,079890	0,023400
X_4	-0,008183	0,012040
X_6	0,004507	0,011940
X_7	0,004789	0,010580
X_8	0,000435	0,012250
X_9	0,054440	0,033860
X_{10}	-0,051250	0,043420
X_{11}	0,049690	0,051570
X_{12}	0,000019	0,000003

Untuk menentukan model terbaik yang digunakan maka dicari nilai AIC yang paling kecil.

Tabel 5.10 AIC model *Generalized Poisson Regression*

Model	Nilai AIC
X_{12}	497,6945
X_9, X_{12}	494,1961
X_3, X_9, X_{12}	481,4682
X_3, X_6, X_9, X_{12}	482,8608
X_3, X_6, X_7, X_9, X_{12}	484,6203
X_3, X_6, X_7, X_8, X_9, X_{12}	486,4703
X_3, X_4, X_6, X_7, X_8, X_9, X_{12}	488,2413
X_3, X_4, X_6, X_7, X_8, X_9, X_{10}, X_{12}	488,7205
X_3, X_4, X_6, X_7, X_8, X_9, X_{10}, X_{11}, X_{12}	489,8243

Berdasarkan tabel diatas dapat dijelaskan bahwa model terbaik berdasarkan kriteria nilai AIC adalah model yang melibatkan semua variabel prediktor, namun karena data yang diamati mengalami kasus overdispersi dan juga terdapat variabel prediktor yang tidak signifikan berpengaruh terhadap variabel respon setelah di uji menggunakan model regresi Poisson sehingga dikatakan tidak layak digunakan. Oleh karena itu, model terbaik untuk pemodelan angka kematian akibat Covid-19 di Indonesia adalah dengan menggunakan *Generalized Poisson Regression* dengan model sebagai berikut:

$$\hat{y}_i = \exp\left(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi}\right) + \varepsilon_i$$

$$\begin{aligned}\hat{y} &= \exp(6{,}716000 - 0{,}079890x_3 - 0{,}008183x_4 + 0{,}004507x_6 + 0{,}004789x_7 \\ &\quad + 0{,}000435x_8 + 0{,}054440x_9 - 0{,}051250x_{10} + 0{,}049690x_{11} + 0{,}000019x_{12})\end{aligned}$$

Dengan model yang signifikan sebagai berikut:

$$\hat{y}_i = \exp\left(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi}\right) + \varepsilon_i$$

$$\begin{aligned}\hat{y} &= \exp(6{,}716000 - 0{,}079890x_3 - 0{,}008183x_4 + 0{,}004507x_6 + 0{,}004789x_7 \\ &\quad + 0{,}000435x_8 + 0{,}054440x_9 - 0{,}051250x_{10} + 0{,}000019x_{12})\end{aligned}$$

Dengan model terbaik yang dilihat dari nilai AIC terkecil :

$$\hat{y}_i = \exp\left(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi}\right) + \varepsilon_i$$

$$\hat{y} = \exp(6{,}62600 - 0{,}08548x_3 + 0{,}04589x_9 + 0{,}00002x_{12})$$

5.6 Intepretasi Hasil

Model *Generalized Poisson Regression* terhadap jumlah kematian akibat Covid-19 di Indonesia sebagai berikut:

$$\begin{aligned}\hat{y} &= \exp(6{,}716000 - 0{,}079890x_3 - 0{,}008183x_4 + 0{,}004507x_6 + 0{,}004789x_7 \\ &\quad + 0{,}0004351x_8 + 0{,}054440x_9 - 0{,}051250x_{10} + 0{,}000019x_{12})\end{aligned}$$

Berdasarkan model di atas dapat dikatakan bahwa variabel yang berpengaruh secara nyata (signifikan) adalah persentase kematian umur 35-45 tahun (X_3) persentase kematian umur 46-60 tahun (X_4) persentase kematian dengan hipertensi (X_6) persentase kematian dengan diabetes melitus (X_7) persentase kematian dengan jantung (X_8) persentase kematian dengan ginjal (X_9)

persentase kematian dengan paru-paru (X_{10}) jumlah tenaga kesehatan (X_{12}). Pada model dapat dilihat bahwa jumlah kematian akibat Covid-19 di Indonesia akan bertambah atau berkurang jika:

1. Persentase kematian umur (35-45) tahun (X_3) bertambah satu satuan akan menyebabkan jumlah kematian menjadi exp(-0,079890) = 0,923218 kali dari nilai awal dengan asumsi peubah lain dianggap tetap.
2. Persentase kematian umur 46-60 tahun (X_4) bertambah satu satuan akan menyebabkan jumlah kematian menjadi exp(-0,008183) = 0,991850 kali dari nilai awal dengan asumsi peubah lain dianggap tetap.
3. Persentase kematian dengan hipertensi (X_6) bertambah satu satuan akan menyebabkan jumlah kematian menjadi exp(0,004507) = 1,004517 kali dari nilai awal dengan asumsi peubah lain dianggap tetap.
4. Persentase kematian dengan diabetes melitus (X_7) bertambah satu satuan menyebabkan jumlah kematian menjadi exp(0,004789) = 1,004800 kali dari nilai awal dengan asumsi peubah lain dianggap tetap.
5. Persentase kematian dengan jantung (X_8) bertambah satu satuan akan menyebabkan jumlah kematian menjadi exp(0,000435) = 1,000435 kali dari nilai awal dengan asumsi peubah lain dianggap tetap.
6. Persentase kematian dengan ginjal (X_9) bertambah satu satuan akan menyebabkan jumlah kematian menjadi exp(0,054440) = 1,055949 kali dari nilai awal dengan asumsi peubah lain dianggap tetap.
7. Persentase kematian dengan paru-paru (X_{10}) bertambah satu satuan akan menyebabkan jumlah kematian menjadi exp(-0,051250) = 0,950041 kali dari nilai awal dengan asumsi peubah lain dianggap tetap.
8. Jumlah tenaga kesehatan (X_{12}) bertambah satu satuan akan menyebabkan jumlah kematian menjadi exp(0,000019) = 1,000019 kali dari nilai awal dengan asumsi peubah lain dianggap tetap.

Model *Generalized Poisson Regression* Terbaik yang di lihat dari nilai AIC terkecil terhadap jumlah kematian akibat Covid-19 di Indonesia:

$$\hat{y} = \exp(6{,}62600 - 0{,}08548x_3 + 0{,}04589x_9 + 0{,}00002x_{12})$$

Pada model dapat dilihat bahwa jumlah kematian akibat Covid-19 di Indonesia akan bertambah atau berkurang jika:

1. Persentase kematian umur (35-45) tahun (X_3) bertambah satu satuan akan menyebabkan jumlah kematian menjadi exp(-0,08548) = 0,91807 kali dari nilai awal dengan asumsi peubah lain dianggap tetap.
2. Persentase kematian dengan ginjal (X_9) bertambah satu satuan akan menyebabkan jumlah kematian menjadi exp(0,04589) = 1,04695 kali dari nilai awal dengan asumsi peubah lain dianggap tetap.
3. Jumlah tenaga kesehatan (X_{12}) bertambah satu satuan akan menyebabkan jumlah kematian menjadi exp(0,00002) = 1,00002 kali dari nilai awal dengan asumsi peubah lain dianggap tetap.