ANALISIS PEUBAH RESPONS KONTINU NON NEGATIF DENGAN REGRESI *INVERSE GAUSSIAN*

SKRIPSI

Diajukan kepada Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Negeri Yogyakarta untuk memenuhi sebagian persyaratan
guna memperoleh gelar Sarjana Sains

Oleh:
Neni Rosalina
06305141046

PROGRAM STUDI MATEMATIKA
JURUSAN PENDIDIKAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI YOGYAKARTA
2011

ANALISIS PEUBAH RESPONS KONTINU NON NEGATIF DENGAN REGRESI *INVERSE GAUSSIAN*

SKRIPSI

Diajukan kepada Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Negeri Yogyakarta untuk memenuhi sebagian persyaratan
guna memperoleh gelar Sarjana Sains

Oleh:
Neni Rosalina
06305141046

PROGRAM STUDI MATEMATIKA
JURUSAN PENDIDIKAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI YOGYAKARTA
2011

PERSETUJUAN

SKRIPSI

ANALISIS PEUBAH RESPONS KONTINU NON NEGATIF DENGAN REGRESI *INVERSE GAUSSIAN*

Oleh :

Neni Rosalina
06305141046

Telah disetujui pada tanggal 1 April 2011

untuk dipertahankan di depan dewan penguji skripsi

Program Studi Matematika

Jurusan Pendidikan Matematika

Fakultas Matematika dan Ilmu Pengetahuan Alam

Universitas Negeri Yogyakarta

Menyetujui,

Pembimbing,

Kismiantini, M. Si.
NIP. 19790816 200112 2 001

SKRIPSI

ANALISIS PEUBAH RESPONS KONTINU NON NEGATIF DENGAN REGRESI *INVERSE GAUSSIAN*

Oleh :
Neni Rosalina
06305141046

Telah dipertahankan di depan dewan penguji skripsi Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Yogyakarta pada tanggal 8 April 2011 dan dinyatakan telah memenuhi persyaratan guna memperoleh gelar Sarjana Sains.

DEWAN PENGUJI

Nama	Jabatan	Tanda Tangan	Tanggal
Kismiantini, M.Si	Ketua Penguji	....	18/4/2011
Retno Subekti, M.Sc	Sekretaris Penguji	....	18/4-2011
Dr. Djamilah B. W.	Penguji Utama	....	13/4-2011
Mathilda S., M.Si	Penguji Pendamping	....	13 April 2011

Yogyakarta, 25 April 2011
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Negeri Yogyakarta
Dekan,

Dr. Ariswan
NIP. 19590914 198803 1 003

PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Neni Rosalina

NIM : 06305141046

Prodi/ Jurusan : Matematika/ Pendidikan Matematika

Fakultas : Matematika dan Ilmu Pengetahuan Alam

Judul TAS : Analisis Peubah Respons Kontinu Non Negatif dengan Regresi *Inverse Gaussian*

Dengan ini saya menyatakan bahwa skripsi ini adalah hasil pekerjaan saya sendiri dan sepanjang pengetahuan saya tidak berisi materi yang dipublikasikan atau ditulis oleh orang lain atau telah digunakan sebagai persyaratan penyelesaian studi di Perguruan Tinggi lain kecuali pada bagian tertentu yang saya ambil sebagai acuan.

Apabila ternyata terbukti pernyataan ini tidak benar, sepenuhnya menjadi tanggung jawab saya dan saya bersedia menerima sanksi sesuai peraturan yang berlaku.

Yogyakarta, 1 April 2011

Yang menyatakan,

Neni Rosalina
NIM. 06305141046

MOTTO

"Ketahuilah bahwa kemenangan akan datang bersama kesabaran, jalan keluar
datang bersama kesulitan dan kemudahan itu ada bersama kesulitan"
(QS. Ath-Thalaq: 7)

Orang-orang melupakan seberapa cepat anda melakukan sebuah pekerjaan - tetapi mereka ingat seberapa baik anda melakukannya. (People forget how fast you did a job ? but they remember how well you did it.)
(Howard Newton)

Masa depan tergantung pada apa yang kita lakukan saat ini. (The future depends on what we do in the present.)
(Mahatma Gandhi)

Syukuri apa yang ada hidup adalah anugrah tetap jalani hidup ini melakukan yang
terbaik. Tuhan pastikan menunjukkan kebesaran dan kuasaNya bagi hambanya
yang sabar dan tak kenal putus asa, jangan menyerah.....jangan
menyerah.....jangan menyerah.......
(D'Masiv)

PERSEMBAHAN

Karya sederhana ini aku persembahkan dengan segenap kasih teruntuk:

- *Tuhan Yang Maha Esa yang telah melimpahkan kasih dan karunia_Nya sehingga penulis dapat menyelesaikan penulisan skripsi ini.*
- *Kedua orang tuaku,Bapak & Ibu atas cinta, doa, dukungan, pengorbanan dan kesempatannya sehingga aku bisa menuntut ilmu sampai saat ini.*
- *Kakak2ku (Mbak Dian, Mas Gembong & Mas Dery) dan keponakan2ku: (Dea & Tia) atas dukungan, kegembiraan dan motivasinya.*
- *Sahabat-sahabatku: Nopek, Ani, Pungky, dan Beby yang tidak pernah bosan memberiku kata "semangat".*
- *Teman-teman kos Karangmalang A21 (Desi, Yuli, Uma, Ari, Siwi, Nida, Titis, Widya dan Heni) atas persahabatan selama ini serta bapak ibu kos Pak Juri, Ical, Mbak Rini dan Mas Noer atas perhatian yang telah diberikan.*
- *Teman-teman KKN 2009 kelompok 56 di Dusun Kare (Kitty, Mbah Dibyo, Yani, Danish, Abel, Reza, dan Mas Eri) atas motivasi dan pertemanannya.*

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan kehadirat Tuhan Yang Maha Esa, yang telah memberikan segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul “Analisis Peubah Respons Kontinu Non Negatif dengan Regresi *Inverse Gaussian*” ini guna memenuhi persyaratan untuk memperoleh gelar Sarjana Sains pada Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Yogyakarta.

Penulis mengucapkan terima kasih kepada:

1. Bapak Dr. Ariswan, sebagai Dekan FMIPA Universitas Negeri Yogyakarta yang telah memberikan kesempatan penulis dalam menyelesaikan studi.
2. Bapak Dr. Hartono, sebagai Ketua Jurusan Pendidikan Matematika FMIPA Universitas Negeri Yogyakarta yang telah memberikan kemudahan pengurusan administrasi selama penyusunan skripsi ini.
3. Ibu Atmini Dhoruri, MS, sebagai Ketua Program Studi Matematika FMIPA Universitas Negeri Yogyakarta yang telah memberikan dukungan untuk kelancaran studi.
4. Ibu Kismiantini M. Si., sebagai pembimbing yang telah memberikan banyak bimbingan, saran, bantuan serta masukan selama penyusunan skripsi ini.
5. Seluruh dosen Jurusan Pendidikan Matematika FMIPA Universitas Negeri Yogyakarta yang telah memberikan ilmu kepada penulis.

6. Teman-teman Matematika Reguler 2006, untuk semua kritik dan pendapatnya kepada penulis.
7. Teman-teman KSR PMI Unit UNY yang selalu memberikan support kepada penulis.
8. Semua pihak yang telah membantu sehingga skripsi ini bisa terselesaikan.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih terdapat banyak kekurangan baik isi maupun susunannya. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun dari berbagai pihak demi perbaikan skripsi ini. Semoga skripsi ini dapat bermanfaat bagi para pembaca.

Yogyakarta, 1 April 2011

Penulis,

Neni Rosalina
06305141046

ANALISIS PEUBAH RESPONS KONTINU NON NEGATIF DENGAN REGRESI *INVERSE GAUSSIAN*

Oleh:
Neni Rosalina
06305141046

ABSTRAK

Analisis regresi merupakan suatu analisis yang digunakan untuk mengetahui hubungan antara peubah respons (Y) dengan satu atau beberapa peubah penjelas (X). Bila peubah respons berupa peubah respons kontinu non negatif dan membentuk kurva menceng ke kanan maka kurang tepat diselesaikan dengan regresi linear klasik. Pada regresi linear klasik, peubah respons diasumsikan sebagai peubah kontinu yang berdistribusi normal sehingga nilainya terletak $-\infty < Y < \infty$. Penanganan permasalahan tersebut dapat diselesaikan dengan regresi *inverse Gaussian*. Tujuan penulisan ini adalah mengetahui cara mengestimasi parameter pada analisis peubah respons kontinu non negatif dengan regresi *inverse Gaussian* menggunakan metode *MLE* dan mengetahui contoh penerapan regresi *inverse Gaussian*.

Metode yang digunakan untuk mengestimasi parameter pada analisis peubah respons kontinu non negatif dengan regresi *inverse Gaussian* adalah metode *MLE (Maximum Likelihood Estimation)*. Langkah-langkah metode *MLE* adalah 1. menentukan fungsi kepadatan peluang distribusi *inverse Gaussian*, 2. menentukan fungsi *likelihood*, 3. Melogaritma-naturalkan fungsi *likelihood* atau disebut fungsi *log-likelihood* dan 4. menurunkan fungsi *log-likelihood* terhadap masing-masing parameter regresi lalu disamadenganka nol. Turunan terhadap fungsi *log-likelihood* ternyata tidak diperoleh estimator parameter regresi yang eksak, sehingga pengestimasian parameter-parameter tersebut harus dilakukan secara bersamaan. Permasalahan tersebut dapat diatasi dengan menggunakan bantuan program SAS 9.1.3 yaitu metode *MLE* yang diselesaikan dengan metode Newton-Raphson.

Contoh penerapan regresi *inverse Gaussian* adalah pada data inflasi di Indonesia periode 1980-2010 dengan peubah respons yaitu inflasi sedangkan peubah penjelasnya adalah bahan makanan, perumahan, sandang, kesehatan dan pendidikan. Histogram data inflasi menunjukkan bahwa peubah respons (inflasi) merupakan peubah kontinu yang memiliki nilai non negatif dan membentuk kurva yang menceng ke kanan sehingga regresi *inverse Gaussian* layak digunakan. Selanjutnya melalui uji kelayakan model dengan uji *goodness of fit* diperoleh kesimpulan bahwa regresi *inverse Gaussian* layak digunakan. Sedangkan pada uji signifikansi masing-masing parameter regresi dengan uji Wald menunjukkan bahwa peubah penjelas bahan makanan, perumahan dan pendidikan signifikan di dalam model, selain itu peubah penjelas sandang dan kesehatan tidak signifikan.

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR LAMPIRAN

BAB I

PENDAHULUAN

A. LATAR BELAKANG

Model linear pada analisis regresi telah diterapkan dalam bidang bisnis, ekonomi, pendidikan, kesehatan, dan biologi. Penerapan model linear pada bidang bisnis misalnya model yang bertujuan untuk menentukan hubungan antara penjualan dengan biaya promosi dan biaya produksi. Bidang ekonomi misalnya model yang bertujuan untuk menganalisis faktor-faktor yang mempengaruhi inflasi baik peubah domestik seperti suku bunga SBI (Sertifikat Bank Indonesia), dan produktivitas, maupun peubah internasional seperti nilai tukar dan inflasi luar negeri. Bidang pendidikan misalnya model yang bertujuan untuk mengetahui adakah pengaruh nilai Ujian Akhir Nasional matematika siswa SLTP dan cara belajar matematika terhadap prestasi belajar matematika siswa SLTP. Pada bidang kesehatan misalnya model untuk mengetahui pengaruh diet dengan berat badan seseorang. Sedangkan pada bidang biologi misalnya model untuk meneliti pengaruh hama terhadap hasil panen petani.

Seringkali analisis regresi bertujuan untuk mengetahui apakah ada hubungan antara dua atau lebih peubah dalam suatu model. Analisis regresi merupakan analisis yang sering digunakan dalam beberapa analisa data yang berkenaan dengan penggambaran hubungan antara peubah respons (Y)

dengan satu atau beberapa peubah penjelas (*X*). Pada persamaan regresi memiliki dua sifat yaitu linear dan non linear (Gujarati, 2006: 124). Pada sifat linear, maka kurva akan membentuk arah naik atau turun dengan garis lurus tergantung pada hubungan antara peubah respons dan peubah penjelas baik sederhana maupun berganda. Istilah linear dapat ditafsirkan dalam dua cara yang berbeda yaitu linear dalam peubah dan linear dalam parameter. Model regresi dikatakan linear dalam peubah penjelas jika peubah di dalam model berpangkat satu dan linear dalam parameter jika tidak ada parameter yang muncul sebagai suatu eksponen atau dikalikan atau dibagi oleh parameter lain. Sedangkan model regresi non linear diantaranya yaitu model kuadratik dan kubik dengan kurva membentuk garis lengkung. Analisis regresi mempunyai tiga kegunaan utama: (1) deskripsi, yaitu menjelaskan secara detail tentang hubungan peubah respons dengan beberapa peubah penjelas (2) kontrol atau kendali misalnya sebuah kantor yang membuka cabang kantor baru di daerah lain, maka yang menjadi pengendali adalah kantor utama dan (3) peramalan misalnya meramalkan besarnya kenaikan inflasi untuk lima tahun yang akan datang pada data yang sudah diketahui (Neter, 1997: 26). Persamaan matematik yang memungkinkan untuk meramalkan nilai-nilai suatu peubah respons dari nilai-nilai satu atau lebih peubah penjelas disebut persamaan regresi. Apabila peubah responsnya berupa peubah kontinu yang berdistribusi normal, maka digunakan model regresi linear klasik.

Asumsi pada regresi linear klasik adalah: (1) normalitas yaitu peubah respons berdistribusi normal, (2) kelinearan yaitu hubungan antara peubah respons dan peubah penjelas harus linear, (3) peubah penjelas X tidak berkorelasi dengan faktor gangguan acak ε, tetapi jika peubah X bersifat nonstokhastik (yaitu, nilainya merupakan bilangan yang telah ditetapkan sebelumya) maka asumsi ini otomatis terpenuhi dan (4) homoskedastisitas yaitu ragam residual dari satu pengamatan ke pengamatan yang lain adalah tetap (Gujarati, 2007:145).

Analisis regresi diterapkan dalam kehidupan sehari-hari. Sebagai contoh bidang ekonomi, peubah yang menjadi pengamatan seperti faktor-faktor yang mempengaruhi inflasi merupakan peubah respons kontinu non negatif serta memiliki tingkat ukuran kemencengan yang cenderung ke kanan, sehingga regresi linear klasik tidak tepat digunakan pada permasalahan tersebut. Kemencengan (*Skewness*/ *Sk*) bernilai positif menyebabkan bentuk kurva distribusi menceng ke kanan, yaitu distribusi terpusat disebagian besar sisi kiri dengan ekor tipis panjang yang terletak di sisi kanan. Sehingga data cenderung terkonsentrasi pada nilai yang rendah. Pada regresi linear klasik, peubah respons diasumsikan sebagai peubah kontinu yang berdistribusi normal sehingga nilai peubah respons terletak $-\infty < Y < \infty$. Dalam hal ini yaitu permasalahan tentang inflasi, peubah respons tidak mungkin bernilai negatif. Sehingga perlu adanya metode analisis regresi yang mampu menyelesaikan permasalahan yang dihadapi yaitu peubah respons kontinu yang non negatif.

Salah satu alternatif regresi yang dapat digunakan untuk mengatasi permasalahan peubah respons kontinu non negatif pada masalah data inflasi adalah model regresi *inverse Gaussian* (Jong & Heller, 2008: 29). Model regresi *inverse Gaussian* merupakan model regresi dengan peubah respons kontinu non negatif yang berdistribusi *inverse Gaussian*. Regresi *inverse Gaussian* dipilih dengan alasan bahwa memuat nilai peubah respons kontinu non negatif dan termasuk dalam distribusi keluarga eksponensial sehingga model linear terampat (*Generalized Linear Model/ GLM*) dapat digunakan untuk menyelesaikan permasalahan inflasi dengan peubah respons kontinu non negatif dan tidak memerlukan pemenuhan asumsi seperti pada model regresi linear klasik. Model *inverse Gaussian* tergolong *GLM* dengan mengasumsikan peubah respons $Y_i \sim IG\ (\mu_i, \sigma^2)$, $E(Y_i) = \boldsymbol{\mu_i} = \boldsymbol{X'\beta}$; dengan penghubung kanonik $g(\mu_i) = {\mu_i}^{-2}$. Biasanya, fungsi hubung log lebih sering digunakan (Jong & Heller, 2008: 125).

GLM dikenalkan oleh McCullagh & Nelder (1989: 19). Ada tiga komponen utama *GLM* yaitu: (1) komponen acak, yaitu komponen dari Y_i yang bebas dan peubah respons Y_i diasumsikan berdistribusi keluarga eksponensial, (2) komponen sistematik yaitu $X_1, X_2, \dots, X_k$ yang menghasilkan penduga linear $\eta = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k$ dan (3) mempunyai fungsi hubung $g(\mu) = \eta$. Metode yang sering digunakan untuk mengestimasi parameter dalam *GLM* adalah metode estimasi kemungkinan maksimum (*Maximum Likelihood Estimation/ MLE*). Metode *MLE* memanfaatkan distribusi peluang bersama (*joint probability distribution*) dari data sampel untuk menduga nilai parameter.

B. RUMUSAN MASALAH

Rumusan masalah berdasarkan latar belakang di atas adalah:

1. Bagaimana cara mengestimasi parameter pada analisis peubah respons kontinu non negatif dengan regresi *inverse Gaussian*?
2. Bagaimana contoh penerapan analisis peubah respons kontinu non negatif dengan regresi *inverse Gaussian*?

C. TUJUAN PENULISAN

Tujuan penulisan berdasarkan rumusan masalah di atas adalah:

1. Menjelaskan cara mengestimasi parameter pada analisis peubah respons kontinu non negatif dengan regresi *inverse Gaussian*.
2. Menerapkan analisis peubah respons kontinu non negatif dengan regresi *inverse Gaussian* untuk menyelesaikan suatu permasalahan.

D. MANFAAT PENULISAN

Manfaat penulisan tugas akhir ini yaitu:

1. Bagi penulis, menambah pengetahuan statistika terutama tentang penerapan model regresi dalam kehidupan sehari-hari.
2. Bagi mahasiswa matematika, dapat menambah pengetahuan terutama bidang pemodelan regresi dengan respons kontinu non negatif.
3. Bagi pembaca pada umumnya dapat menggunakan regresi *inverse Gaussian* untuk menganalisis peubah respons kontinu non negatif pada data-data penelitian.

BAB II

KAJIAN PUSTAKA

Pada kajian pustaka ini dibahas tentang peluang, peubah acak dan distribusi peluang kontinu, matriks, distribusi keluarga eksponensial, model linear terampat (*Generalized Linear Model*/ *GLM*), *Maximum Likelihood Estimation* (*MLE*), metode Newton-Raphson, regresi linear ganda, uji *Goodness of Fit* dan uji signifikansi koefisien regresi dengan uji Wald.

A. Peluang

Definisi 2.1 (Walpole & Myers, 1995: 2)

Ruang sampel adalah himpunan semua hasil yang mungkin dari suatu percobaan. Ruang sampel biasa dinyatakan dengan lambang *S*.

Definisi 2.2 (Walpole & Myers, 1995: 5)

Suatu kejadian adalah himpunan bagian dari ruang sampel.

Definisi 2.3 (Walpole & Myers, 1995: 6)

Irisan dua kejadian A dan B, dinyatakan dengan lambang $A \cap B$, ialah kejadian yang unsurnya termasuk dalam A dan B.

Definisi 2.4 (Walpole & Myers, 1995: 6)

Kejadian A dan B saling meniadakan atau terpisah bila $A \cap B = \emptyset$, yakni, bila A dan B tidak memiliki unsur persekutuan.

Definisi 2.5 (Walpole & Myers, 1995: 7)

Gabungan dua kejadian *A* dan *B*, dinyatakan dengan lambang $A \cup B$, ialah kejadian yang memiliki semua unsur yang termasuk A atau B atau keduanya.

Definisi 2.6 (Bain & Engelhardt, 1992: 9)

Jika sebuah percobaan mempunyai ruang sampel S dan A_1, A_2, A_3 ... mewakili kejadian *A*, maka *P (A)* adalah suatu bilangan real yang disebut peluang dari kejadian *A* atau peluang *A*, yang memiliki ketiga sifat berikut:

a. $0 \leq P(A) \leq 1$, untuk tiap kejadian *A* dari *S* (2.1)

b. $P(S) = 1$ (2.2)

c. Jika $A_1, A_2, A_3, \ldots$, barisan kejadian saling asing maka

$$P\left(\bigcup_{i=1}^{\sim} A_i\right) = \sum_{i=1}^{\sim} P(A_i) \tag{2.3}$$

Saling asing apabila $A_i \cap A_j = \emptyset$, untuk setiap pasang (i, j) dengan $i \neq j$.

B. Peubah Acak dan Distribusi Peluang Kontinu

Definisi 2.7 (Walpole & Myers, 1995: 51)

Peubah acak adalah suatu fungsi yang mengaitkan suatu bilangan real pada setiap unsur dalam ruang sampel.

Peubah acak biasanya dilambangkan dengan huruf kapital, misalnya *Y*, sedangkan nilainya dilambangkan dengan huruf kecil padanannya, misalnya *y*.

Definisi 2.8 (Bain & Engelhardt, 1992: 64)

Peubah acak Y disebut peubah acak kontinu jika daerah nilai dari Y adalah suatu interval.

Definisi 2.9 (Bain & Engelhardt, 1992: 65)

Fungsi $f(y)$ adalah fungsi kepadatan peluang peubah acak kontinu Y, bila memenuhi:

a. $f(y) \geq 0$ untuk semua $y \in R$ (2.4)

b. $\int_{-\sim}^{\sim} f(y)\, dy = 1$ (2.5)

c. $P(a \leq Y \leq b) = \int_{a}^{b} f(y)\, dy$ (2.6)

Definisi 2.10 (Bain & Engelhardt, 1992: 66)

Fungsi distribusi kumulatif $F(y)$ suatu peubah acak kontinu Y dengan fungsi kepadatan peluang $f(y)$ ditunjukkan oleh:

$$F(y) = P(Y \leq y) = \int_{-\sim}^{y} f(t)\, dt, \qquad -\sim < y < \sim \quad (2.7)$$

Definisi 2.11 (Bain & Engelhardt, 1992: 67)

Apabila Y suatu peubah acak kontinu dengan fungsi kepadatan peluang $f(y)$, dan apabila $\int_{-\sim}^{\sim} y\, f(y)\, dy$ konvergen mutlak, maka nilai harapan Y yang disimbolkan $E(Y)$ didefinisikan sebagai berikut:

$$E(Y) = \int_{-\sim}^{\sim} y\, f(y)\, dy \quad (2.8)$$

Nilai harapan dari Y dinyatakan juga dengan simbol μ_Y. Apabila hanya terdapat satu peubah acak, nilai harapan dapat disimbolkan μ.

Definisi 2.12 (Bain & Engelhardt, 1992: 73)

Ragam dari peubah acak Y adalah

$$Var\ (Y) = E[(Y-\mu)^2] \tag{2.9}$$

Definisi 2.13 (Bain & Engelhardt, 1992: 73)

Momen ke-k dari peubah acak Y, yang biasa dinyatakan dengan simbol μ_k' adalah bilangan yang ditentukan dengan rumus:

$$\mu_k' = E(Y^k) \tag{2.10}$$

Momen ke-k dari peubah acak Y, terhadap nilai harapan, yang dinyatakan dengan simbol μ_k adalah bilangan yang ditentukan dengan rumus

$$\mu_k = E\left[\left\{\left(Y - E(Y)\right)\right\}^k\right]$$

$$\mu_k = E[(Y-\mu)^k] \tag{2.11}$$

Teorema 2.1 (Bain & Engelhardt, 1992: 74)

Misalkan Y peubah acak maka $Var\ (Y) = E(Y^2) - \mu^2$.

Bukti:

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

$$= E(Y^2) - 2\mu \cdot \mu + \mu^2$$

$$= E(Y^2) - 2\mu^2 + \mu^2$$

$$Var\ (Y) = E(Y^2) - \mu^2 \text{ (Terbukti)}$$

Definisi 2.14 (Bain & Engelhardt, 1992: 78)

Yang disebut fungsi pembangkit momen (FPM) dari peubah acak Y adalah fungsi $M_Y(t)$ yang didefinisikan:

$$M_Y(t) = E(e^{tY})$$

untuk setiap *t* dalam interval *(-h, h)* untuk suatu konstanta positif *h*.

Misalkan Y suatu peubah acak kontinu, dengan fungsi kepadatan peluang $f(y)$, maka

$$M_Y(t) = \int_{-\sim}^{\sim} e^{tY} f(y) dy \qquad (2.12)$$

Perluasan e^{tY} sebagai suatu deret dalam *t* diperoleh:

$$e^{tY} = 1 + tY + \frac{t^2Y^2}{2!} + \cdots + \frac{t^rY^r}{r!} + \cdots$$

Untuk mendapatkan nilai harapan, dari persamaan:

$$M_Y(t) = E[e^{tY}] = 1 + E(Y) \cdot t + E(Y^2) \cdot \frac{t^2}{2!} + \cdots + E(Y^r) \cdot \frac{t^r}{r!} + \cdots$$

sehingga diperoleh:

$$M_Y(t) = 1 + M_1' \cdot t + M_2' \cdot \frac{t^2}{2!} + \cdots + M_r' \cdot \frac{t^r}{r!} + \cdots$$

Bila $M_Y(t)$ ditulis sebagai sebuah deret dalam *t*, maka koefisien dari $\frac{t^r}{r!}$ dalam perluasan adalah momen ke-r disekitar momen pusat. Satu cara untuk menggunakan fungsi pembangkit momen adalah:

a. $M_Y(t)$ diperoleh secara analitis untuk distribusi tertentu.

b. $M_Y(t)$ diperluas sebagai sebuah deret dalam t dan diperoleh koefisien dari $\frac{t^r}{r!}$ sebagai momen pusat ke-r.

Definisi 2.15 (Gujarati, 2007: 56)

Kemencengan (*Skewness/ Sk*) adalah derajat ketidaksimetrisan suatu bentuk kurva distribusi. Pada distribusi yang tidak simetris, data akan terkonsentrasi pada salah satu sisi kurva sehingga bentuk kurva yang diperoleh akan menceng.

Kemencengan suatu kurva dapat dilihat dari hubungan antara tiga nilai sentral. Nilai sentral adalah suatu nilai yang mewakili semua nilai pengamatan dalam suatu data. Nilai sentral dianggap sebagai gambaran dari kondisi suatu data. Beberapa nilai sentral yang umum digunakan adalah rata-rata hitung, median dan modus. Notasi yang dipakai untuk menunjukkan rata-hitung adalah $\bar{x}$. Pada statistika, rata-rata hitung pada sampel dinotasikan $\bar{x}$ sedangkan rata-rata hitung pada populasi dinotasikan μ. Median dinotasikan dengan *Md* sedangkan modus dinotasikan dengan *Mo*.

a. Hubungan Rata-rata Hitung, Median, dan Modus

Secara empiris, rata-rata hitung, median dan modus memiliki hubungan yang tergambar pada rumus berikut (Atmaja, 2009: 17):

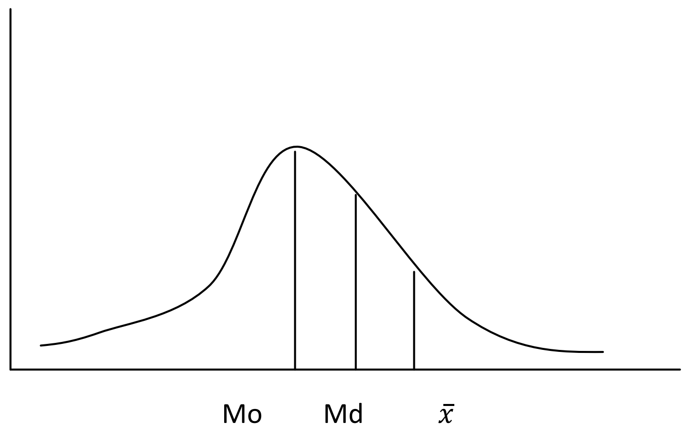
$$\bar{x} - Mo = 3(\bar{x} - Md) \quad (2.13)$$

dengan,

$\bar{x}$: rata-rata hitung
Md : median
Mo : modus

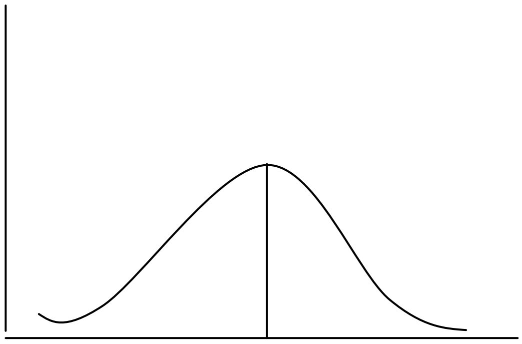
Pada kurva distribusi frekuensi, hubungan ketiga nilai sentral yaitu rata-rata hitung, median dan modus dapat digambarkan sebagai berikut:

i. Pada kurva yang menceng ke kanan ("ekor" kurva ada di sebelah kanan), $\bar{x} > Md > Mo$.



Gambar 1. Kurva menceng ke kanan

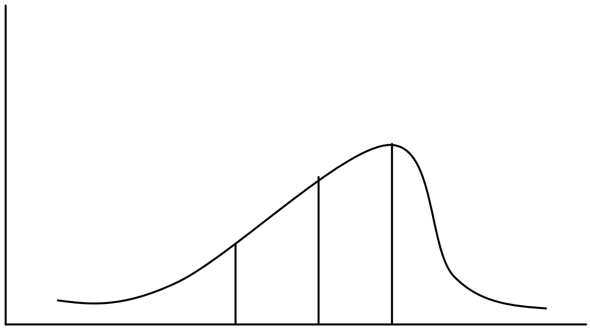
ii. Pada kurva yang simetris, $Mo = Md = \bar{x}$.



$\bar{x} = Mo = Md$

Gambar 2. Kurva Simetris

iii. Pada kurva yang menceng ke kiri, $Mo > Md > \bar{x}$.



Gambar 3. Kurva menceng ke kiri

Hubungan antara rata-rata hitung, median dan modus dapat dimanfaatkan untuk menghitung kemencengan suatu kurva distribusi frekuensi.

a. Pengukuran Kemencengan

i. Kemencengan Suatu Distribusi Frekuensi

Suatu distribusi dikatakan menceng ke kanan apabila sebagian besar nilai yang memiliki frekuensi rendah kebanyakan berada disebelah kanan nilai rata-rata, atau dikatakan "ekornya"nya menjulur ke kanan. Kurva seperti ini ditunjukkan pada Gambar 1.

Distribusi dapat berbentuk simetris, yang berarti luas kurva di sebelah kiri nilai rata-rata sama dengan luas kurva di sebelah kanan nilai rata-rata. Distribusi frekuensi yang simetris digambarkan pada Gambar 2.

Kurva distribusi dapat pula menceng ke kiri. Suatu distribusi frekuensi dikatakan menceng ke kiri jika nilai-nilai pengamatan yang berfrekuensi rendah lebih banyak berada disebelah kiri nilai rata-rata, atau "ekor"nya menjulur ke kiri. Gambar 3 menunjukkan distribusi frekuensi yang menceng ke kiri.

ii. Metode Pengukuran Kemencengan

Mengukur kemencengan suatu distribusi pada data populasi dapat digunakan rumus Koefisien Karl Pearson sebagai berikut (Atmaja, 2009: 26):

$$Sk = \frac{(\bar{x} - M_o)}{s} \tag{2.14}$$

dengan,

Sk : kemencengan
$\bar{x}$: rata-rata hitung
M_o : modus
s : standar deviasi

Rumus standar deviasi adalah sebagai berikut (Atmaja, 2009: 21):

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} \quad (2.15)$$

dengan,

x_i : nilai pengamatan ke-i
$\bar{x}$: rata-rata hitung
n : banyaknya pengamatan

Nilai kemencengan positif artinya distribusi frekuensi menceng ke kanan, nilai kemencengan negatif artinya distribusi frekuensi menceng ke kiri, dan nilai kemencengan sama dengan nol artinya distribusi frekuensi simetris.

Mencari nilai kemencengan dapat menggunakan rumus Karl Pearson sebagai berikut, dengan diketahui hubungan antara $\bar{x}, M_o, dan\ M_d$ pada persamaan (2.13) adalah:

$$\bar{x} - M_o = 3(\bar{x} - M_d)$$

$$M_o = \bar{x} - 3(\bar{x} - M_d)$$

$$Sk = \frac{\bar{x} - M_o}{s}$$

$$Sk = \frac{\bar{x} - [\bar{x} - 3(\bar{x} - M_d)]}{s}$$

$$= \frac{\bar{x} - \bar{x} + 3(\bar{x} - M_d)}{s}$$

$$Sk = \frac{3(\bar{x} - M_d)}{s} \quad (2.16)$$

Definisi 2.16 (Gujarati, 2007: 57)

Keruncingan (*Kurtosis/ K*) adalah suatu ukuran yang digunakan untuk menentukan runcing atau tidaknya suatu kurva distribusi sehingga dapat diketahui apakah kumpulan data terkonsentrasi disekitar rata-rata atau menyebar.

Telah diketahui bahwa momen pertama dari fungsi kepadatan peluang dari peubah acak Y diukur dengan $E(Y) = \mu_Y$, atau nilai harapan dari Y, sedangkan momen kedua yakni ragam diukur dengan $E(Y - \mu_Y)^2$. Dengan cara yang sama diperoleh momen keempat yakni $E(Y - \mu_Y)^4$. Salah satu pengukuran keruncingan menggunakan momen ke-4, diberikan oleh:

$$K = \frac{\mu_4}{{\mu_2}^2} \tag{2.17}$$

$$K = \frac{E(Y - \mu_Y)^4}{[E(Y - \mu_Y)^2]^2}$$

dengan,

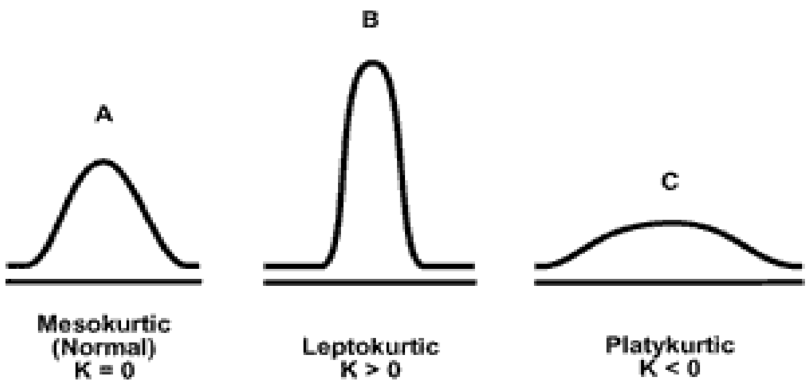
K: keruncingan
μ_i: momen pusat ke-i.

Ada tiga bentuk keruncingan kurva distribusi:

a. Leptokurtic $(K > 0)$, bentuk kurva distribusi ini menunjukkan data terkonsentrasi pada interval tertentu sekitar rata-rata. Keruncingan (puncaknya) relatif tinggi atau lebih tinggi dari keruncingan pada kurva distribusi normal.

b. Mesokurtic $(K = 0)$, bentuk kurva distribusinya simetris sehingga dianggap menggambarkan distribusi normal.

c. Platikurtic $(K < 0)$, bentuk kurva distribusi ini menunjukkan data tersebar ke seluruh daerah kurva. Keruncingannya (puncaknya) mendatar atau lebih rendah daripada puncak pada kurva distribusi normal.



Gambar 4. Keruncingan

C. Matriks

Matriks sering digunakan untuk menyingkat penulisan dari data multivariat. Menurut Hadley (1992: 51) matriks berordo $n \times k$ ialah sebanyak *nk* bilangan yang tersusun dalam *n* baris dan *k* kolom, yang lazim dinyatakan dengan notasi $A = (a_{ij})$; $i = 1,2, \dots, n$; $j = 1,2, \dots, k$; atau dengan notasi:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nk} \end{bmatrix} = a_{ij}, i = 1,2, \dots, n; j = 1,2, \dots, k.$$

a. Matriks Diagonal

Matriks diagonal menurut Hadley (1992: 64) adalah matriks persegi yang semua unsurnya yang di bawah diagonal utama dan yang di atas diagonal utama adalah nol. Tanda $\mathbf{D}(a_{ij})$ digunakan untuk menyatakan matriks diagonal yang unsur-unsur diagonal utamanya berturut-turut adalah $a_{11}, a_{22}, \dots, a_{nn}$. Matriks diagonal dinyatakan juga dengan tanda $\mathbf{diag}(a_{11}, a_{22}, \dots, a_{nn})$.

$$\text{Jadi, Diag}\left(a_{ij}\right) = \mathbf{diag}(a_{11}, a_{22}, \dots, a_{nn}) = \begin{bmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{nn} \end{bmatrix}$$

b. Determinan

Jika $\boldsymbol{A}$ adalah matriks kuadrat, maka minor entri (a_{ij}) dinyatakan oleh M_{ij} dan didefinisikan menjadi determinan dari submatriks yang tetap setelah baris ke-i dan kolom ke-j dicoret dari $\boldsymbol{A}$. Bilangan $(-1)^{i+j}M_{ij}$ dinyatakan dengan C_{ij} dan dinamakan kofaktor entri a_{ij}. Determinan matriks $\boldsymbol{A}$ yang berukuran $n \times n$ dapat dihitung dengan mengalikan entri-entri dalam suatu baris (atau kolom) dengan kofaktor-kofaktornya dan menambahkan hasil kali yang dihasilkan; yakni, untuk setiap $1 \leq i \leq n \; dan \; 1 \leq j \leq n$, maka

$$\det(A) = \sum_{i,j=1}^{n} a_{ij} \, C_{ij}$$

c. Invers Matriks

Jika **A** adalah matriks persegi, dan jika dapat dicari matriks **B** sehingga **AB** = **BA** = **I**, maka menurut Anton (1987: 34), matriks **A** mempunyai invers yaitu **B**. Jadi **B** dinamakan invers dari **A**, atau dapat ditulis $\boldsymbol{A^{-1}} = \boldsymbol{B}.$

Jika **A** matriks persegi dan C_{ij} adalah kofaktor a_{ij}; yakni bilangan $(-1)^{i+j}M_{ij}$, maka menurut Anton (1987: 81) matriks $\begin{bmatrix} C_{11} & C_{12} & \dots & C_{1n} \\ C_{21} & C_{22} & \dots & C_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{n1} & C_{n2} & \dots & C_{nn} \end{bmatrix}$ dinamakan matriks kofaktor $\boldsymbol{A}_{n\times n}$. Transpos matriks ini dinamakan Adjoin **A** dan dinyatakan adj **(A)**. Menurut Anton (1987: 82) jika matriks $\boldsymbol{A}_{n\times n}$ mempunyai invers, maka

$$\boldsymbol{A}^{-1} = \frac{1}{\det(\boldsymbol{A})} adj\ (\boldsymbol{A}).$$

d. Matriks Informasi

Misalkan θ merupakan vektor parameter berukuran p, $\theta = (\theta_1, \theta_2, \dots, \theta_p)$. Metode *MLE* merupakan salah satu metode untuk menentukan nilai estimasi parameter yang memberikan nilai maksimum fungsi *log-likelihood*. Pada metode *MLE*, untuk memperoleh fungsi *log-likelihood* $\ln L(\theta)$ maksimum, maka turunan parsial keduanya $\frac{\partial^2 L(\theta)}{\partial \theta^2}$ seharusnya bernilai negatif (Mc Cullagh & Nelder, 1989: 28). Misalkan $\boldsymbol{D}(\theta)$ adalah matriks turunan kedua fungsi *log-likelihood*, maka $\boldsymbol{D}(\theta)$ seharusnya juga bernilai negatif. Menurut Jong & Heller (2008: 69) $\boldsymbol{D}(\theta)$

inilah yang disebut matriks informasi. Jadi, matriks informasi adalah matriks turunan kedua fungsi *log-likelihood* yang bernilai negatif.

D. Distribusi Keluarga Eksponensial

Suatu peubah acak Y dengan fungsi kepadatan peluang $f(y)$ dan parameter θ dikatakan menjadi anggota distribusi keluarga eksponensial jika dapat dinyatakan dalam bentuk umum distribusi keluarga eksponensial. Bentuk umum distribusi keluarga eksponensial menurut Jong & Heller (2008: 35) adalah:

$$f(y|\theta,\phi) = c(y,\phi)exp\left(\frac{y\theta - b(\theta)}{a(\phi)}\right)$$

Suatu peubah acak Y dengan fungsi kepadatan peluang $f(y)$ dan parameter θ dikatakan menjadi anggota distribusi keluarga eksponensial jika dapat dinyatakan dalam bentuk umum distribusi keluarga eksponensial. Bentuk umum distribusi keluarga eksponensial pada *GLM* menurut Mc Cullagh & Nelder (1989: 28) adalah

$$f(y|\theta,\phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right\} \tag{2.18}$$

dengan θ dan ϕ adalah parameter kanonik dan parameter dispersi. Pemilihan fungsi $b(\theta)$ dan $c(y,\phi)$ yaitu dengan menetapkan fungsi tersebut sebagai fungsi peluang seperti normal, binomial, atau gama.

Nilai harapan $E(Y)$ dan ragam $Var(Y)$ fungsi $b(\theta)$ dari distribusi keluarga eksponensial dapat dicari dengan rumus:

$$E(Y) = \dot{b}(\theta) \tag{2.19}$$

$$Var(Y) = \phi\,\ddot{b}(\theta) \tag{2.20}$$

dengan $\dot{b}(\theta)$ dan $\ddot{b}(\theta)$ adalah turunan pertama dan kedua dari $b(\theta)$ terhadap θ.

a. Ragam

Ragam fungsi $b(\theta)$ pada distribusi keluarga eksponensial adalah sebagai berikut:

$$\ddot{b}(\theta) = \frac{\partial \dot{b}(\theta)}{\partial \theta} = \frac{\partial \mu}{\partial \theta} \equiv V(\mu) \tag{2.21}$$

dengan $\dot{b}(\theta)$ dan $\ddot{b}(\theta)$ adalah turunan pertama dan kedua dari $b(\theta)$ terhadap θ, μ adalah nilai tengah pada peubah penjelas. Sehingga dari persamaan (2.21) di atas dapat ditulis ragam $Var(Y) = \phi V(\mu)$.

b. Pembuktian Nilai Harapan dan Ragam

Untuk menunjukkan hubungan-hubungan pada persamaan (2.19) dan (2.20), ditetapkan $\dot{f}(y)$ dan $\ddot{f}(y)$ sebagai turunan pertama dan kedua dari $f(y)$ pada persamaan (2.18) berkenaan dengan θ. Kemudian

$$\dot{f}(y) = f(y)\left\{\frac{y - \dot{b}(\theta)}{\phi}\right\}, \ddot{f}(y) = f(y)\left\{\frac{y - \dot{b}(\theta)}{\phi}\right\}^2 - f(y)\frac{\ddot{b}(\theta)}{\phi}$$

Masing-masing kedua ruas diintegralkan berkenaan dengan y didapat

$$\int_{-\sim}^{\sim} \dot{f}(y)dy = \int_{-\sim}^{\sim} f(y)\left\{\frac{y-\dot{b}(\theta)}{\phi}\right\}dy$$

$$\frac{\partial}{\partial(\theta)}\int_{-\sim}^{\sim} f(y)\,dy = \int_{-\sim}^{\sim} \frac{y\,f(y)}{\phi}\,dy - \int_{-\sim}^{\sim} \frac{\dot{b}(\theta)f(y)}{\phi}\,dy \qquad (2.22)$$

$$\int_{-\sim}^{\sim} \ddot{f}(y)dy = \int_{-\sim}^{\sim}\left[f(y)\left\{\frac{y-\dot{b}(\theta)}{\phi}\right\}^2 - f(y)\frac{\ddot{b}(\theta)}{\phi}\right]dy$$

$$\frac{\partial^2}{\partial(\theta^2)}\int_{-\sim}^{\sim} f(y)\,dy = \int_{-\sim}^{\sim} f(y)\left\{\frac{y-\dot{b}(\theta)}{\phi}\right\}^2 dy - \int_{-\sim}^{\sim} f(y)\frac{\ddot{b}(\theta)}{\phi}dy \qquad (2.23)$$

Menurut persamaan (2.6) yaitu $\int_{-\sim}^{\sim} f(y)\,dy = 1$ dan persamaan (2.9) yaitu $E(Y) = \int_{-\sim}^{\sim} y\,f(y)\,dy$ maka persamaan (2.22) menjadi

$$\frac{\partial 1}{\partial(\theta)} = \frac{E(Y)}{\phi} - \frac{\dot{b}(\theta)}{\phi}$$

$$0 = \frac{E(Y)}{\phi} - \frac{\dot{b}(\theta)}{\phi}$$

$$\frac{E(Y)}{\phi} = \frac{\dot{b}(\theta)}{\phi}$$

$$E(Y) = \dot{b}(\theta).$$

Menurut persamaan (2.6) yaitu $\int_{-\sim}^{\sim} f(y)\,dy = 1$, persamaan (2.9) yaitu $\mu = E(Y) = \int_{-\sim}^{\sim} y\,f(y)\,dy$ dan (2.10) yaitu $Var(Y) = \int_{-\sim}^{\sim} (y-\mu)^2 f(y)\,dy$ maka persamaan (2.23) menjadi

$$\frac{\partial^2 1}{\partial(\theta^2)} = \frac{Var\ (Y)}{\phi^2} - \frac{\ddot{b}(\theta)}{\phi}$$

$$0 = \frac{Var\ (Y)}{\phi^2} - \frac{\ddot{b}(\theta)}{\phi}$$

$$\frac{Var\ (Y)}{\phi^2} = \frac{\ddot{b}(\theta)}{\phi}$$

$$Var\ (Y) = \frac{\ddot{b}(\theta)}{\phi}\phi^2$$

$$Var\ (Y) = \phi\ddot{b}(\theta).$$

Jadi, terbukti bahwa $E(Y) = \dot{b}(\theta)$ dan $Var\ (Y) = \phi\ddot{b}(\theta)$ yaitu sesuai dengan persamaan (2.19) dan (2.20).

Pada Tabel 1 akan ditampilkan distribusi keluarga eksponensial dengan parameternya.

Tabel 1. Distribusi keluarga eksponensial dan parameternya

(Jong & Heller, 2008: 36)

Distribusi	θ	$b(\theta)$	ϕ	$E(Y)$	$Var(\mu) = \frac{Var(Y)}{\phi}$
Binomial, $B(n,\pi)$	$ln\frac{\pi}{1-\pi}$	$n\ ln(1+e^{\theta})$	1	$n\pi$	$n\pi(1-\pi)$
Poisson, $P(\mu)$	$ln\ \mu$	e^{θ}	1	μ	μ
Normal, $N(\mu,\sigma^2)$	μ	$\frac{1}{2}\theta^2$	σ^2	μ	1
Gamma, $G(\mu,V)$	$-\frac{1}{\mu}$	$-\ln(-\theta)$	$\frac{1}{v}$	μ	μ^2
Inverse Gaussian, $IG(\mu,\sigma^2)$	$-\frac{1}{2\mu^2}$	$-\sqrt{-2\theta}$	σ^2	μ	μ^3
Negative Binomial, $NB(\mu,\kappa)$	$ln\frac{\kappa\mu}{1+\kappa\mu}$	$-\frac{1}{\kappa}\ln(1-\kappa e^{\theta})$	1	μ	$\mu(1+\kappa\mu)$

Keterangan:

θ : parameter kanonik pada peubah respons
ϕ : parameter dispersi pada peubah respons
$E(Y)$: nilai harapan (*Y*) pada peubah respons
$Var\ (Y)$: ragam (*Y*) pada peubah respons

E. Model Linear Terampat (*Generalized Linear Model*/ *GLM*)

GLM digunakan untuk menilai dan mengukur hubungan antara peubah respons dengan peubah penjelas. Menurut Jong & Heller (2008: 64), *GLM* merupakan perluasan dari proses pemodelan linear yang mengijinkan penentuan model dari suatu data dengan peubah acak tidak harus menyebar normal, asalkan sebaran tersebut termasuk dalam distribusi keluarga eksponensial, antara lain yaitu *Inverse Gaussian*, Gamma dan Poisson. Uji hipotesis yang diterapkan pada *GLM* tidak memerlukan asumsi kenormalan dari peubah responsnya ataupun kehomogenan ragam. Sehingga dengan demikian *GLM* tak hanya dapat digunakan pada peubah respons yang berdistribusi Normal tetapi juga untuk peubah respons yang berdistribusi selain Normal dan ragam yang tidak konstan atau homogen.

Model umum *GLM* dengan $Y_i \sim NID\ (X_i^T\beta, \sigma^2); g(\mu) = \eta_i = X_i^T\beta;\ \eta_i = \mu_i$ dan Y_i termasuk dalam distribusi keluarga eksponensial adalah

$$Y_i = X_i\beta + \varepsilon \tag{2.24}$$

dengan

Y_i: peubah respons
X_i: peubah penjelas
β: parameter
ε: galat

Misalkan terdapat vektor pengamatan *y* yang mempunyai sejumlah *n* komponen pengamatan yaitu $y_1, y_2, \dots, y_n$ yang menyebar normal dengan nilai harapan μ. Selanjutnya dalam kasus model linear klasik dapat dirumuskan (Mc Cullagh & Nelder, 1989: 26)

$$\mu = \sum_{j=1}^{k} x_j\beta_j$$

Jika *i* menunjukkan pengamatan dan *j* menunjukkan peubah, maka bentuk fungsi di atas menjadi:

$$\mu_i = \sum_{j=1}^{k} x_{ij}\beta_j\,; i = 1,2, \dots, n$$

dengan,

x_{ij}: nilai pengamatan ke-*i* peubah ke-*j*, untuk $x_{ij} = 1$
k : banyaknya peubah
β_j : parameter yang nilainya tidak diketahui dan harus ditaksir dari data

Jika nilai harapan μ dinyatakan dalam bentuk matriks maka:

$$\boldsymbol{\mu} = \boldsymbol{X\beta}$$

dengan,

$\boldsymbol{\mu}$: vektor nilai harapan berukuran $n \times 1$, dengan $\mu_i = E(Y|X = x_i), x_i = 1, x_{i1}, \dots, x_{ik}$
$\boldsymbol{X}$: matriks berukuran $n \times k$
$\boldsymbol{\beta}$: vektor parameter berukuran $k \times 1$.

Untuk memahami tentang *GLM*, perlu diketahui terlebih dahulu tentang bentuk dan fungsi hubung pada *GLM*.

1. Bentuk *GLM*

Misalkan Y adalah peubah respons, maka bentuk *GLM* (Jong & Heller, 2008: 64):

$$f(y) = c(y,\phi)exp\left(\frac{y\theta - b(\theta)}{a(\phi)}\right), \quad \boldsymbol{g}(\boldsymbol{\mu_i}) = \boldsymbol{X'\beta} \qquad (2.25)$$

Pada persamaan (2.25), $f(y)$ merupakan fungsi kepadatan peluang peubah respons pada distribusi keluarga eksponensial. Sedangkan persamaan kedua yaitu $\boldsymbol{g}(\boldsymbol{\mu_i})$ adalah fungsi hubung yang menggambarkan hubungan antara penduga linear terhadap nilai harapan μ_i.

Perbedaan antara model linear klasik dengan *GLM*.

i. Untuk komponen-komponen model linear klasik

   a. Komponen acak Y diasumsikan menyebar normal dengan ragam konstan dan $E(Y) = \mu$.

b. Komponen sistematis: misalkan $x_1, x_2, \dots, x_k$ adalah pengamatan sebanyak k yang menghasilkan suatu penduga linear (*linear predictor*) η yang diberikan dalam bentuk

$$\eta = \sum_{j=1}^{k} x_j \beta_j$$

c. *Link* atau penghubung digunakan untuk menghubungkan komponen acak dengan komponen sistematis yaitu $\eta = \mu$, jika *i* merupakan indeks untuk pengamatan maka fungsi hubung ditulis:

$$\eta_i = g(\,\mu_i)$$

ii. Untuk komponen-komponen dalam *GLM*, yaitu:

a. Komponen acak; peubah respons, $Y_1, Y_2, \dots, Y_n$ dengan nilai harapan $E(Y) = \mu_i$ diasumsikan merupakan bagian dari distribusi keluarga eksponensial.

b. Komponen sistematis; $X_1, X_2, \dots, X_k$ menghasilkan penduga linear η dengan $\eta = \beta_0 + X_1\beta_1 + \cdots + X_k\beta_k$. Sekumpulan parameter $\boldsymbol{\beta}_{(i\times 1)}$ dan peubah penjelas $\boldsymbol{X}_{(i\times 1)}$.

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_i \end{bmatrix} \qquad \boldsymbol{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_i \end{bmatrix} \qquad i = 1,2,\dots,n$$

c. Fungsi hubung monoton g sedemikian sehingga $g(\mu_i) = \boldsymbol{\eta} = \boldsymbol{X_i^T}\boldsymbol{\beta}$, dengan $\mu_i = E(Y_i)$. Fungsi hubung menentukan model yang menghubungkan $\mu_i = E(Y_i)$ dengan fungsi linear $\boldsymbol{X_i^T}\boldsymbol{\beta}$.

Suatu fungsi hubung disebut fungsi penghubung kanonik apabila $g(\mu_i) = \eta = \theta$, dengan θ adalah parameter kanonik dalam persamaan (2.25).

2. Fungsi Hubung

Menurut Mc Cullagh & Nelder (1989: 32), fungsi hubung adalah suatu fungsi yang menghubungkan fungsi penduga linear η dengan nilai harapan μ_i. Dalam model linear klasik, fungsi hubung bisa berupa fungsi yang identik atau kanonik. Suatu fungsi hubung dikatakan fungsi penghubung kanonik bila parameter kanoniknya sama dengan fungsi hubungnya, yaitu

$$\eta = \theta \qquad (2.26)$$

dengan θ adalah parameter kanonik. Berikut fungsi penghubung kanonik untuk beberapa distribusi.

Tabel 2. Fungsi hubung (Jong & Heller, 2008: 67)

Fungsi hubung	$g(\mu)$	Penghubung kanonik untuk distribusi
Identity	μ	Normal
Log	$\ln \mu$	Poisson
Power	μ^p	Gamma ($p = -1$) *Inverse Gaussian* ($p = -2$)
Square root	$\sqrt{\mu}$	
Logit	$\ln \frac{\mu}{1-\mu}$	Binomial

F. *Maximum Likelihood Estimation* (*MLE*)

Salah satu metode dalam estimasi parameter adalah *Maximum Likelihood Estimation* (*MLE*). Prinsip dari *MLE* adalah menemukan estimator $\hat{\theta}$ yang memaksimumkan fungsi *likelihood* dengan disamadengankan nol. Berikut ini akan diberikan pengertian mengenai fungsi *likelihood* dan *MLE*.

Definisi 2.17 (Bain & Engelhardt, 1992: 293):

Misalkan $Y_1, Y_2, \dots, Y_n$ sampel acak dengan fungsi kepadatan peluang $f(y_i; \theta)$, $i = 1,2, \dots, n$. Apabila L fungsi peluang bersama dari $Y_1, Y_2, \dots, Y_n$ yang dipandang sebagai fungsi dari θ maka: $L(\theta) = f(y_1; \theta) \cdot f(y_2; \theta) \cdot \dots \cdot f(y_n; \theta)$ disebut fungsi *likelihood*.

Definisi 2.18 (Bain & Engelhardt, 1992: 294):

Misalkan $L(\theta) = f(y_1; \theta) \cdot f(y_2; \theta) \cdot \dots \cdot f(y_n; \theta)$, $\theta \epsilon \Omega$, adalah fungsi kepadatan peluang pada $Y_1, Y_2, \dots, Y_n$. Untuk sebuah himpunan $(y_1, y_2, \dots, y_n)$ yang diberikan, sebuah nilai $\hat{\theta}$ pada Ω dengan $L(\theta)$, disebut *MLE* pada θ. $\hat{\theta}$ adalah θ yang memenuhi: $f\left(y_1, y_2, \dots, y_n; \hat{\theta}\ \right) = \max_{\theta \epsilon \Omega} f(y_1, y_2, \dots, y_n; \theta)$.

Ada beberapa hal yang harus diperhatikan untuk menentukan estimator parameter dengan metode *MLE*, yaitu:

a. Jika setiap himpunan hasil pengamatan $(y_1, y_2, \dots, y_n)$ bersesuaian dengan tepat satu nilai $\tau(\theta)$, maka cara memperoleh taksiran yaitu dengan menentukan suatu fungsi $\tau\left(\hat{\theta}\right) = l(y_1, y_2, \dots, y_n$, yang domainnya adalah himpunan dari semua himpunan hasil pengamatan sampel. Fungsi itu disebut estimator *MLE* untuk $\tau(\theta)$.

b. Jika Ω merupakan suatu interval, dan jika $L(\theta)$ diferensiabel dan mencapai maksimum disuatu nilai dalam Ω, maka *MLE* merupakan penyelesaian dari persamaan:

$$\frac{\partial}{\partial\theta}[L(\theta)] = 0 \; atau \; \frac{\partial}{\partial\theta}\ln[L(\theta)] = 0$$

Jika terdapat peubah acak Y_i dengan $i = 1,2,\dots,n$ yang saling bebas dengan fungsi kepadatan peluang $f(y;\theta)$ maka fungsi peluang bersama berbentuk sebagai berikut:

$$L(\theta) = f(y_1;\theta) \cdot f(y_2;\theta) \cdot \dots \cdot f(y_n;\theta)$$

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

Nilai parameter θ dapat diperoleh dengan metode memaksimumkan fungsi kepadatan bersama atau disebut *MLE*. Hal tersebut dilakukan dengan metode turunan pertama dari fungsi *likelihood*-nya terhadap setiap parameternya sama dengan nol. Namun biasanya sulit untuk mencari turunan fungsi *likelihood* sehingga yang dilakukan adalah menentukan nilai maksimum dari logaritma natural fungsi *likelihood* tersebut atau disebut dengan fungsi *log-likelihood*.

Fungsi *log-likelihood* merupakan fungsi kepadatan bersama yang diubah menjadi bentuk logaritma, tujuannya untuk mempermudah di dalam menaksir parameter. Fungsi *log-likelihood* dapat ditulis dalam bentuk:

$$\ln L(\theta) = \prod_{i=1}^{n} \ln f(y_i;\theta)$$

Langkah-langkah untuk menentukan estimator parameter dengan *MLE*:

1. Menentukan fungsi *likelihood*

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

2. Menentukan fungsi *log-likelihood*

$$\ln L(\theta) = \prod_{i=1}^{n} \ln f(y_i; \theta)$$

3. Jika $f\left(y_1, y_2, \dots, y_n; \hat{\theta}\right) = \max_{\theta \epsilon \Omega} f(y_1, y_2, \dots, y_n; \theta)$, maka untuk memperoleh nilai $\hat{\theta}$ tersebut yang memaksimumkan $L(\theta)$ harus diturunkan dengan langkah-langkah sebagai berikut:

   i. nilai $\hat{\theta}$ diperoleh dari turunan pertama dengan disamadenganka nol

$$\left.\frac{\partial}{\partial\theta} L(\theta)\right|_{\theta=\hat{\theta}} = 0 \qquad (2.27)$$

   ii. nilai $\hat{\theta}$ dikatakan memaksimumkan $L(\theta)$ jika

$$\left.\frac{\partial^2}{\partial\theta^2} L(\theta)\right|_{\theta=\hat{\theta}} < 0 \qquad (2.28)$$

Selain dengan cara memaksimumkan fungsi *likelihood*, nilai $\hat{\theta}$ juga dapat diperoleh dengan memaksimumkan *log-likelihood*, $\ln L(\theta)$. Karena dengan memaksimumkan fungsi *likelihood* juga akan memaksimumkan fungsi *log-likelihood*, $\ln L(\theta)$. Sebab $\ln L(\theta)$ merupakan fungsi yang monoton naik, maka untuk memperoleh $\hat{\theta}$ dengan memaksimumkan fungsi *log-likelihood* dapat dilakukan dengan langkah-langkah yang sama yaitu:

i. nilai $\hat{\theta}$ diperoleh dari turunan pertama dengan disamadenganankan nol:

$$\frac{\partial}{\partial\theta}\ln L(\theta)\bigg|_{\theta=\hat{\theta}} = 0 \tag{2.29}$$

ii. nilai $\hat{\theta}$ dikatakan memaksimumkan $\ln L(\theta)$ jika

$$\frac{\partial^2}{\partial\theta^2}\ln L(\theta)\bigg|_{\theta=\hat{\theta}} < 0 \tag{2.30}$$

Dalam penggunaannya, ln $L(\theta)$ lebih sering digunakan karena lebih mudah penggunaannya dibandingkan dengan $L(\theta)$.

4. Menyelesaikan fungsi *log-likelihood* yang diperoleh pada langkah 2 atau 3 dan mendapatkan $\hat{\theta}$ sebagai estimator *MLE*-nya.

G. Metode Newton-Raphson

Apabila langkah mengestimasi parameter menggunakan metode *MLE* menghasilkan fungsi *likelihood* yang nonlinear, maka penyelesaian fungsi tersebut untuk memperoleh nilai taksiran parameternya digunakan metode Newton Raphson (Jong & Heller, 2008: 69). Metode ini merupakan metode perhitungan yang iteratif, sehingga akan lebih mudah dikerjakan dengan bantuan komputer.

Metode Newton Raphson menurut Chapra & Canale (1988: 74) didasarkan pada deret Taylor sebagai berikut:

$$f(x_{i+1}) = f(x_i) + f'(x_i)(x_{i+1} - x_i) + \frac{f''(x_i)}{2!}(x_{i+1} - x_i)^2 + \cdots$$

$$+ \frac{f^{(n)}(x_i)}{n!}(x_{i+1} - x_i)^n$$

Fungsi *likelihood* dengan parameter θ dapat diselesaikan sehingga memperoleh nilai estimator $\hat{\theta}$ dengan menggunakan metode Newton Raphson. Rumus estimasi parameter $\hat{\theta}$ pada iterasi ke-*(t+1)* dalam proses iterasi $(t = 0,1,2, \ldots)$ adalah sebagai berikut:

$$\hat{\theta}_{(t+1)} = \hat{\theta}_{(t)} - \boldsymbol{D}(\theta_t)^{-1}\boldsymbol{d}(\theta_t) \tag{2.31}$$

dengan,

$\hat{\theta}_{(t+1)}$: estimasi parameter θ pada iterasi ke-$(t+1)$

$\hat{\theta}_{(\theta_t)}$: estimasi parameter θ pada iterasi ke-(t)

$\boldsymbol{d}(\theta_t)$: matriks turunan pertama fungsi *likelihood*, sehingga entri dari $\boldsymbol{d}(\theta_t)$ adalah $\frac{\partial L(\theta_t)}{\partial \theta_t}$

$\boldsymbol{D}(\theta_t)$: matriks turunan kedua fungsi *likelihood*, sehingga entri dari $\boldsymbol{D}(\theta_t)$ adalah $\frac{\partial^2 L(\theta_t)}{\partial (\theta_t)^2}$

Akan dibuktikan bahwa $\hat{\theta}_{(t+1)} = \hat{\theta}_{(t)} - \boldsymbol{D}(\theta_t)^{-1}\boldsymbol{d}(\theta_t)$.

Bukti:

Diketahui $\boldsymbol{d}(\theta_t) = \left(\frac{\partial L(\theta_t)}{\partial (\theta_t)}\right)$, dan $\boldsymbol{D}(\theta_t) = \left(\frac{\partial^2 L(\theta_t)}{\partial L(\theta_t)^2}\right)$

Misalkan $\boldsymbol{Q}(\theta_t)$ merupakan deret Taylor orde kedua, maka

$$\boldsymbol{Q}(\theta_t) = L(\theta_t) + \boldsymbol{d}(\theta_t)(\theta_{t+1} - \theta_t) + \frac{1}{2}\boldsymbol{D}(\theta_t)(\theta_{t+1} - \theta_t)^2$$

Agar nilai $\boldsymbol{Q}(\theta_t)$ maksimum, maka $\frac{\partial \boldsymbol{Q}(\theta_t)}{\partial (\theta_{t+1})} = 0$

$$\frac{\partial \boldsymbol{Q}(\theta_t)}{\partial(\theta_{t+1})} = \boldsymbol{d}(\theta_t) + \frac{1}{2}\boldsymbol{D}(\theta_t)\left(\hat{\theta}_{t+1} - \hat{\theta}_t\right)^2 = 0$$

$$\boldsymbol{d}(\theta_t) + \frac{1}{2}\boldsymbol{D}(\theta_t)\left(\hat{\theta}_{t+1} - \hat{\theta}_t\right)^2 = 0$$

$$\boldsymbol{d}(\theta_t) + \boldsymbol{D}(\theta_t)\left(\hat{\theta}_{t+1} - \hat{\theta}_t\right) = 0$$

$$\boldsymbol{d}(\theta_t) + \hat{\theta}_{t+1}\boldsymbol{D}(\theta_t) - \hat{\theta}_t\boldsymbol{D}(\theta_t) = 0$$

$$\hat{\theta}_{t+1}\boldsymbol{D}(\theta_t) = \hat{\theta}_t\boldsymbol{D}(\theta_t) - \boldsymbol{d}(\theta_t)$$

$$\hat{\theta}_{t+1} = \hat{\theta}_t - \frac{\boldsymbol{d}(\theta_t)}{\boldsymbol{D}(\theta_t)}$$

$$\hat{\theta}_{t+1} = \hat{\theta}_t - \boldsymbol{D}(\theta_t)^{-1}\boldsymbol{d}(\theta_t)$$

Terbukti.

Menurut Chapra & Canale (1988: 134) proses iterasi dengan menggunakan metode Newton-Raphson terus dilakukan hingga didapatkan nilai $\hat{\theta}$ yang konvergen, yaitu sampai $\left|\frac{\hat{\theta}_{t+1}-\hat{\theta}_t}{\hat{\theta}_t}\right| < \delta$, dengan δ bilangan yang kecil, $\delta > 0$.

H. Regresi Linear Ganda

Regresi linear ganda digunakan untuk memodelkan hubungan linear antara peubah respons dengan dua atau lebih peubah penjelas. Model regresi linear ganda yang melibatkan *k* peubah penjelas, (Jong & Heller, 2008: 44) adalah:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + \varepsilon_i \qquad (2.32)$$

dengan,

$\beta_0, \beta_1, \beta_2, \dots, \beta_\mathrm{k}$: parameter
Y_i : peubah respons pada pengamatan ke-*i*
$X_{\mathrm{i}1}, X_{\mathrm{i}2}, \dots, X_{ik}$: peubah penjelas pada pengamatan ke-*i*
ε_i : galat pada pengamatan ke-*i*
$\varepsilon_i \sim N(0, \sigma^2)$

Persamaan (2.32) dapat ditulis kembali dalam notasi matriks yang tepat menjadi:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{2.33}$$

atau

$$\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix}, \boldsymbol{X} = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1k} \\ 1 & X_{21} & X_{22} & \cdots & X_{2k} \\ 1 & X_{31} & X_{32} & \cdots & X_{3k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{nk} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}, \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

dengan,

$\boldsymbol{Y}$: vektor kolom dari peubah respons berukuran $n \times 1$
$\boldsymbol{X}$: matriks dari peubah penjelas berukuran $n \times (k+1)$, yang bersifat tetap
$\boldsymbol{\beta}$: vektor parameter regresi berukuran $k \times 1$
$\boldsymbol{\varepsilon}$: galat
$\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \boldsymbol{\sigma^2 I})$

Untuk mencari nilai estimator parameter pada model regresi linear ganda dengan menggunakan metode *MLE* dilakukan langkah-langkah sebagai berikut:

1. Menentukan nilai harapan dan ragam dari peubah respons

   $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2)$

   $Y_i \sim N(\mu_i, \sigma^2)$

i. Nilai harapan $[E(Y_i)]$

$E(Y_i) = E(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + \varepsilon_i)$

$\quad = E(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik}) + E(\varepsilon_i)$

Karena X_{ik} merupakan peubah yang tetap, maka $\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik}$ adalah suatu konstanta, sehingga:

$E(Y_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik}$

$E(Y_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} = \mu_i$

ii. Ragam $[Var(Y_i)]$

$Var(Y_i) = Var(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + \varepsilon_i)$

$\quad = Var(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik}) + Var(\varepsilon_i)$

Karena X_{ik} merupakan peubah yang tetap, maka $\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik}$ adalah suatu konstanta dan diasumsikan suku galat ε_i mempunyai ragam yang sama dengan σ^2, maka:

$Var(Y_i) = Var(\varepsilon_i) = \sigma^2$

2. Menentukan fungsi kepadatan peluang peubah respons

Misalkan $(Y_1, Y_2, \ldots, Y_n)$ adalah sampel acak berukuran *n* dari berdistribusi normal, $Y_i \sim N(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik}, \sigma^2)$ maka fungsi kepadatan peluang dari Y_i adalah

$$f(y_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y_i - \mu_i}{\sigma}\right)^2}$$

$$f(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{y_i - (\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik})}{\sigma}\right)^2}$$

3. Menentukan fungsi *likelihood* $(L = L(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik}, \sigma^2))$

$$L(\boldsymbol{\beta}) = f(y_1) \cdot f(y_2) \cdot \ldots f(y_n)$$

$$= \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{\left[-\frac{1}{2\sigma^2}(y_i-\beta_0-\beta_1X_{i1}-\beta_2X_{i2}-\cdots-\beta_kX_{ik})^2\right]}$$

$$= \frac{1}{\left(\sigma\sqrt{2\pi}\right)^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\beta_0-\beta_1X_{i1}-\beta_2X_{i2}-\cdots-\beta_kX_{ik})^2}$$

$$L(\boldsymbol{\beta}) = e^{\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\beta_0-\beta_1X_{i1}-\beta_2X_{i2}-\cdots-\beta_kX_{ik})^2\right\}} \frac{1}{\left(\sigma\sqrt{2\pi}\right)^n}$$

4. Melogaritma-naturalkan fungsi *likelihood* atau disebut dengan fungsi *log-likelihood*

$$\ln L(\boldsymbol{\beta}) = -\frac{n}{2}(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1X_{i1} - \beta_2X_{i2} - \cdots - \beta_kX_{ik})^2$$

5. Menurunkan fungsi *log-likelihood* dengan disamadengankan nol terhadap parameter

   Prinsip dari metode *MLE* adalah mancari nilai estimator β dengan memaksimumkan fungsi *likelihood*. Untuk itu, agar lebih mudah dengan melogaritma-naturalkan yaitu pada langkah 4 atau disebut fungsi *log-likelihood*, kemudian menurunkan fungsi *log-likelihood* tersebut terhadap masing-masing parameter, yaitu $\beta_0, \beta_1, \ldots, \beta_k$ (Bain & Engelhardt, 1992: 507).

   i. Mencari estimator β_0

$$\frac{\partial \ln L(\boldsymbol{\beta})}{\partial\, \beta_0} = -\frac{1}{2\sigma^2}\sum_{i=1}^{n} 2(-y_i + \beta_0 + \beta_1X_{i1} + \beta_2X_{i2} + \cdots + \beta_{k-1}X_{ik-1}) = 0$$

$$-\frac{1}{\sigma^2}\sum_{i=1}^{n}\left(-y_i+\hat{\beta}_0+\hat{\beta}_1X_{i1}+\hat{\beta}_2X_{i2}+\cdots+\hat{\beta}_{k-1}X_{ik-1}\right)=0$$

$$-\sum_{i=1}^{n}\left(-y_i+\hat{\beta}_0+\hat{\beta}_1X_{i1}+\hat{\beta}_2X_{i2}+\cdots+\hat{\beta}_{k-1}X_{ik-1}\right)=0$$

$$\sum_{i=1}^{n}y_i-\sum_{i=1}^{n}\hat{\beta}_0-\hat{\beta}_1\sum_{i=1}^{n}X_{i1}-\hat{\beta}_2\sum_{i=1}^{n}X_{i2}-\cdots-\hat{\beta}_{k-1}\sum_{i=1}^{n}X_{ik-1}=0$$

$$\sum_{i=1}^{n}\hat{\beta}_0+\hat{\beta}_1\sum_{i=1}^{n}X_{i1}+\hat{\beta}_2\sum_{i=1}^{n}X_{i2}+\cdots+\hat{\beta}_{k-1}\sum_{i=1}^{n}X_{ik-1}=\sum_{i=1}^{n}y_i$$

$$n\hat{\beta}_0+\hat{\beta}_1\sum_{i=1}^{n}X_{i1}+\hat{\beta}_2\sum_{i=1}^{n}X_{i2}+\cdots+\hat{\beta}_{k-1}\sum_{i=1}^{n}X_{ik-1}=\sum_{i=1}^{n}y_i$$

ii. Mencari estimator β_1

$$\frac{\partial\ln L(\boldsymbol{\beta})}{\partial\,\beta_1}=-\frac{1}{2\sigma^2}\sum_{i=1}^{n}2(-X_{i1}y_i+\beta_0X_{i1}+\beta_1{X_{i1}}^2+\beta_2X_{i1}X_{i2}+\cdots$$
$$+\beta_{k-1}X_{i1}X_{ik-1})=0$$

$$-\frac{1}{\sigma^2}\sum_{i=1}^{n}\left(-X_{i1}y_i+\hat{\beta}_0X_{i1}+\hat{\beta}_1{X_{i1}}^2+\hat{\beta}_2X_{i1}X_{i2}+\cdots+\hat{\beta}_{k-1}X_{i1}X_{ik-1}\right)=0$$

$$-\sum_{i=1}^{n}\left(-X_{i1}y_i+\hat{\beta}_0X_{i1}+\hat{\beta}_1{X_{i1}}^2+\hat{\beta}_2X_{i1}X_{i2}+\cdots+\hat{\beta}_{k-1}X_{i1}X_{ik-1}\right)=0$$

$$\sum_{i=1}^{n}\left(X_{i1}y_i-\hat{\beta}_0X_{i1}-\hat{\beta}_1{X_{i1}}^2-\hat{\beta}_2X_{i1}X_{i2}-\cdots-\hat{\beta}_{k-1}X_{i1}X_{ik-1}\right)=0$$

$$\sum_{i=1}^{n}\left(\hat{\beta}_0X_{i1}+\hat{\beta}_1{X_{i1}}^2+\hat{\beta}_2X_{i1}X_{i2}+\cdots+\hat{\beta}_{k-1}X_{i1}X_{ik-1}\right)=\sum_{i=1}^{n}X_{i1}y_i$$

$$\hat{\beta}_0 \sum_{i=1}^{n} X_{i1} + \hat{\beta}_1 \sum_{i=1}^{n} X_{i1}{}^2 + \hat{\beta}_2 \sum_{i=1}^{n} X_{i1}X_{i2} + \cdots + \hat{\beta}_{k-1} \sum_{i=1}^{n} X_{i1}X_{ik-1} = \sum_{i=1}^{n} X_{i1}y_i$$

iii. Mencari estimator β_2

$$\frac{\partial \ln L(\boldsymbol{\beta})}{\partial \beta_2} = -\frac{1}{2\sigma^2}\sum_{i=1}^{n} 2\big(-X_{i2}y_i + \beta_0 X_{i2} + \beta_1 X_{i1}X_{i2} + \beta_2 X_{i2}{}^2 + \cdots$$

$$+ \beta_{k-1}X_{i2}X_{ik-1}\big) = 0$$

$$-\frac{1}{\sigma^2}\sum_{i=1}^{n}\big(-X_{i2}y_i + \hat{\beta}_0 X_{i2} + \hat{\beta}_1 X_{i1}X_{i2} + \hat{\beta}_2 X_{i2}{}^2 + \cdots + \hat{\beta}_{k-1}X_{i2}X_{ik-1}\big) = 0$$

$$-\sum_{i=1}^{n}\big(-X_{i2}y_i + \hat{\beta}_0 X_{i2} + \hat{\beta}_1 X_{i1}X_{i2} + \hat{\beta}_2 X_{i2}{}^2 + \cdots + \hat{\beta}_{k-1}X_{i2}X_{ik-1}\big) = 0$$

$$\sum_{i=1}^{n}\big(X_{i2}y_i - \hat{\beta}_0 X_{i2} - \hat{\beta}_1 X_{i1}X_{i2} - \hat{\beta}_2 X_{i2}{}^2 - \cdots - \hat{\beta}_{k-1}X_{i2}X_{ik-1}\big) = 0$$

$$\sum_{i=1}^{n}\big(\hat{\beta}_0 X_{i2} + \hat{\beta}_1 X_{i1}X_{i2} + \hat{\beta}_2 X_{i2}{}^2 + \cdots + \hat{\beta}_{k-1}X_{i2}X_{ik-1}\big) = \sum_{i=1}^{n} X_{i2}y_i$$

$$\hat{\beta}_0 \sum_{i=1}^{n} X_{i2} + \hat{\beta}_1 \sum_{i=1}^{n} X_{i1}X_{i2} + \hat{\beta}_2 \sum_{i=1}^{n} X_{i2}{}^2 + \cdots + \hat{\beta}_{k-1} \sum_{i=1}^{n} X_{i2}X_{ik-1} = \sum_{i=1}^{n} X_{i2}y_i$$

iv. Mencari estimator β_k

$$\frac{\partial \ln L(\boldsymbol{\beta})}{\partial \beta_k} = -\frac{1}{2\sigma^2}\sum_{i=1}^{n} 2(-X_{ik-1}y_i + \beta_0 X_{ik-1} + \beta_1 X_{i1}X_{ik-1} + \beta_2 X_{i2}X_{ik-1}$$

$$+ \cdots + \beta_k X_{ik-1}^2) = 0$$

$$-\frac{1}{\sigma^2}\sum_{i=1}^{n}\big(-X_{ik-1}y_i + \hat{\beta}_0 X_{ik-1} + \hat{\beta}_1 X_{i1}X_{ik-1} + \hat{\beta}_2 X_{i2}X_{ik-1} + \cdots + \hat{\beta}_k X_{ik-1}^2\big) = 0$$

$$-\sum_{i=1}^{n}\big(-X_{ik-1}y_i + \hat{\beta}_0 X_{ik-1} + \hat{\beta}_1 X_{i1}X_{ik-1} + \hat{\beta}_2 X_{i2}X_{ik-1} + \cdots + \hat{\beta}_k X_{ik-1}^2\big) = 0$$

$$\sum_{i=1}^{n}\left(X_{ik-1}y_i - \hat{\beta}_0 X_{ik-1} - \hat{\beta}_1 X_{i1}X_{ik-1} - \hat{\beta}_2 X_{i2}X_{ik-1} - \cdots - \hat{\beta}_k X_{ik-1}^2\right) = 0$$

$$\sum_{i=1}^{n}\left(\hat{\beta}_0 X_{ik-1} + \hat{\beta}_1 X_{i1}X_{ik-1} + \hat{\beta}_2 X_{i2}X_{ik-1} + \cdots + \hat{\beta}_k X_{ik-1}^2\right) = \sum_{i=1}^{n} X_{ik-1}y_i$$

$$\hat{\beta}_0 \sum_{i=1}^{n} X_{ik-1} + \hat{\beta}_1 \sum_{i=1}^{n} X_{i1}X_{ik-1} + \hat{\beta}_2 \sum_{i=1}^{n} X_{i2}X_{ik-1} + \cdots + \hat{\beta}_k \sum_{i=1}^{n} X_{ik-1}^2 = \sum_{i=1}^{n} X_{ik-1}y_i$$

Setelah diperoleh estimator pada model regresi, kemudian dilakukan uji kelayakan model dengan uji *goodness of fit*.

I. Uji *Goodness of Fit*

Uji *goodness of fit* digunakan untuk mengetahui apakah model layak atau tidak layak digunakan. Langkah-langkah pengujian dengan uji *goodness of fit* (Jong & Heller, 2008:71):

1. Merumuskan hipotesis

   H_0: Model layak digunakan.

   H_1: Model tidak layak digunakan.

2. Memilih taraf signifikansi: α
3. Menentukan statistik uji

   Statistik uji yang digunakan adalah *Deviance* (lihat Tabel 3)

4. Kriteria Keputusan: H_0 ditolak jika *Deviance* $> \chi^2_{\alpha(n-k)}$

   n: banyaknya pengamatan, *k*: banyaknya parameter.

5. Perhitungan

6. Kesimpulan: Jika nilai *deviance* $\leq \chi^2_{\alpha(n-k)}$, maka dapat disimpulkan bahwa model layak digunakan.

Tabel 3. *Deviance* untuk respons distribusi Keluarga Eksponensial (Jong & Heller, 2008:73)

Distribusi	*Deviance* Δ
Normal	$\frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i-\hat{\mu}_i)^2$
Poisson	$2\sum_{i=1}^{n}\left\{y_i \ln\left(\frac{y_i}{\hat{\mu}_i}\right)-(y_i-\hat{\mu}_i)\right\}$
Binomial	$2\sum_{i=1}^{n}n_i\left\{y_i \ln\left(\frac{y_i}{\hat{\mu}_i}\right)+(n_i-y_i)\ln\left(\frac{n_i-y_i}{1-\hat{\mu}_i}\right)\right\}$
Gamma	$2v\sum_{i=1}^{n}\left\{-\ln\left(\frac{y_i}{\hat{\mu}_i}\right)+\frac{y_i-\hat{\mu}_i}{\hat{\mu}_i}\right\}$
Inverse Gaussian	$\frac{1}{\sigma^2}\sum_{i=1}^{n}\frac{(y_i-\hat{\mu}_i)^2}{\hat{\mu}_i^2 y_i}$
Negatif Binomial	$2\sum_{i=1}^{n}\left\{y_i \ln\left(\frac{y_i}{\hat{\mu}_i}\right)-\left(y_i+\frac{1}{\kappa}\right)\ln\left(\frac{y_i+\frac{1}{\kappa}}{\hat{\mu}_i+\frac{1}{\kappa}}\right)\right\}$

J. Uji Signifikansi Koefisien Regresi dengan Uji Wald

Setelah dilakukan uji dengan *deviance* dilakukan uji signifikansi dari peubah penjelas, dengan melihat apakah terdapat peubah penjelas yang tidak signifikan di dalam model. Jika terdapat peubah penjelas yang tidak signifikan, perlu dilakukan reduksi terhadap peubah penjelas tersebut. Statistik uji yang digunakan untuk menyelidiki tingkat signifikansi dari peubah penjelas yaitu dengan uji Wald. Uji Wald akan membandingkan estimator *likelihood* parameter β_j dengan estimator standar errornya. Peubah penjelas yang dikeluarkan terlebih dahulu yaitu peubah yang tidak signifikan yang memiliki nilai *p-value* paling besar, lalu dibuat model baru tanpa memasukkan peubah yang direduksi tersebut. Langkah-langkah uji Wald adalah sebagai berikut (Agresti, 2007: 11):

1. Merumuskan Hipotesis

   $H_0: \beta_j = 0$

   $H_1: \beta_j \neq 0$

2. Memilih tingkat signifikansi: α
3. Menentukan statistik uji

   Statistik uji yang digunakan adalah: uji Wald

$$W = \left\{\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}\right\}^2 \sim \chi^2_1, j = 0,1,2,3,\dots,k$$

4. Kriteria Keputusan:

   H_0 ditolak jika $W < \chi^2_{\left(1-\frac{\alpha}{2}\right)(1)}$ atau $W > \chi^2_{\frac{\alpha}{2}(1)}$.

   Atau H_0 ditolak jika *p-value* $< \alpha$.

5. Perhitungan
6. Kesimpulan: Jika nilai $W > \chi^2_{\left(1-\frac{\alpha}{2}\right)(1)}$ dan $W < \chi^2_{\frac{\alpha}{2}(1)}$, maka H_0 diterima sehingga dapat disimpulkan bahwa peubah penjelas tidak signifikan di dalam model yang berarti bahwa peubah penjelas dikeluarkan dari model.

BAB III

PEMBAHASAN

Dalam teori peluang, distribusi *inverse Gaussian* (juga dikenal dengan distribusi *Wald*) adalah salah satu keluarga eksponensial dua parameter distribusi peluang kontinu dengan nilai peubah pada interval $(0, \infty)$.

Model regresi *inverse Gaussian* termasuk dalam *GLM* dengan peubah respons berdistribusi *inverse Gaussian*. Komponen utama model regresi *inverse Gaussian* adalah: (1) komponen acak, yaitu komponen dari Y_i yang bebas dan fungsi distribusi peluang Y_i termasuk dalam distribusi *inverse Gaussian* dengan $E(Y_i) = \boldsymbol{\mu_i} = \boldsymbol{X}'\boldsymbol{\beta}$; (2) komponen sistematik, yaitu $X_1, X_2, \dots, X_k$ yang menghasilkan penduga linear η dengan $\eta = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k$; dan (3) fungsi hubung $g(\mu_i) = {\mu_i}^{-2}$, menggambarkan hubungan antara penduga linear η dengan nilai harapan μ_i. Biasanya, fungsi hubung log lebih sering digunakan pada model regresi *inverse Gaussian* (Jong & Heller, 2008: 125).

Metode yang digunakan untuk memperoleh nilai estimasi parameter pada regresi *inverse Gaussian* adalah metode *MLE*. Pada penulisan ini yang dibahas adalah cara mengestimasi parameter pada analisis peubah respons kontinu non negatif dengan regresi *inverse Gaussian* menggunakan metode *MLE* dan contoh penerapannya.

A. Estimasi Parameter pada Analisis Peubah Respons Kontinu Non Negatif dengan Regresi *Inverse Gaussian*

Estimasi parameter pada analisis peubah respons kontinu non negatif dengan regresi *inverse Gaussian* dapat dicari dengan menggunakan metode *MLE*. Untuk mencari nilai estimasi parameter dengan *MLE*, perlu diketahui terlebih dahulu fungsi kepadatan peluang regresi *inverse Gaussian*.

1. Fungsi Kepadatan Peluang Regresi *Inverse Gaussian*

Peubah respons Y_i pada model regresi *inverse Gaussian* merupakan peubah non negatif dengan fungsi kepadatan peluang sebagai berikut $Y_i \sim IG(\mu_i, \sigma^2)$, (Jong & Heller, 2008: 29):

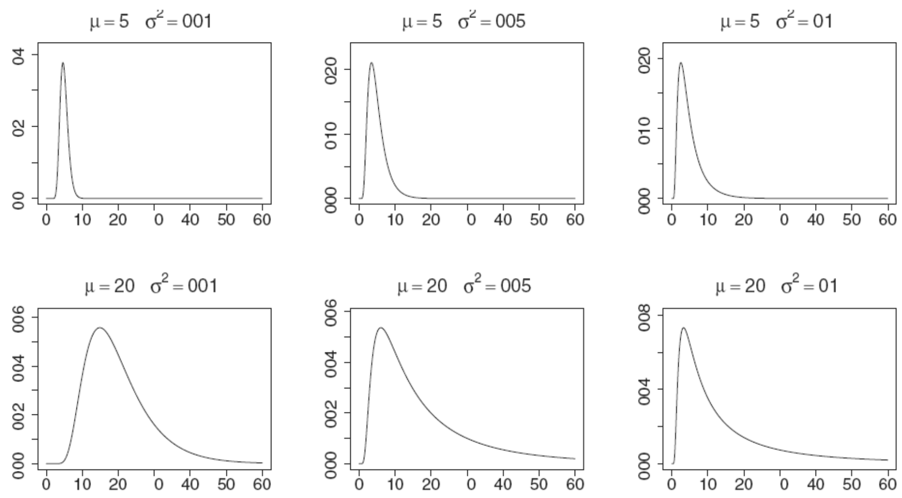
$$f(y_i) = \frac{1}{\sigma\sqrt{2\pi y_i{}^3}} \exp\left\{-\frac{1}{2y_i}\left(\frac{y_i - \mu_i}{\mu_i \sigma}\right)^2\right\}, y_i > 0, \sigma > 0, \mu_i > 0 \quad (3.1)$$

dengan $i = 1,2, \dots, n$.

Keterangan:

Y_i : peubah acak pada pengamatan ke-i
$f(y_i)$: fungsi kepadatan peluang distribusi *inverse Gaussian*
μ_i, σ : parameter pada peubah respons

Dalam hal distribusi *inverse Gaussian*, μ_i dan σ adalah kedua parameter distribusi *inverse Gaussian* dan $\exp(a)$ berarti e^a. Regresi *inverse Gaussian* memuat nilai peubah respons kontinu non negatif dan merupakan salah satu distribusi yang termasuk ke dalam distribusi keluarga eksponensial. Berikut diberikan contoh gambar fungsi kepadatan peluang distribusi *inverse Gaussian*:



Gambar 5. Fungsi kepadatan peluang distribusi *inverse Gaussian*

(Jong & Heller, 2008: 30)

Nilai estimasi regresi *inverse Gaussian* dapat dicari dengan mudah jika fungsi kepadatan peluang *inverse Gaussian* terlebih dahulu diubah ke dalam bentuk umum distribusi keluarga eksponensial yaitu seperti pada persamaan (2.18). Bentuk khusus fungsi kepadatan peluang dari peubah acak Y_i yang berdistribusi *inverse Gaussian* yaitu sebagai berikut:

$$f(y_i) = \frac{1}{\sigma\sqrt{2\pi {y_i}^3}} \exp\left\{-\frac{1}{2y_i}\left(\frac{y_i-\mu_i}{\mu_i\sigma}\right)^2\right\}$$

$$= \frac{(2\pi y^3)^{-\frac{1}{2}}}{\sigma} \exp\left\{-\frac{1}{2y_i}\left(\frac{{y_i}^2-2\mu_i y_i+{\mu_i}^2}{{\mu_i}^2\sigma^2}\right)\right\}$$

$$= \left\{-\frac{1}{2}\ln(2\pi y^3) - \ln\sigma\right\} \exp\left\{-\frac{1}{2y_i}\left(\frac{{y_i}^2-2\mu_i y_i+{\mu_i}^2}{{\mu_i}^2\sigma^2}\right)\right\}$$

$$= \left\{-\frac{1}{2}\ln(2\pi y^3) - \ln\sigma\right\} \exp\left\{-\frac{{y_i}^2}{2y_i{\mu_i}^2\sigma^2}+\frac{2\mu_i y_i}{2y_i{\mu_i}^2\sigma^2}-\frac{{\mu_i}^2}{2y_i{\mu_i}^2\sigma^2}\right\}$$

$$= \left\{-\frac{1}{2}\ln(2\pi y^3) - \ln\sigma\right\}\exp\left\{-\frac{y_i}{2{\mu_i}^2\sigma^2} + \frac{1}{\mu_i\sigma^2} - \frac{1}{2y_i\sigma^2}\right\}$$

$$= \exp\left\{-\frac{y_i}{2{\mu_i}^2\sigma^2} + \frac{1}{\mu_i\sigma^2} - \frac{1}{2y_i\sigma^2} - \frac{1}{2}\ln(2\pi y^3) - \ln\sigma\right\}$$

$$= \exp\left\{\frac{-\frac{y_i}{2{\mu_i}^2} + \frac{1}{\mu_i}}{\sigma^2} - \frac{1}{2y_i\sigma^2} - \frac{1}{2}\ln(2\pi y^3) - \ln\sigma\right\}$$

$$= \exp\left\{\frac{-\frac{y_i}{2{\mu_i}^2} - \left(-\sqrt{\frac{1}{{\mu_i}^2}}\right)}{\sigma^2} - \frac{1}{2y_i\sigma^2} - \frac{1}{2}\ln(2\pi y^3) - \ln\sigma\right\}$$

$$= \exp\left\{\frac{y_i\left(-\frac{1}{2{\mu_i}^2}\right) - \left(-\sqrt{2\left(-\frac{1}{2{\mu_i}^2}\right)}\right)}{\sigma^2} - \frac{1}{2y_i\sigma^2} - \frac{1}{2}\ln(2\pi y^3) - \ln\sigma\right\}$$

$$= \exp\left\{\frac{y_i\left(-\frac{1}{2{\mu_i}^2}\right) - \left(-\sqrt{2\theta}\right)}{\sigma^2} - \frac{1}{2y_i\sigma^2} - \frac{1}{2}\ln(2\pi y^3) - \ln\sigma\right\}$$

$$f(y_i) = \exp\left\{\frac{y_i\theta - b(\theta)}{a(\phi)} + c(y_i,\phi)\right\}$$

dengan, $\theta = -\frac{1}{2{\mu_i}^2}$; $b(\theta) = -\left(-\sqrt{2\theta}\right)$; $a(\phi) = \phi = \sigma^2$; $c(y_i,\phi) = -\frac{1}{2y_i\sigma^2} - \frac{1}{2}\ln(2\pi y^3) - \ln\sigma$.

Pada tabel 1 halaman 24 diketahui bahwa nilai harapan dan ragam pada regresi *inverse Gaussian* adalah $E(Y_i) = \mu_i$ dan $Var\ (Y_i) = \sigma^2{\mu_i}^3$.

a. Pembuktian nilai harapan dan ragam pada regresi *inverse Gaussian*

   i. Nilai harapan atau $E(Y_i) = \mu_i$

      Bukti:

Pada persamaan (2.19) sudah dibuktikan bahwa $E(Y_i) = \dot{b}(\theta)$.

$$E(Y_i) = \dot{b}(\theta)$$

$$= \frac{\partial b(\theta)}{\partial\theta}$$

$$= \frac{\partial\left[-\left(-\sqrt{2\theta}\right)\right]}{\partial\theta}$$

$$E(Y_i) = -\left\{\frac{1}{2}(-2\theta)^{-\frac{1}{2}}(-2)\right\}$$

$$= \frac{1}{\sqrt{-2\theta}}, dengan\ \theta = -\frac{1}{2{\mu_i}^2}$$

$$= \frac{1}{\sqrt{-2(-\frac{1}{2{\mu_i}^2})}}$$

$$= \frac{1}{\sqrt{\frac{1}{{\mu_i}^2}}}$$

$$= \frac{1}{\frac{1}{\mu_i}}$$

$E(Y_i) = \mu_i$ (Terbukti)

ii. Ragam atau $Var\ (Y_i) = \sigma^2{\mu_i}^3$

Bukti:

Pada persamaan (2.20) sudah dibuktikan bahwa $Var(Y_i) = \phi\ddot{b}(\theta)$.

$$Var(Y_i) = \phi\ddot{b}(\theta)$$

$$= \frac{\partial\dot{b}(\theta)}{\partial\theta}$$

$$= \phi\left\{-\frac{1}{2}(-2\theta)^{-\frac{3}{2}}(-2)\right\}$$

$$= \phi(-2\theta)^{-\frac{3}{2}}, dengan\ \theta = -\frac{1}{2{\mu_i}^2}$$

$$= \left\{-2\left(-\frac{1}{2{\mu_i}^2}\right)^{-\frac{3}{2}}\right\}\sigma^2$$

$$= ({\mu_i}^{-2})^{-\frac{3}{2}}\sigma^2$$

$$Var(Y_i) = {\mu_i}^3\sigma^2 \text{ (Terbukti)}$$

Salah satu tujuan dari analisis regresi adalah untuk menentukan pola hubungan antara peubah respons dengan peubah penjelas, maka dalam regresi *inverse Gaussian* hubungan tersebut dapat dituliskan dalam bentuk:

$$E(Y_i) = \mu_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik}$$

atau

$$E(Y_i) = \boldsymbol{\mu_i} = \boldsymbol{X'\beta} \qquad (3.2)$$

Regresi *inverse Gaussian* merupakan model regresi yang berdistribusi *inverse Gaussian* dengan peubah respons berbentuk eksponensial. Fungsi hubung log lebih sering digunakan dalam regresi *inverse Gaussian*, sehingga untuk memperoleh estimator parameter μ_i maka bentuk persamaan dugaan regresi *inverse Gaussian* adalah sebagai berikut:

$$\hat{\mu}_i = e^{\mu_i}$$

dengan,

$\hat{\mu}_i$: estimator parameter μ_i
μ_i: $\hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \cdots + \hat{\beta}_k X_{ik}$

Fungsi hubung dikatakan fungsi penghubung kanonik bila parameter kanoniknya sama dengan fungsi hubungnya, yaitu $\eta = \theta$; dengan θ adalah parameter kanonik. Pada regresi *inverse Gaussian* diketahui bahwa $\theta = -\frac{1}{2\mu_i{}^2}$, sehingga karena nilai $\mu_i > 0$ (harus positif) maka digunakan fungsi hubung $g(\mu_i) = \eta_i = \mu_i{}^{-2}$ atau $\boldsymbol{\eta_i} = \mu_i{}^{-2} = \boldsymbol{X}'\boldsymbol{\beta}.$ Dengan demikian model regresi *inverse Gaussian* dapat ditulis dalam bentuk

$$E(Y_i) = \boldsymbol{\mu_i} = \boldsymbol{X}'\boldsymbol{\beta}, \text{ dan } \; g(\mu_i) = \mu_i{}^{-2} \qquad (3.3)$$

dengan $\boldsymbol{\beta}$ merupakan parameter yang tidak diketahui dalam model dan harus diestimasi.

2. Estimasi Parameter

Metode yang digunakan untuk mengestimasi parameter peubah respons distribusi *inverse Gaussian* pada penulisan ini adalah metode *MLE*. Diketahui fungsi kepadatan peluang distribusi *inverse Gaussian* adalah:

$$f(y_i) = \frac{1}{\sigma\sqrt{2\pi y_i{}^3}} \exp\left\{-\frac{1}{2y_i}\left(\frac{y_i - \mu_i}{\mu_i \sigma}\right)^2\right\}, y_i > 0, \sigma > 0, \mu_i > 0$$

dengan $i = 1,2, \dots, n$.

Y_i adalah peubah acak dengan nilai harapan μ_i dan ragam $\sigma^2\mu_i{}^3$ (Jong & Heller, 2008: 29), dengan μ_i dan $\sigma^2\mu_i{}^3$ tidak diketahui.

Metode *MLE* dapat dilakukan jika distribusi dari data diketahui. Langkah pertama yang dilakukan adalah menentukan fungsi *likelihood* dari model regresi *inverse Gaussian*. Dengan mengasumsikan $y_1, y_2, y_3, \dots, y_n$ adalah sekumpulan peubah acak *inverse Gaussian* yang saling bebas atau y_i berdistribusi *inverse Gaussian*, maka diperoleh fungsi *likelihood*.

a. Fungsi *likelihood* peubah respons distribusi *inverse Gaussian* adalah

$$L(\mu_i, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi {y_i}^3}} \exp\left\{-\frac{1}{2y_i}\left(\frac{y_i - \mu_i}{\mu_i \sigma}\right)^2\right\}$$

$$= \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi {y_i}^3}} e^{\left\{-\frac{1}{2y_i}\left(\frac{y_i - \mu_i}{\mu_i \sigma}\right)^2\right\}}$$

$$= \frac{1}{(2\pi\sigma^2 {y_i}^3)^{\frac{n}{2}}} e^{-\left(\frac{1}{2\sigma^2 {\mu_i}^2 y_i}\right)\sum_{i=1}^{n}(y_i - \mu_i)^2} \tag{3.4}$$

dengan $\boldsymbol{\mu_i} = \boldsymbol{X'\beta}$, $i = 1,2,3, \dots, n$. Selanjutnya, dari fungsi *likelihood*, kedua ruas di logaritma-naturalkan.

b. Logaritma natural dari kedua ruas atau disebut fungsi *log-likelihood* menjadi:

$$\ln L(\mu_i, \sigma^2) = -\frac{n}{2}\ln(2\pi {y_i}^3) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2 {\mu_i}^2 y_i}\sum_{i=1}^{n}(y_i - \mu_i)^2$$

$$= -\frac{n}{2}\ln(2\pi {y_i}^3) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2y_i}\sum_{i=1}^{n}\left(\frac{y_i - \mu_i}{\mu_i \sigma}\right)^2$$

$$= -\frac{n}{2}\ln(2\pi {y_i}^3) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2y_i}\sum_{i=1}^{n}\left(\frac{{y_i}^2 - 2y_i\mu_i + {\mu_i}^2}{{\mu_i}^2\sigma^2}\right)$$

$$= -\frac{n}{2}\ln(2\pi {y_i}^3) - \frac{n}{2}\ln\sigma^2 - \sum_{i=1}^{n}\frac{{y_i}^2 - 2y_i\mu_i + {\mu_i}^2}{2y_i{\mu_i}^2\sigma^2}$$

$$= -\frac{n}{2}\ln(2\pi {y_i}^3) - \frac{n}{2}\ln\sigma^2 - \sum_{i=1}^{n}\frac{{y_i}^2}{2y_i{\mu_i}^2\sigma^2} + \sum_{i=1}^{n}\frac{2y_i\mu_i}{2y_i{\mu_i}^2\sigma^2} - \sum_{i=1}^{n}\frac{{\mu_i}^2}{2y_i{\mu_i}^2\sigma^2}$$

$$= -\frac{n}{2}\ln(2\pi {y_i}^3) - \frac{n}{2}\ln\sigma^2 - \sum_{i=1}^{n}\frac{y_i}{2{\mu_i}^2\sigma^2} + \sum_{i=1}^{n}\frac{1}{\mu_i\sigma^2} - \sum_{i=1}^{n}\frac{1}{2y_i\sigma^2}$$

$$= -\frac{1}{2}\sum_{i=1}^{n}\frac{y_i}{{\mu_i}^2\sigma^2} + \sum_{i=1}^{n}\frac{1}{\mu_i\sigma^2} - \frac{1}{2}\sum_{i=1}^{n}\frac{1}{y_i\sigma^2} - \frac{n}{2}\ln(2\pi {y_i}^3) - \frac{n}{2}\ln\sigma^2$$

Karena $\boldsymbol{\mu_i} = \boldsymbol{X'\beta}$, maka

$$\ln L(\sigma,\boldsymbol{\beta}) = -\frac{1}{2}\sum_{i=1}^{n}\frac{y_i}{(\boldsymbol{X'\beta})^2\sigma^2} + \sum_{i=1}^{n}\frac{1}{(\boldsymbol{X'\beta})\sigma^2} - \frac{1}{2}\sum_{i=1}^{n}\frac{1}{y_i\sigma^2} - \frac{n}{2}\ln(2\pi {y_i}^3) - \frac{n}{2}\ln\sigma^2 \quad (3.5)$$

Kemudian, persamaan (3.5) diturunkan terhadap σ dan β dan disamadengankan nol.

c. Turunan dari $\ln L(\sigma,\boldsymbol{\beta})$ yang disamadengankan nol

   i. Turunan $\ln L(\sigma,\boldsymbol{\beta})$ terhadap σ

   Sebelum menurunkan $\ln L(\sigma,\boldsymbol{\beta})$ terhadap σ, persamaan (3.5) diubah terlebih dahulu ke dalam bentuk sebagai berikut

$$\ln L(\sigma,\boldsymbol{\beta}) = -\frac{1}{2}\sum_{i=1}^{n}\frac{y_i\sigma^{-2}}{(\boldsymbol{X}'\boldsymbol{\beta})^2} + \sum_{i=1}^{n}\frac{\sigma^{-2}}{(\boldsymbol{X}'\boldsymbol{\beta})} - \frac{1}{2}\sum_{i=1}^{n}\frac{\sigma^{-2}}{y_i} - \frac{n}{2}\ln(2\pi {y_i}^3)$$

$$-\frac{n}{2}\ln\sigma^2 \qquad (3.6)$$

Turunan persamaan (3.6) terhadap σ yang disamadengankan nol adalah

$$\frac{\partial \ln L(\sigma,\boldsymbol{\beta})}{\partial\sigma} = \sum_{i=1}^{n}\frac{y_i\sigma^{-3}}{(\boldsymbol{X}'\boldsymbol{\beta})^2} - 2\sum_{i=1}^{n}\frac{\sigma^{-3}}{(\boldsymbol{X}'\boldsymbol{\beta})} + \sum_{i=1}^{n}\frac{\sigma^{-3}}{y_i} - \frac{1}{\sigma^2} = 0$$

$$0 = \sum_{i=1}^{n}\frac{y_i}{(\boldsymbol{X}'\boldsymbol{\beta})^2\sigma^3} - 2\sum_{i=1}^{n}\frac{1}{(\boldsymbol{X}'\boldsymbol{\beta})\sigma^3} + \sum_{i=1}^{n}\frac{1}{y_i\sigma^3} - \frac{1}{\sigma^2} \qquad (3.7)$$

ii. Turunan $\ln L(\sigma,\boldsymbol{\beta})$ terhadap $\boldsymbol{\beta}$

Sebelum menurunkan $\ln L(\sigma,\boldsymbol{\beta})$ terhadap $\boldsymbol{\beta}$, persamaan (3.5) diubah terlebih dahulu ke dalam bentuk sebagai berikut

$$\ln L(\sigma,\boldsymbol{\beta}) = -\frac{1}{2}\sum_{i=1}^{n}\frac{y_i}{\left({\boldsymbol{X}'}^2\boldsymbol{\beta}^2\right)\sigma^2} + \sum_{i=1}^{n}\frac{1}{(\boldsymbol{X}'\boldsymbol{\beta})\sigma^2} - \frac{1}{2}\sum_{i=1}^{n}\frac{1}{y_i\sigma^2} - \frac{n}{2}\ln(2\pi {y_i}^3)$$

$$-\frac{n}{2}\ln\sigma^2$$

$$\ln L(\sigma,\boldsymbol{\beta}) = -\frac{1}{2}\sum_{i=1}^{n}\frac{y_i\boldsymbol{\beta}^{-2}}{{\boldsymbol{X}'}^2\sigma^2} + \sum_{i=1}^{n}\frac{\boldsymbol{\beta}^{-1}}{\boldsymbol{X}'\sigma^2} - \frac{1}{2}\sum_{i=1}^{n}\frac{1}{y_i\sigma^2} - \frac{n}{2}\ln(2\pi {y_i}^3)$$

$$-\frac{n}{2}\ln\sigma^2 \qquad (3.8)$$

Turunan persamaan (3.8) terhadap $\boldsymbol{\beta}$ yang disamadengankan nol adalah

$$\frac{\partial \ln L(\sigma, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \frac{3}{2}\sum_{i=1}^{n} \frac{y_i \boldsymbol{\beta}^{-3}}{\boldsymbol{X}'^2 \sigma^2} - \sum_{i=1}^{n} \frac{\boldsymbol{\beta}^{-2}}{\boldsymbol{X}' \sigma^2} = 0$$

$$0 = \frac{3}{2}\sum_{i=1}^{n} \frac{y_i}{(\boldsymbol{X}'^2 \boldsymbol{\beta}^3)\sigma^2} - \sum_{i=1}^{n} \frac{1}{(\boldsymbol{X}' \boldsymbol{\beta}^2)\sigma^2} \tag{3.9}$$

Persamaan (3.7) dan (3.9) di atas tidak memberikan penyelesaian karena kedua persamaaan di atas saling terkait satu sama lain. Misalnya pada persamaan (3.7) setelah fungsi *log-likelihood* diturunkan terhadap σ ternyata turunannya masih mempunyai parameter lain yaitu $\boldsymbol{\beta}$. Begitu juga dengan persamaan (3.9) sehingga pengestimasian kedua parameter ini harus dilakukan secara bersamaan. Dari persamaan (3.7) dan (3.9) tidak diperoleh estimator yang eksak. Oleh karena itu digunakan bantuan komputer dengan program SAS 9.1.3. Dalam program SAS 9.1.3, nilai estimator dicari dengan menggunakan metode *MLE* yang diselesaikan dengan metode numerik iterasi yang disebut sebagai metode Newton-Raphson.

Metode Newton-Raphson adalah metode numerik untuk menyelesaikan persamaan non-linear secara iteratif seperti persamaan *likelihood* yang memaksimumkan suatu fungsi. Dasar dari metode ini adalah pendekatan deret Taylor linear. Dalam metode Newton-Raphson dibutuhkan turunan pertama dan kedua dari fungsi *log-likelihood*nya. Untuk mengestimasi σ dan $\boldsymbol{\beta}$ dengan metode Newton-Raphson diperlukan estimasi awal dari σ dan $\boldsymbol{\beta}$.

Karena model regresi *inverse Gaussian* menggunakan fungsi hubung log, dan setelah diperoleh nilai estimator parameter regresi *inverse Gaussian* dengan metode *MLE* pada program SAS 9.1.3 maka persamaan regresi *inverse Gaussian* dugaan dengan peubah respons berdistribusi *inverse Gaussian* adalah sebagai berikut:

$$\hat{\mu} = e^{\mu}$$

$$\hat{\mu} = e^{\hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \cdots + \hat{\beta}_k X_{ik}}$$

B. Contoh Penerapan Regresi *Inverse Gaussian*

Contoh penerapan regresi *inverse Gaussian* adalah data inflasi di Indonesia pada periode tahun 1980-2010. Data inflasi diambil dari Badan Pusat Statistik (BPS) Republik Indonesia. Tabel 4 mencantumkan data inflasi, bahan makanan, perumahan, sandang, kesehatan dan pendidikan. Dari data inflasi, ingin diketahui hubungan antara bahan makanan, perumahan, sandang, kesehatan dan pendidikan terhadap inflasi. Pada tabel 4, data inflasi sebagai peubah respons *Y* sedangkan bahan makanan, perumahan, sandang, kesehatan dan pendidikan sebagai peubah penjelas *X*. Histogram data inflasi pada gambar 6 halaman 58 menunjukkan bahwa kurva menceng ke kanan, sehingga data inflasi tersebut merupakan regresi *inverse*

Gaussian yang menyatakan hubungan antara peubah respons dengan peubah penjelas.

Dalam ekonomi, inflasi adalah kenaikan harga barang dan jasa secara umum dimana barang dan jasa tersebut merupakan kebutuhan pokok masyarakat atau turunnya daya jual mata uang suatu negara (http://www.bps.go.id). Inflasi terjadi karena adanya kenaikan harga yang ditunjukkan oleh kenaikan indeks pada kelompok bahan makanan, perumahan, sandang, kesehatan dan pendidikan.

Data inflasi di Indonesia periode tahun 1980-2010 digunakan untuk mengetahui hubungan antara peubah respons yaitu inflasi dengan peubah penjelas yaitu bahan makanan, perumahan, sandang, kesehatan dan pendidikan. Pada data inflasi untuk mengetahui apakah regresi *inverse Gaussian* layak digunakan dalam model atau tidak layak digunakan dengan uji *goodness of fit*. Sedangkan untuk mengetahui signifikansi koefisien regresi digunakan uji Wald.

Perhitungan nilai inflasi didasarkan pada Indeks Harga Konsumen (IHK) pada tahun yang bersangkutan. IHK merupakan nomor indeks yang mengukur harga rata-rata dari barang dan jasa yang dikonsumsi oleh rumah tangga (http://id.wikipedia.org/wiki). Pembagian kelompok-kelompok IHK pada inflasi periode tahun 1980-2010 yaitu: bahan makanan, perumahan, sandang, kesehatan dan pendidikan.

Diberikan data inflasi yaitu pada tabel 4 dengan penjelasan sebagai berikut. Sebagai contoh pada tahun 1980 terjadi inflasi sebesar 16 persen. Inflasi ini terjadi karena adanya kenaikan harga yang ditunjukkan oleh kenaikan indeks pada kelompok bahan makanan 16,3 persen; kelompok perumahan 18,3 persen; kelompok sandang 12,7 persen; kelompok kesehatan 14,6 persen dan kelompok pendidikan 13,8 persen. Pada tahun 1998 terjadi inflasi sebesar 77,66 persen. Inflasi pada tahun 1998 sangat tinggi, hal ini dikarenakan pada tahun 1997 di Indonesia terjadi krisis moneter yang menyebabkan tingginya nilai inflasi. Inflasi tahun 1998 sangat tinggi yang terjadi karena adanya kenaikan harga yang tinggi yang ditunjukkan oleh kenaikan indeks pada kelompok bahan makanan 118,4 persen; kelompok perumahan 47,5 persen; kelompok sandang 98,7 persen; kelompok kesehatan 86,1 persen dan kelompok pendidikan 9,7 persen.

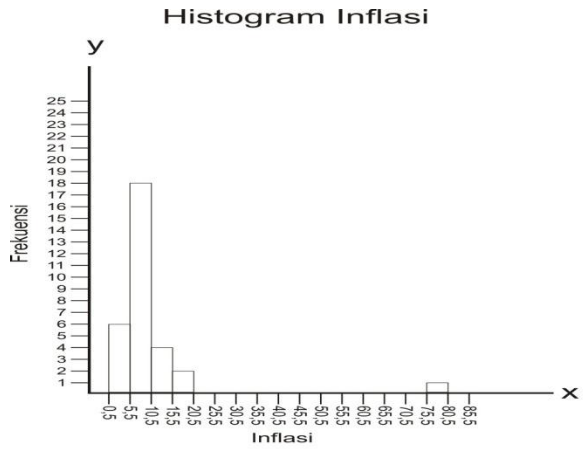
Peubah respons pada contoh penerapan *inverse Gaussian* berupa nilai inflasi (Y_i) sedangkan peubah penjelas berupa bahan makanan (X_1), perumahan (X_2), sandang (X_3), kesehatan (X_4), dan pendidikan (X_5). Semua data periode tahun 1980-2010 dalam satuan persen. Berikut diberikan data inflasi pada periode tahun 1980-2010 di Indonesia.

Tabel 4. Data Inflasi Periode tahun 1980-2010 di Indonesia

Tahun	Inflasi	Bahan Makanan	Perumahan	Sandang	Kesehatan	Pendidikan
1980	16	16,3	18,3	12,7	14,6	13,8
1981	7,1	8	7,7	3,8	4,1	7,4
1982	6,7	7,3	14,3	3,4	11,4	6,9
1983	11,5	10	12,9	4,3	10,5	9,9
1984	8,8	6,3	12,8	3	7,6	8,9
1985	4,3	2,1	7	3,3	5,4	8,5
1986	8,8	13,6	4,6	9,5	9,3	4,4
1987	8,9	11,7	6	7,7	5,6	7,8
1988	5,5	7,8	4,3	3,5	2,4	7
1989	6	6,7	6,1	4,7	5,7	6,1
1990	9,5	7	12,4	4,8	9,2	6,4
1991	9,5	9,7	7,7	5,2	5,4	8,4
1992	4,9	6	4,6	7,2	3	8
1993	9,8	5,1	15,5	8	13,8	10,3
1994	9,2	13,9	9,1	6,1	13,5	9,5
1995	8,6	13,3	5,7	6,5	7,8	12,4
1996	6,5	6,1	4,7	5,8	11	7,6
1997	11,1	18,5	6,1	7,7	13,4	14,8
1998	77,6	118,4	47,5	98,7	86,1	9,7
1999	2	-5,3	5,2	6,5	3,9	11
2000	9,4	4	10,1	10,2	9,6	27,4
2001	12,6	12	13,6	8,1	8,9	17,4
2002	10	9,1	12,7	2,7	5,6	16,5
2003	5,1	-1,7	9,4	7,1	5,7	21,5
2004	6,4	6,4	7,4	4,9	4,8	17,5
2005	17,11	13,91	13,94	6,92	6,13	8,24
2006	6,6	12,94	4,83	6,84	5,87	8,13
2007	6,59	11,26	4,88	8,42	4,31	8,83
2008	11,06	16,35	10,92	7,33	7,96	6,66
2009	2,78	3,88	1,83	6	3,89	3,89
2010	6,96	15,64	4,08	6,51	2,19	3,29

Sumber: BPS Republik Indonesia

Keterangan: data dalam satuan persen.



Gambar 6. Histogram peubah respons (inflasi)

Pada histogram inflasi diatas terlihat bahwa peubah respons (inflasi) adalah peubah kontinu yang memiliki nilai yang positif dan sebagian besar nilai dengan frekuensi rendah kebanyakan berada di sebelah kanan sehingga dapat diketahui bahwa kurva menceng ke kanan (“ekor” kurva ada disebelah kanan). Sehingga data cenderung terkonsentrasi pada nilai yang rendah. Hal ini menunjukkan bahwa model regresi *inverse Gaussian* layak digunakan.

Dengan hasil *output* menggunakan SAS 9.1.3 sebagai berikut:

Tabel 5. Ringkasan *output* SAS 9.1.3 dengan fungsi hubung log

Parameter	Db	$\hat{\beta}$	*Standard Error*	χ^2	*p-value*
Intersep	1	0,7472	0,1211	38,09	<0,0001
Bahan Makanan	1	0,0704	0,0065	116,49	<0,0001
Perumahan	1	0,0548	0,0118	21,67	<0,0001
Sandang	1	-0,0145	0,0189	0,59	0,4439
Kesehatan	1	0,0130	0,0126	1,05	0,3054
Pendidikan	1	0,0199	0,0074	7,15	0,0075
Goodness of Fit	*DF*	*Value*	*Value/ DF*		
Deviance	25	0,1184	0,0047		

Dari hasil output SAS 9.1.3 diperoleh model regresi *inverse Gaussian* dengan fungsi hubung log. Persamaan regresi *inverse Gaussian* dugaan adalah sebagai berikut:

$$\hat{\mu} = \exp(0{,}7472 + 0{,}0704\mathrm{X}_1 + 0{,}0548\mathrm{X}_2 - 0{,}0145\mathrm{X}_3 + 0{,}0130\mathrm{X}_4 + 0{,}0199\mathrm{X}_5)$$

$$\hat{\mu} = e^{(0{,}7472+0{,}0704\mathrm{X}_1+0{,}0548\mathrm{X}_2-0{,}0145\mathrm{X}_3+0{,}0130\mathrm{X}_4+0{,}0199\mathrm{X}_5)},\ e = 2{,}178$$

dengan X_1, X_2, X_3, X_4 & X_5 adalah peubah penjelas yaitu bahan makanan, perumahan, sandang, kesehatan dan pendidikan.

Koefisien β_0 sebesar $e^{0{,}7472} = 2{,}11$ menyatakan bahwa setiap kenaikan (karena tanda positif) peubah penjelas tiap satu satuan persen akan meningkatkan rata-rata tingkat nilai inflasi sebesar $2{,}11\%$.

Koefisien X_1 sebesar $e^{0,0704} = 1,073$ menyatakan bahwa setiap kenaikan (karena tanda positif) bahan makanan tiap satu satuan persen akan meningkatkan rata-rata tingkat nilai inflasi sebesar $1,073$ %.

Koefisien X_2 sebesar $e^{0,0548} = 1,056$ menyatakan bahwa setiap kenaikan (karena tanda positif) perumahan tiap satu satuan persen akan meningkatkan rata-rata tingkat nilai inflasi sebesar $1,056$ %.

Koefisien X_3 sebesar $e^{-0,0145} = 0,986$ menyatakan bahwa setiap kenaikan (karena tanda positif) sandang tiap satu satuan persen akan menaikan rata-rata tingkat nilai inflasi sebesar $0,986$ %.

Koefisien X_4 sebesar $e^{0,0130} = 1,013$ menyatakan bahwa setiap kenaikan (karena tanda positif) kesehatan tiap satu satuan persen akan meningkatkan rata-rata tingkat nilai inflasi sebesar $1,013$ %.

Koefisien X_5 sebesar $e^{0,0199} = 1,02$ menyatakan bahwa setiap kenaikan (karena tanda positif) pendidikan tiap satu satuan persen akan meningkatkan rata-rata tingkat nilai inflasi sebesar $1,02$ %.

Pada permasalahan data inflasi akan diketahui apakah model regresi *inverse Gaussian* layak digunakan dalam model atau tidak, sehingga perlu adanya uji *goodness of fit*.

a. Langkah-langkah pengujian pada regresi *inverse Gaussian* dengan uji *goodness of fit* (Jong & Heller, 2008:71)

1. Merumuskan hipotesis

   H_0: Model regresi *inverse Gaussian* layak digunakan.

   H_1: Model regresi *inverse Gaussian* tidak layak digunakan.

2. Memilih tingkat signifikansi α

   Digunakan tingkat signifikansi $\alpha = 5\% = 0{,}05$.

3. Menentukan statistik uji

   Statistik uji yang digunakan adalah *Deviance* (Tabel 3).

   Untuk regresi *Inverse Gaussian*, *deviance* yang digunakan yaitu:

$$Deviance\ (\Delta) = \frac{1}{\sigma^2}\sum_{i=1}^{n}\frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i^2 y_i}$$

4. Kriteria keputusan

   H_0 ditolak jika *Deviance* $> \chi^2_{\alpha(n-k)}$.

   Dengan *n*: banyaknya pengamatan, *k*: banyaknya parameter.

5. Menarik kesimpulan

   Jika nilai *deviance* $\leq \chi^2_{\alpha(n-k)} = \chi^2_{0,05(25)} = 37{,}652$, maka dapat disimpulkan bahwa model regresi *inverse Gaussian* layak digunakan. Atau secara langsung dapat dilihat melalui nilai *Value/DF* dari hasil *output* analisis menggunakan bantuan program SAS 9.1.3, jika *deviance* $\leq \chi^2_{\alpha(n-k)}$ maka model regresi *inverse Gaussian* layak digunakan.

   Berdasarkan Tabel 5 halaman 58, *Goodness of Fit* dengan kriteria *deviance* diperoleh nilai *Value/DF* untuk model regresi *inverse Gaussian* yaitu sebesar 0,0047 lebih kecil dari $\chi^2_{0,05(25)} \approx 37{,}652$ maka H_0 diterima. Sehingga dapat dikatakan bahwa model regresi *inverse Gaussian* layak digunakan.

Setelah dilakukan uji kelayakan model dengan uji *goodness of fit*, selanjutnya untuk mengetahui signifikansi koefisien regresi terhadap peubah penjelas, maka dilakukan uji signifikansi yaitu dengan uji Wald.

b. Langkah-langkah Uji Signifikansi Koefisien Regresi *Inverse Gaussian* dengan Uji Wald

1. Merumuskan hipotesis

   $H_0: \beta_j = 0, j = 0,1,2,3,4,5$

   $H_1: \beta_j \neq 0, j = 0,1,2,3,4,5$

2. Memilih tingkat signifikansi: α.

   Digunakan tingkat signifikansi $\alpha = 5\% = 0{,}05$.

3. Menentukan statistik uji

   Statistik uji yang digunakan untuk uji koefisien regresi *inverse Gausssian* adalah uji Wald.

$$W = \left\{\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}\right\}^2 \sim \chi_1^2, j = 0,1,2,3,4,5$$

4. Kriteria Keputusan:

   H_0 ditolak jika $W < \chi^2_{(0{,}975)(1)} = 0{,}000982$ atau $W > \chi^2_{(0{,}025)(1)} = 5{,}024$.

5. Perhitungan

Koefisien	$W = \left\{\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}\right\}^2 = \chi^2_{hit}$	Kesimpulan
β_0	$\left\{\frac{0,7472}{0,1211}\right\}^2 = 38,09$	H_o ditolak
β_1	$\left\{\frac{0,0704}{0,0065}\right\}^2 = 116,49$	H_o ditolak
β_2	$\left\{\frac{0,0548}{0,0118}\right\}^2 = 21,67$	H_o ditolak
β_3	$\left\{\frac{-0,0145}{0,0189}\right\}^2 = 0,59$	H_o diterima
β_4	$\left\{\frac{0,0130}{0,0126}\right\}^2 = 1,05$	H_o diterima
β_5	$\left\{\frac{0,0199}{0,0074}\right\}^2 = 7,15$	H_o ditolak

6. Kesimpulan

Dari tabel di atas diketahui bahwa untuk koefisien $\beta_0, \beta_1, \beta_2, \beta_5$ memiliki nilai $W > \chi^2_{(0,025)(1)} = 5,024$ sehingga H_o ditolak. Berarti bahwa bahan makanan, perumahan dan pendidikan memiliki koefisien regresi signifikan di dalam model sehingga bahan makanan, perumahan dan pendidikan tetap di dalam model. Sedangkan untuk koefisien β_3 dan β_4 memiliki nilai $W > \chi^2_{\left(1-\frac{\alpha}{2}\right)(1)}$ dan $W < \chi^2_{\frac{\alpha}{2}(1)}$ maka H_0 diterima. Berarti bahwa untuk sandang dan kesehatan memiliki koefisien regresi yang tidak signifikan di dalam model, sehingga sandang dan kesehatan dikeluarkan dari model.

BAB IV

KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan pembahasan mengenai analisis peubah respons kontinu non negatif dengan regresi *inverse Gaussian* maka dapat diambil kesimpulan sebagai berikut:

1. Cara mengestimasi parameter pada analisis peubah respons kontinu non negatif dengan distribusi *inverse Gaussian* yaitu dengan menggunakan metode *MLE*. Langkah-langkah metode *MLE* adalah:
   a. Menentukan fungsi kepadatan peluang peubah respons yang berdistribusi *inverse Gaussian* yaitu

   $$f(y_i) = \frac{1}{\sigma\sqrt{2\pi {y_i}^3}} \exp\left\{-\frac{1}{2y_i}\left(\frac{y_i - \mu_i}{\mu_i \sigma}\right)^2\right\}, y_i > 0, \sigma > 0, \mu_i > 0 \quad (3.1)$$

   dengan $i = 1,2, \dots, n$.

   Keterangan:

   Y_i : peubah acak pada pengamatan ke-*i*
   $f(y_i)$: fungsi kepadatan peluang distribusi *inverse Gaussian*
   μ_i, σ : parameter

   b. Menentukan fungsi *likelihood* peubah respons yang berdistribusi *inverse Gaussian* yaitu sebagai berikut:

   $$L(\mu_i, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi {y_i}^3}} \exp\left\{-\frac{1}{2y_i}\left(\frac{y_i - \mu_i}{\mu_i \sigma}\right)^2\right\}, \text{dengan } \boldsymbol{\mu_i} = \boldsymbol{X}'\boldsymbol{\beta}$$

c. Melogaritma-naturalkan kedua ruas atau disebut fungsi *log-likelihood*. Bentuk fungsi *log-likelihood* peubah respons yang berdistribusi *inverse Gaussian* adalah sebagai berikut:

$$\ln L(\sigma, \boldsymbol{\beta}) = -\frac{1}{2}\sum_{i=1}^{n}\frac{y_i}{(\boldsymbol{X}'\boldsymbol{\beta})^2\sigma^2} + \sum_{i=1}^{n}\frac{1}{(\boldsymbol{X}'\boldsymbol{\beta})\sigma^2} - \frac{1}{2}\sum_{i=1}^{n}\frac{1}{y_i\sigma^2} - \frac{n}{2}\ln(2\pi {y_i}^3) - \frac{n}{2}\ln\sigma^2$$

d. Menurunkan kedua ruas fungsi *log-likelihood* dengan disamadengankan nol.

e. Setelah menurunkan terhadap fungsi *log-likelihood* ternyata tidak diperoleh estimator yang eksak, sehingga pengestimasian kedua parameter harus dilakukan secara bersamaan. Oleh karena itu digunakan bantuan komputer dengan program SAS 9.1.3. Dalam program SAS 9.1.3, nilai estimator dicari dengan menggunakan metode *MLE* yang diselesaikan dengan metode numerik iterasi yang disebut sebagai metode Newton-Raphson. Fungsi hubung yang digunakan pada model regresi *inverse Gaussian* adalah fungsi hubung log, maka persamaan dugaan regresi *inverse Gaussian* adalah:

$$\hat{\mu} = e^{\hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \cdots + \hat{\beta}_k X_{ik}}$$

2. Contoh penerapan regresi *inverse Gaussian* yaitu pada data inflasi di Indonesia. Pada histogram data inflasi terlihat bahwa peubah respons (inflasi) merupakan peubah kontinu yang memiliki nilai positif, sehingga kurva menceng ke kanan. Hal ini berarti bahwa regresi *inverse Gaussian* layak digunakan pada data inflasi. Hasil output pada tabel 5 halaman 58

diperoleh kesimpulan untuk uji Wald, peubah penjelas yaitu bahan makanan, perumahan dan pendidikan memiliki koefisien regresi yang signifikan di dalam model yang berarti bahan makanan, perumahan dan pendidikan tetap di dalam model. Sedangkan untuk sandang dan kesehatan memiliki koefisien regresi yang tidak signifikan di dalam model yang berarti bahwa sandang dan kesehatan dikeluarkan dari model.

B. Saran

Dalam penulisan skripsi ini, penulis melakukan analisis peubah respons kontinu non negatif dengan regresi *inverse Gaussian* yang termasuk dalam *GLM (Generalized Linear Model)* dengan metode *MLE (Maximum Likelihood Estimation)*. Bagi pembaca yang berminat dengan permasalahan analisis regresi khususnya regresi peubah kontinu non negatif, penulis menyarankan untuk mencari nilai estimasi parameter dengan *MLE* yang diselesaikan dengan *IRLS (Iteratively Weighted Least Squares)* menggunakan program S-Plus.

DAFTAR PUSTAKA

Agresti, Alan. 2007. *An Introduction to Categorical Data Analysis Second Edition*. Florida: Departement of Statistics University of Florida.

Anton, Howard. 1987. *Aljabar Linear Elementer* (Silaban, P. & Susila I. N., terjemahan). Jakarta: Erlangga.

Atmaja, L. S. 2009. *Statistika untuk Bisnis dan Ekonomi*. Yogyakarta: Andi.

Bain, L. J. & Engelhard, M. 1992. *Introduction to the Probability and Mathematical Statistics*. California: Duxbury Press.

Chapra, S. C. & Chanale, R. P. 1996. *Metode Numerik* (Susila, I. N., terjemahan). Jakarta: Erlangga.

Gujarati, D. N. 2007. *Dasar-dasar Ekonometrika jilid 1 edisi Ke-3* (Mulyadi, J. A., terjemahan). Jakarta: Erlangga.

Jong, P. D. & Heller, G. Z. 2008. *Generalized Linear Models for Insurance Data*. Cambridge: Cambridge University Press.

Hadley, G. 1992. *Aljabar Linear*. Jakarta: Erlangga.

McCullagh, P. & Nelder, J. A. 1989. *Generalized Linear Models* 2^{nd} Edition. London: Chapman & Hall.

Neter, John. 1997. *Model Linear Terapan edisi ketiga* (Sumantri, Bambang., terjemahan). Bogor: IPB.

SAS Institute Inc. 2004. SAS/STAT® 9.1 *User's Guide*. North Carolina: SAS Institute Inc.

Walpole, R. E. & Myers, R. H. 1995. *Ilmu Peluang dan Statistika untuk Insinyur dan Ilmuwan Edisi ke-4* (Sembiring, R. K., terjemahan). Bandung: ITB.

BPS. 2011. *Inflasi Indonesia Menurut Kelompok Komoditi*. http://www.bps.go.id. [8 Maret 2011].

Wikipedia. 2011. *Indeks Harga Konsumen*. http://id.wikipedia.org/wiki. [12 April 2011].

L
A
M
P
I
R
A
N

Lampiran 1. Data Inflasi di Indonesia periode Tahun 1980-2010

Inflasi Indonesia Menurut Kelompok Komoditi,
Periode Tahun 1980-2010

Tahun	Inflasi	Bahan Makanan	Perumahan	Sandang	Kesehatan	Pendidikan
1980	16	16,3	18,3	12,7	14,6	13,8
1981	7,1	8	7,7	3,8	4,1	7,4
1982	6,7	7,3	14,3	3,4	11,4	6,9
1983	11,5	10	12,9	4,3	10,5	9,9
1984	8,8	6,3	12,8	3	7,6	8,9
1985	4,3	2,1	7	3,3	5,4	8,5
1986	8,8	13,6	4,6	9,5	9,3	4,4
1987	8,9	11,7	6	7,7	5,6	7,8
1988	5,5	7,8	4,3	3,5	2,4	7
1989	6	6,7	6,1	4,7	5,7	6,1
1990	9,5	7	12,4	4,8	9,2	6,4
1991	9,5	9,7	7,7	5,2	5,4	8,4
1992	4,9	6	4,6	7,2	3	8
1993	9,8	5,1	15,5	8	13,8	10,3
1994	9,2	13,9	9,1	6,1	13,5	9,5
1995	8,6	13,3	5,7	6,5	7,8	12,4
1996	6,5	6,1	4,7	5,8	11	7,6
1997	11,1	18,5	6,1	7,7	13,4	14,8
1998	77,6	118,4	47,5	98,7	86,1	9,7
1999	2	-5,3	5,2	6,5	3,9	11
2000	9,4	4	10,1	10,2	9,6	27,4
2001	12,6	12	13,6	8,1	8,9	17,4
2002	10	9,1	12,7	2,7	5,6	16,5
2003	5,1	-1,7	9,4	7,1	5,7	21,5
2004	6,4	6,4	7,4	4,9	4,8	17,5
2005	17,11	13,91	13,94	6,92	6,13	8,24
2006	6,6	12,94	4,83	6,84	5,87	8,13
2007	6,59	11,26	4,88	8,42	4,31	8,83
2008	11,06	16,35	10,92	7,33	7,96	6,66
2009	2,78	3,88	1,83	6	3,89	3,89
2010	6,96	15,64	4,08	6,51	2,19	3,29

Badan Pusat Statistik Republik Indonesia (Statistics Indonesia)
Jl. Dr. Sutomo 6-8 Jakarta 10710 Indonesia, Telp (62-21) 3841195, 3842508, 3810291, Faks (62-21) 3857046,
Mailbox : bpshq@bps.go.id

Lampiran 2. Syntak Regresi *Inverse Gaussian* dengan SAS 9.1.3

```
data regresi;
input inflasi bahanmakanan perumahan sandang kesehatan pendidikan;
datalines;
16      16.3   18.3   12.7   14.6   13.8
7.1      8     7.7    3.8    4.1    7.4
6.7     7.3    14.3   3.4    11.4   6.9
11.5    10     12.9   4.3    10.5   9.9
8.8     6.3    12.8   3      7.6    8.9
4.3     2.1    7      3.3    5.4    8.5
8.8     13.6   4.6    9.5    9.3    4.4
8.9     11.7   6      7.7    5.6    7.8
5.5     7.8    4.3    3.5    2.4    7
6       6.7    6.1    4.7    5.7    6.1
9.5     7      12.4   4.8    9.2    6.4
9.5     9.7    7.7    5.2    5.4    8.4
4.9     6      4.6    7.2    3      8
9.8     5.1    15.5   8      13.8   10.3
9.2     13.9   9.1    6.1    13.5   9.5
8.6     13.3   5.7    6.5    7.8    12.4
6.5     6.1    4.7    5.8    11     7.6
11.1    18.5   6.1    7.7    13.4   14.8
77.6    118.4  47.5   98.7   86.1   9.7
2       -5.3   5.2    6.5    3.9    11
9.4     4      10.1   10.2   9.6    27.4
12.6    12     13.6   8.1    8.9    17.4
10      9.1    12.7   2.7    5.6    16.5
5.1     -1.7   9.4    7.1    5.7    21.5
6.4     6.4    7.4    4.9    4.8    17.5
17.11   13.91  13.94  6.92   6.13   8.24
6.6     12.94  4.83   6.84   5.87   8.13
6.59    11.26  4.88   8.42   4.31   8.83
11.06   16.35  10.92  7.33   7.96   6.66
2.78    3.88   1.83   6      3.89   3.89
6.96    15.64  4.08   6.51   2.19   3.29
;

proc genmod data = regresi;
model inflasi = bahanmakanan perumahan sandang kesehatan pendidikan/dist = IG

      link = log

      type3;
run;
```

Lampiran 3. *Output* Regresi *Inverse Gaussian* dengan SAS 9.1.3

```
      The SAS System      09:34 Thursday, March 31, 2011

               The GENMOD Procedure

                 Model Information

                 Data Set                   WORK.REGRESI
                 Distribution           Inverse Gaussian
                 Link Function                       Log
                 Dependent Variable              inflasi


                 Number of Observations Read          31
                 Number of Observations Used          31


               Criteria For Assessing Goodness Of Fit

 Criterion                DF           Value        Value/DF

 Deviance                 25          0,1184          0,0047
 Scaled Deviance          25         31,0000          1,2400
 Pearson Chi-Square       25          0,0988          0,0040
 Scaled Pearson X2        25         25,8622          1,0345
 Log Likelihood                     -55,4536


 Algorithm converged.


                               Analysis Of Parameter Estimates

                     Standard Wald 95% Confidence Chi-Square Pr>Chi-Square
Parameter     DF Estimate Error             Limits
Intercept     1  0,7472   0,1211  0,5099    0,9845    38,09   <0,0001
bahanmakanan  1  0,0704   0,0065  0,0576    0,0832    116,49  <0,0001
perumahan     1  0,0548   0,0118  0,0317    0,0779    21,67   <0,0001
sandang       1 -0,0145   0,0189 -0,0516    0,0226    0,59     0,4439
kesehatan     1  0,0130   0,0126 -0,0118    0,0377    1,05     0,3054
pendidikan    1  0,0199   0,0074  0,0053    0,0344    7,15     0,0075
Scale         1  0,0618   0,0078  0,0482    0,0793

  NOTE: The scale parameter was estimated by maximum likelihood.
```

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
bahanmakanan	1	43.15	<0,0001
perumahan	1	16,33	<0,0001
sandang	1	0,58	0,4468
kesehatan	1	1,07	0,3020

Lampiran 4. Tabel Khi-Kuadrat (χ^2_α)

TABLE C: Chi-Square distributions

cum probability	0.025	0.80	0.90	0.95	0.975	0.99	0.995	0.999	0.9995
right tail	0.975	0.2	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
df									
1	0.00098	1.64	2.71	3.84	5.02	6.63	7.88	10.83	12.12
2	0.051	3.22	4.61	5.99	7.38	9.21	10.60	13.82	15.20
3	0.216	4.64	6.25	7.81	9.35	11.34	12.84	16.27	17.73
4	0.48	5.99	7.78	9.49	11.14	13.28	14.86	18.47	20.00
5	0.83	7.29	9.24	11.07	12.83	15.09	16.75	20.51	22.11
6	1.24	8.56	10.64	12.59	14.45	16.81	18.55	22.46	24.10
7	1.69	9.80	12.02	14.07	16.01	18.48	20.28	24.32	26.02
8	2.18	11.03	13.36	15.51	17.53	20.09	21.95	26.12	27.87
9	2.70	12.24	14.68	16.92	19.02	21.67	23.59	27.88	29.67
10	3.25	13.44	15.99	18.31	20.48	23.21	25.19	29.59	31.42
11	3.82	14.63	17.28	19.68	21.92	24.73	26.76	31.26	33.14
12	4.40	15.81	18.55	21.03	23.34	26.22	28.30	32.91	34.82
13	5.01	16.98	19.81	22.36	24.74	27.69	29.82	34.53	36.48
14	5.63	18.15	21.06	23.68	26.12	29.14	31.32	36.12	38.11
15	6.26	19.31	22.31	25.00	27.49	30.58	32.80	37.70	39.72
16	6.91	20.47	23.54	26.30	28.85	32.00	34.27	39.25	41.31
17	7.56	21.61	24.77	27.59	30.19	33.41	35.72	40.79	42.88
18	8.23	22.76	25.99	28.87	31.53	34.81	37.16	42.31	44.43
19	8.91	23.90	27.20	30.14	32.85	36.19	38.58	43.82	45.97
20	9.59	25.04	28.41	31.41	34.17	37.57	40.00	45.31	47.50
21	10.28	26.17	29.62	32.67	35.48	38.93	41.40	46.80	49.01
22	10.98	27.30	30.81	33.92	36.78	40.29	42.80	48.27	50.51
23	11.69	28.43	32.01	35.17	38.08	41.64	44.18	49.73	52.00
24	12.40	29.55	33.20	36.42	39.36	42.98	45.56	51.18	53.48
25	13.12	30.68	34.38	37.65	40.65	44.31	46.93	52.62	54.95
30	16.79	36.25	40.26	43.77	46.98	50.89	53.67	59.70	62.16
40	24.43	47.27	51.81	55.76	59.34	63.69	66.77	73.40	76.10
50	32.36	58.16	63.17	67.50	71.42	76.15	79.49	86.66	89.56
60	40.48	68.97	74.40	79.08	83.30	88.38	91.95	99.61	102.7
80	57.15	90.41	96.58	101.9	106.6	112.3	116.3	124.8	128.3
100	74.22	111.7	118.5	124.3	129.6	135.8	140.2	149.4	153.2