ANALISIS PEUBAH RESPONS KONTINU NON NEGATIF
DENGAN REGRESI GAMMA DAN REGRESI *INVERSE GAUSSIAN* [1]

Kismiantini
Jurusan Pendidikan Matematika, FMIPA Universitas Negeri Yogyakarta
Email : kismi@uny.ac.id

Abstrak

Peubah respons kontinu pada bidang asuransi seperti peubah waktu antara pelaporan klaim dengan penyelesaian biasanya merupakan bilangan non-negatif dan mempunyai ukuran kemiringan ke kanan. Sehingga apabila ingin mengetahui hubungan antara peubah respons (peubah tak bebas) kontinu non negatif dengan peubah bebas maka regresi linear klasik tidak tepat digunakan. Pada makalah ini akan mengkaji alternatif regresi yang mampu mengatasi permasalahan peubah respons kontinu non negatif dengan menggunakan regresi gamma dan regresi *inverse gaussian*.

Kata-kata kunci: peubah non kontinu, regresi gamma, regresi *inverse gaussian*.

1. Pendahuluan

Pada bidang asuransi, peubah yang menjadi pangamatan seperti laju klaim, waktu antara pelaporan klaim dengan penyelesaian dan rata-rata biaya klaim merupakan peubah respons kontinu non negatif serta mempunyai tingkat ukuran kemiringan yang cenderung ke kanan. Sehingga regresi linear klasik tidak tepat digunakan dalam permasalahan tersebut. Pada regresi linear klasik, peubah respons diasumsikan sebagai peubah kontinu yang berdistribusi normal sehingga nilai $-\infty < Y < \infty$.

Alternatif regresi yang dapat digunakan untuk mengatasi permasalahan peubah respons non kontinu adalah regresi gamma dan regresi *inverse Gaussian* (Jong & Heller, 2008). Regresi ini dipilih dengan alasan bahwa memuat nilai peubah respons non negatif dan termasuk dalam dalam keluarga eksponensial sehingga model linear terampat (*Generalized Linear Model*) dapat digunakan untuk menyelesaikan permasalahan tersebut dan tidak memerlukan pemenuhan asumsi seperti pada model regresi linear klasik.

2. Keluarga Eksponensial

Peubah pengamatan Y diasumsikan memiliki fungsi peluang keluarga eksponensial yang dapat dimodelkan sebagai berikut (Jong & Heller, 2008):

$$f(y|\theta,\phi) = c(y,\phi)\exp\left(\frac{y\theta - a(\theta)}{\phi}\right) \tag{1}$$

dengan θ dan ϕ adalah parameter kanonik dan parameter dispersi dan $a(\phi)$, $b(\theta)$ dan $c(y, \phi)$ adalah suatu fungsi yang diketahui.

[1] Makalah ini disampaikan pada Seminar Nasional Penelitian, Pendidikan dan Penerapan MIPA yang diselenggarakan oleh FMIPA Universitas Negeri Yogyakarta pada tanggal 16 Mei 2009

Fungsi log-likelihood adalah log $f(y|\theta,\phi)$, ditulis sebagai $l(\theta,\phi|y)$, merupakan suatu fungsi dari θ dan ϕ dengan y diketahui. Nilai harapan dan ragam Y dapat ditentukan dengan mengevaluasi fungsi turunan dari $l(\theta,\phi|y)$ yang bersesuaian yaitu :

$$E(Y) = \mu = b'(\theta) \text{ dan } Var(Y) = \sigma^2 = b''(\theta)a(\phi)$$

dengan $b'(\theta)$ dan $b''(\theta)$ adalah turunan pertama dan kedua dari $b(\theta)$.

Beberapa distribusi peluang yang termasuk dalam distribusi keluarga eksponesial adalah distribusi binomial, Poisson, normal, gamma, *inverse Gaussian* dan *negative binomial*.

Tabel 1. Distribusi keluarga eksponensial dan parameternya

Distribusi	θ	$a(\theta)$	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, $G(\mu,\nu)$	$-\frac{1}{\mu}$	$-\ln(-\theta)$	$\frac{1}{\nu}$	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	$\mu(1+\kappa\mu)$

3. Model Linear Terampat (*Generalized Linear Model*)

Bila peubah respons Y tidak lagi mengikuti sebaran normal namun seperti Gamma atau *Inverse Gaussian* (asalkan termasuk dalam keluarga eksponensial) dan ragam Y merupakan fungsi dari nilai tengahnya sehingga dapat dipastikan bahwa ragam tidak homogen maka digunakanlah suatu model yang disebut model linear terampat (*Generalized Linear Model/GLM*). *GLM* merupakan pengembangan dari model linear klasik dengan peubah respons Y merupakan suatu komponen yang bebas dengan nilai tengah μ.

Ada tiga komponen utama dalam *GLM* (McCullagh & Nelder, 1989):

1. Komponen acak, yaitu komponen dari Y yang bebas dan fungsi sebaran peluang Y termasuk dalam keluarga sebaran eksponensial dengan $E(Y) = \mu$.
2. Komponen sistematik, yaitu $X_1, X_2, \ldots, X_p$ yang menghasilkan penduga linear η dimana $\eta = \beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p$.
3. Fungsi penghubung (*link function*) $g(.)$, menggambarkan hubungan antara penduga linear η dengan nilai tengah μ. Hubungan ini dapat ditulis dengan $\eta = g(\mu)$.

Dalam model linear klasik, komponen (1) menyebar normal dan komponen (3) merupakan fungsi identitas. Sedangkan dalam *GLM*, komponen (1) mungkin berasal dari salah satu anggota keluarga sebaran eksponensial lainnya dan komponen (3) merupakan fungsi monoton lainnya. Berikut Tabel 2 tentang fungsi hubung kanonik untuk berbagai distribusi.

Tabel 2. Fungsi hubung

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

4. Pendugaan Parameter pada *GLM*

Pendugaan parameter $\boldsymbol{\beta}$ dan ϕ dapat dilakukan dengan metode *Maximum Likelihood Estimation* (*MLE*) yaitu dengan memaksimumkan fungsi log-likehood sebagai berikut :

$$\ell(\beta,\phi)=\sum_{i=1}^{n}\ln f\left(y_i;\beta,\phi\right)=\sum_{i=1}^{n}\left\{\ln c\left(y_i,\phi\right)+\frac{y_i\theta_i-a(\theta_i)}{\phi}\right\} \qquad (2)$$

dengan mengasumsikan bahwa y_i adalah peubah respons keluraga eksponensial yang bebas.

Lalu *MLE* untuk masing-masing parameter β_j adalah dengan menurunkan $\ell(\beta,\phi)$ terhadap β_j yaitu

$$\frac{\partial \ell}{\partial \beta_j}=\sum_{i=1}^{n}\frac{\partial \ell}{\partial \theta_i}\frac{\partial \theta_i}{\partial \beta_j},$$

dengan

$$\frac{\partial \ell}{\partial \theta_i}=\frac{y_i-a(\theta_i)}{\phi}=\frac{y_i-\mu_i}{\phi}, \quad \frac{\partial \ell}{\partial \beta_j}=\frac{\partial \theta_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \beta_j}=\frac{\partial \theta_i}{\partial \eta_i}x_{ij}$$

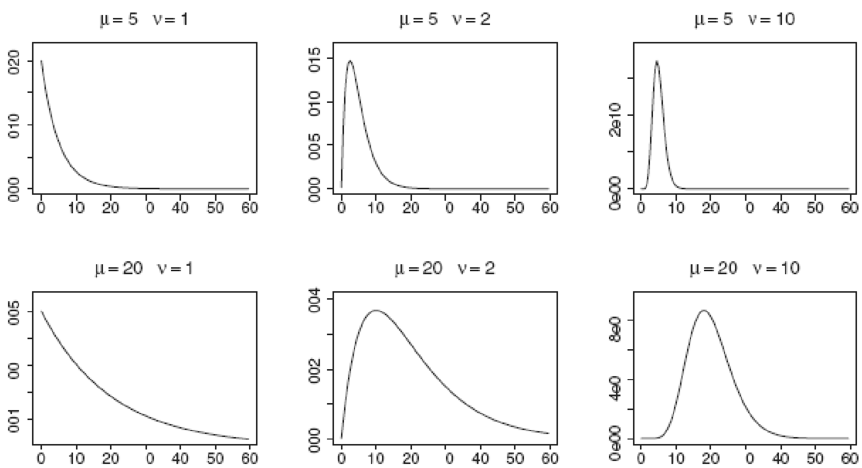
diketahui bahwa $\eta_i = x_i'\beta$ dan x_{ij} adalah komponen i dari x_j, kemudian $\partial\ell/\partial\beta_j = 0$ (yang merupakan kondisi orde pertama dalam memaksimumkan fungsi *likelihood*) maka akan didapatkan penduga dari $\boldsymbol{\beta}$ dan ϕ. Namun pada kondisi order pertama tersebut biasanya sulit diperoleh secara langsung kecuali untuk kasus normal dengan fungsi hubung identitas. Untuk mengatasi hal tersebut dapat digunakan dengan iterasi Newton-Raphson (Jong & Heller, 2008).

5. Regresi Gamma

Peubah respons dalam model regresi gamma merupakan peubah kontinu non negatif ($Y_i > 0$) dengan ukuran kemiringan ke kanan yang ditentukan oleh besarnya ν dan fungsi padat peluangnya sebagai berikut $Y_i \sim Gamma(\mu_i, \nu)$, (McCullagh & Nelder, 1989):

$$f(y_i) = \frac{1}{\Gamma(\nu) y_i} \left(\frac{\nu y_i}{\mu_i} \right)^{\nu} \exp\left(-\frac{\nu y_i}{\mu_i} \right)^{i}, \; y_i > 0, \nu > 0, \mu_i > 0, \; i = 1, 2, \ldots, n \qquad (3)$$

Berdasarkan fungsi padat peluang tersebut diperoleh $E(Y_i) = \mu_i$ dan $Var(Y_i) = \mu_i^2 / \nu = \mu_i^2 \sigma^2$. Berikut Gambar 1 tentang parameter ν menyatakan bentuk distribusinya.



Gambar 1. Distribusi gamma

Pada model regresi gamma, parameter $\nu = \sigma^{-2}$ diasumsikan konstan untuk semua pengamatan, sehingga fungsi densitasnya mempunyai bentuk yang sama untuk semua pengamatan.

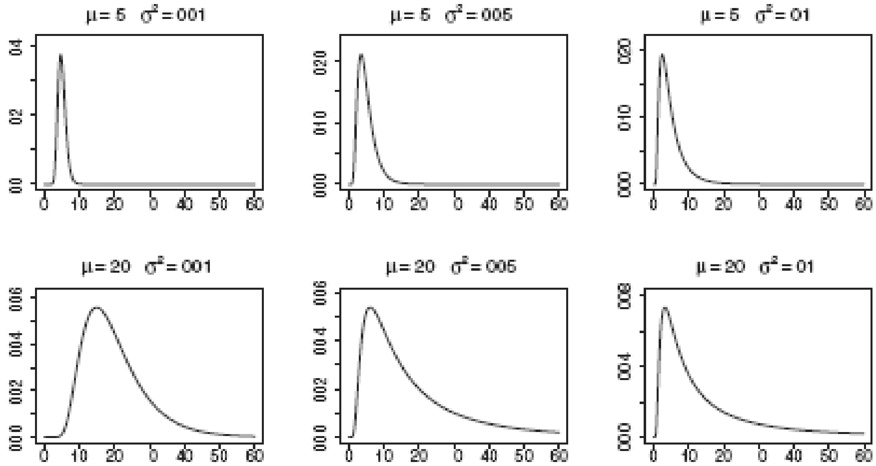
Model regresi gamma merupakan suatu model yang tergolong dalam model linear terampat (*Generalized Linear Model*), dengan mengasumsikan bahwa peubah respons $Y_i \sim Gamma(\mu_i, \nu)$, $E(Y_i) = \mu_i = \mathbf{X}'\boldsymbol{\beta}$ dengan fungsi hubung adalah $g(\mu_i) = \mu_i^{-1}$, kadangkala juga menggunakan fungsi hubung log natural (Czado, 2004).

6. Regresi *Inverse Gaussian*

Peubah respons Y_i pada model regresi *inverse Gaussian* juga merupakan peubah non negatif dengan fungsi padat peluang sebagai berikut $Y_i \sim IG(\mu_i, \sigma^2)$, (Dinh *et al.*, 2003) :

$$f(y_i) = \frac{1}{\sigma \sqrt{2\pi y_i^3}} \exp\left\{ -\frac{1}{2 y_i} \left(\frac{y_i - \mu_i}{\mu_i \sigma} \right)^2 \right\}, \; y_i > 0, \sigma > 0, \mu_i > 0 \qquad (4)$$

Nilai harapan dan ragam dari peubah respons Y_i adalah $E(Y_i) = \mu_i$, $Var(Y_i) = \sigma^2 \mu_i^3$. Parameter σ merupakan parameter bentuk.



Gambar 2. Distribusi *inverse Gaussian*

Model regresi *inverse Gaussian* ini juga tergolong dalam model linear terampat, dengan mengasumsikan peubah respons $Y_i \sim IG(\mu_i, \sigma^2)$, $E(Y_i) = \mu_i = \mathbf{X}'\boldsymbol{\beta}$ dan fungsi hubungnya adalah $g(\mu_i) = \mu_i^{-2}$, kadangkala juga menggunakan fungsi hubung log natural (Jong & Heller, 2008).

7. Uji *Goodness of Fit*

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

Hipotesis :

H_0 : Model layak

H_1 : Model tidak layak

Taraf nyata : α

Statistik Uji : Deviance (lihat Tabel 3)

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

n : banyaknya pengamatan, p : banyaknya parameter.

Perhitungan

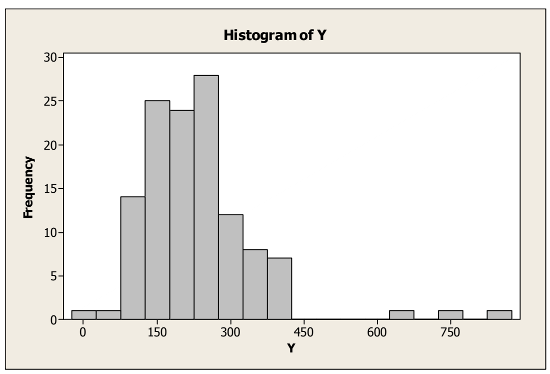
Kesimpulan : Jika nilai *deviance* $\leq \chi^2_{\alpha(n-p)}$, maka dapat disimpulkan bahwa model layak.

Tabel 3. *Deviance* untuk distribusi gamma dan *inverse Gaussian*

Distribusi	Deviance Δ
Gamma	$2\nu\sum_{i=1}^{n}\left\{-\ln\left(\frac{y_i}{\hat{\mu}_i}\right)+\frac{y_i-\hat{\mu}_i}{\hat{\mu}_i}\right\}$
Inverse Gaussian	$\frac{1}{\sigma^2}\sum_{i=1}^{n}\frac{(y_i-\hat{\mu}_i)^2}{\hat{\mu}_i^2 y_i}$

8. Ilustrasi Regresi Gamma dan Regresi *Inverse Gaussian*

Sebagai ilustrasi digunakan data yang diambil dari Baxter *et al.* (1980) dalam McCullagh & Nelder (1989) yaitu tentang rata-rata biaya klaim kepemilikan mobil pribadi yang rusak bila kendaraan tersebut diasuransikan. Peubah pengamatan (respons) adalah rata-rata biaya klaim, sedangkan peubah-peubah penjelas yang diamati adalah umur pemegang polis (terdiri 8 golongan yaitu umur 17-20, 21-24, 25-29, 30-34, 35-39, 40-49, 50-59, 60+ tahun), kelompok mobil (terdiri 4 level yaitu A, B, C dan D), umur kendaraan (terdiri 4 level yaitu 0-3, 4-7, 8-9, 10+ tahun). Data ini ada sebanyak 123 pengamatan.



Gambar 3. Histogram peubah respons (rata-rata biaya klaim)

Berdasarkan Gambar 3 diperoleh bahwa data tentang rata-rata biaya klaim bernilai non negatif dan memiliki kemiringan ke kanan sehingga peluang bahwa model regresi gamma dan model regresi *inverse Gaussian* layak digunakan cukup besar.

Analisis menggunakan SAS 9.1 meliputi: PROC GENMOD dengan fungsi hubung log natural untuk mendapatkan $\hat{\boldsymbol{\beta}}$ dari model regresi gamma dan model regresi *inverse Gaussian*. Pada kasus ini digunakan fungsi hubung log natural karena apabila digunakan fungsi hubung *default* (μ_i^{-2}) untuk model regresi *inverse Gaussian* maka penduga dari koefisien regresinya tidak dapat diperoleh. Pada SAS yang dimaksud nilai *deviance* adalah nilai *scalled deviance* (Jong & Heller, 2008). Berikut beberapa hasil dari ouput SAS 9.1.

Persamaan regresi gamma dugaan :

$$\hat{\mu} = \exp\begin{pmatrix} 5.1649 + 0.2836X_1 + 0.1754X_2 + 0.2888X_3 - 0.0046X_4 - 0.0475X_5 + 0.0028X_6 \\ + 0.0454X_7 - 0.4723X_8 - 0.4144X_9 - 0.3175X_{10} + 0.7355X_{11} + 0.6241X_{12} + 0.3132X_{13} \end{pmatrix}$$

Persamaan regresi *inverse Gaussian* dugaan :

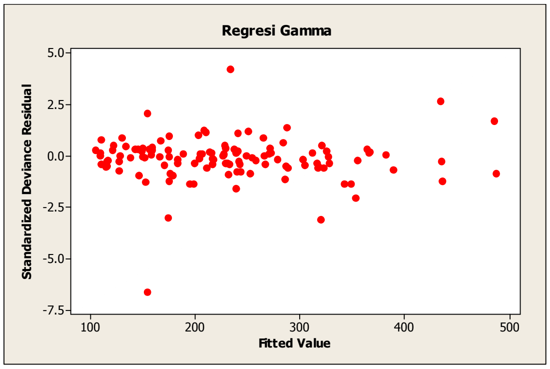
$$\hat{\mu} = \exp\begin{pmatrix} 5.1113 + 0.1919X_1 + 0.1821X_2 + 0.3556X_3 - 0.0242X_4 - 0.0078X_5 - 0.0026X_6 \\ + 0.0588X_7 - 0.4388X_8 - 0.3924X_9 - 0.2897X_{10} + 0.7718X_{11} + 0.6592X_{12} + 0.3590X_{13} \end{pmatrix}$$

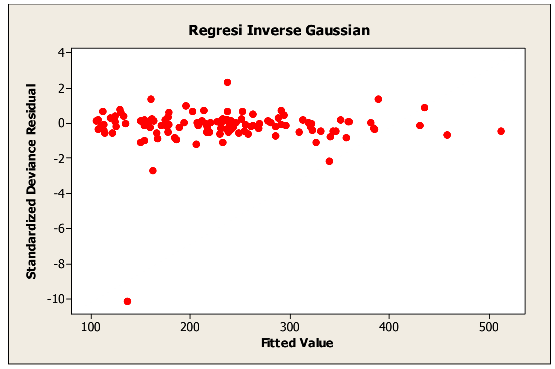
dengan $X_1, X_2, \ldots, X_7$ merupakan peubah indikator dari umur pemegang polis, X_8, X_9, X_{10} peubah indikator dari kelompok mobil dan X_{11}, X_{12}, X_{13} peubah indikator dari umur kendaraan.

Tabel 4. Ringkasan output SAS 9.1 dengan fungsi hubung log

Parameter	db	$\hat{\beta}$		*Standard Error*		χ^2		*p-value*	
		Gamma	IG	Gamma	IG	Gamma	IG	Gamma	IG
Intersep	1	5.1649	5.1113	0.1051	0.1530	2413.0300	1115.8000	<.0001	<.0001
Umur Pemegang Polis									
17-20	1	0.2836	0.1919	0.1129	0.1666	6.3100	1.3300	0.0120	0.2494
21-24	1	0.1754	0.1821	0.1081	0.1561	2.6300	1.3600	0.1047	0.2433
25-29	1	0.2888	0.3556	0.1069	0.1648	7.3000	4.6500	0.0069	0.0310
30-34	1	-0.0046	-0.0242	0.1063	0.1452	0.0000	0.0300	0.9658	0.8675
35-39	1	-0.0475	-0.0078	0.1080	0.1485	0.1900	0.0000	0.6602	0.9584
40-49	1	0.0028	-0.0026	0.1062	0.1457	0.0000	0.0000	0.9793	0.9859
50-59	1	0.0454	0.0588	0.1062	0.1481	0.1800	0.1600	0.6692	0.6911
60+ tahun	0	0	0	0	0	.	.	.	.
Kelompok Mobil									
A	1	-0.4723	-0.4388	0.0781	0.1225	36.5600	12.8400	<.0001	0.0003
B	1	-0.4144	-0.3924	0.0782	0.1228	28.1000	10.2100	<.0001	0.0014
C	1	-0.3175	-0.2897	0.0789	0.1270	16.2100	5.2000	<.0001	0.0226
D	0	0	0	0	0	.	.	.	.
Umur Kendaraan									
0-3	1	0.7355	0.7718	0.0788	0.1156	87.1700	44.5400	<.0001	<.0001
4-7	1	0.6241	0.6592	0.0789	0.1123	62.4900	34.4500	<.0001	<.0001
8-9	1	0.3132	0.3590	0.0791	0.1023	15.6900	12.3200	<.0001	0.0004
10+	0	0	0	0	0	.	.	.	.
Goodness of fit	*DF*	*Value*		*Value/DF*					
		Gamma	IG	Gamma	IG				
Deviance	109	124.8483	123.0000	1.1454	1.1284				

Berdasarkan Tabel 4, secara keseluruhan uji bagi masing-masing parameter (dilihat dari *p-value*) memberikan hasil yang sama bila dipilih $\alpha = 0.05$ untuk model regresi gamma dan model regresi *inverse Gaussian*, kecuali pada parameter umur pemegang polis antara 17-20 tahun. *Goodness of fit* dengan kriteria *deviance* diperoleh nilai *Value/DF* untuk masing-masing model lebih kecil dari $\chi^2_{0,05(109)} \approx 140.2$ sehingga dapat dikatakan bahwa model regresi gamma maupun model *regresi inverse Gaussian* layak digunakan. Demikian pula dari Gambar 4, titik-titik pengamatan tidak membentuk suatu pola tertentu (acak) untuk kedua plot, sehingga dapat diambil kesimpulan bahwa model regresi gamma maupun model *regresi inverse Gaussian* layak digunakan.





Gambar 4. Plot nilai dugaan dengan sisaan *deviance* dibakukan

9. Simpulan

Model regresi gamma dan model regresi *inverse Gaussian* sama baiknya (dengan fungsi hubung log natural) dalam menangani permasalahan peubah respons kontinu non negatif dan mempunyai ukuran kemiringan ke kanan

.

10. Daftar Pustaka

Czado, C. 2004. Gamma regression. [terhubung berkala]. http://www.m4.ma.tum.de/courses/GLM/lec8.pdf [3 Desember 2008].

Dinh, K.T., Nguyen N.T., & Nguyen, T.T. 2003. A regression characterization of inverse Gaussian distributions and application to edf goodness-of-fit test. *IJMMS* 9: 587-592.

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

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

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