

ANALISIS REGRESI TOBIT KUANTIL BAYESIAN PADA DATA TERSENSOR PASIEN PENDERITA PENYAKIT KANKER PAYUDARA

Oleh: Novi Destiyanto¹ Moh.Yamin Darsyah² Abdul Karim³
^{1,2,3}Program Studi Statistika, Universitas Muhammadiyah Semarang
e-mail: novi.destianto3@gmail@gmail.com

Article history

Submission: 17 September 2020
Revised: 23 September 2020
Accepted: 24 September 2020

Keyword:

Kata Kunci: data tersensor, regresi tobit, regresi Kuantil, Tobit Kuantil Bayesian (TKB), MetropolisHastings, pasien pengidap kanker payudara

Abstract

Analisis regresi berkaitan dengan studi mengenai ketergantungan satu variabel, yaitu variabel dependen, terhadap satu atau lebih variabel lainnya, yaitu variabel penjelas, dengan tujuan untuk mengestimasi dan memperkirakan nilai rerata atau rata-rata (populasi) variabel dependen dari nilai yang diketahui atau nilai tetap dari variabel penjelas (dalam sampling berulang). Estimasi parameter model regresi Tobit umumnya menggunakan metode Maximum Likelihood Estimation (MLE) yang berbasis conditional mean, sehingga data yang jauh dari mean, misalnya kuantil 0.05 dan 0.95, seringkali tidak terepresentasi dengan baik oleh estimator yang ada. Hal ini dapat diatasi dengan model regresi kuantil tobit atau regresi kuantil tersensor yang berbasis conditional quantile. Estimator model ini dikenal sebagai estimator Powell. Selain itu, diperkenalkan model regresi tobit kuantil dengan pendekatan bayesian (TKB). Kelebihan pendekatan bayesian adalah kemudahan untuk mengestimasi distribusi posterior dengan teknik MCMC (Metropolis-Hastings), serta mampu mengakomodasi adanya informasi prior. Untuk mendapatkan estimator TKB, diasumsikan error model mengikuti distribusi Asimetrik Laplace, sehingga bisa didapatkan fungsi likelihood untuk menghitung posterior. Dari simulasi yang telah dilakukan, diperoleh hasil bahwa estimator TKB dan Powell tidak cukup baik untuk mengestimasi parameter model kuantil bawah seperti kuantil 0.05 dan 0.25. Selain itu, jika prediktor dalam model cukup banyak dan ukuran sampel yakni 50, performa estimator TKB lebih baik daripada estimator Powell. Di samping itu, bila dibandingkan dengan estimator tobit standar, estimator TKB dan Powell memiliki performa yang lebih baik untuk model dengan error berdistribusi bukan normal. Adapun pemodelan regresi TKB terhadap data pasien pengidap kanker payudara menunjukkan bahwa besarnya estimator yang dihasilkan bervariasi antar kuantil dan model terbaik untuk masing-masing kuantil juga melibatkan prediktor yang berbeda.

1. Pendahuluan

Ilmu statistika sering digunakan dalam berbagai bidang ilmu pekerjaan misalnya pada bidang ekonomi, keuangan, pariwisata, geologi, oceanografi, sosial, politik, industri, pendidikan, dan kesehatan. Terdapat beberapa model regresi, namun pembedanya sendiri dilihat berdasarkan variabel terikat Y yang digunakan. Istilah regresi pertama kali dikenalkan oleh Francis Galton. Pada karya tulisnya yang terkenal, Galton menemukan bahwa adanya hubungan tinggi badan anak dengan tinggi badan orang tuanya.

Menurut Gujarati (2013), analisis regresi berkaitan dengan studi mengenai ketergantungan satu variabel, yaitu variabel dependen, terhadap satu atau lebih variabel lainnya, yaitu variabel penjelas dengan tujuan untuk mengestimasi dan/atau memperkirakan nilai rerata atau rata-rata (populasi) variabel dependen dari nilai yang diketahui atau nilai tetap dari variabel penjelas (dalam sampling berulang). Pada analisis regresi linier klasik, umumnya variabel terikat yang digunakan hanya berdistribusi diskrit atau kontinu saja. Namun kenyataan di lapangan banyak ditemukan kasus dimana variabel terikat mempunyai distribusi gabungan diskrit dan kontinu. Variabel terikat yang mempunyai sifat gabungan diskrit dan kontinu ini dinamakan data tersensor (Greene, 2012). Dalam pengertian yang lebih luas, data tersensor merupakan data dimana informasi tentang variabel terikat yang diteliti tidak memberikan informasi yang lengkap (Gujarati, 2013). Analisis regresi klasik tidak dapat digunakan dalam kasus seperti ini, karena akan terjadi pelanggaran asumsi-asumsi regresi dan bias parameter. Salah satu jenis regresi lain yang dapat digunakan untuk mengestimasi data tersensor adalah regresi tobit (Tobin, 1958).

Regresi tobit pertama kali dikemukakan oleh James Tobin pada tahun 1958, pada analisis regresi tobit, estimasi yang digunakan bukan estimasi *Ordinary Least Square* (OLS). Metode OLS tidak dapat mengestimasi parameter regresi tobit karena jenis data yang digunakan adalah data tersensor. Jika dipaksakan mengestimasi menggunakan metode ini, maka akan terjadi pelanggaran asumsi regresi dan bias parameter. Estimasi yang dapat digunakan untuk menganalisis regresi tobit adalah *Maximum Likelihood Estimation* (MLE).

Menurut Greene (2008) menyebutkan bahwa variabel terikat yang bersifat campuran

(mixture) memiliki struktur data dengan skala diskrit untuk yang bernilai nol, dan berskala kontinu untuk yang tidak bernilai nol. Data tersebut disebut juga data tersensor. Sebaran data tersensor adalah sebaran normal tersensor, yang mengikuti asumsi $N(\mu, \sigma^2)$. Penggunaan regresi tobit (regresi tersensor) pada data campuran akan mengurangi efek bias jika dibandingkan dengan data yang diolah menggunakan regresi linier klasik.

Regresi kuantil adalah salah satu metode analisis regresi yang dapat menggambarkan hubungan satu atau beberapa variabel prediktor terhadap satu variabel respon pada berbagai titik kuantil (conditional quantile) dari distribusi variabel respon tersebut, sehingga metode ini dapat digunakan pada kondisi data yang heterogen. Hal ini berbeda dengan analisis regresi linier yang hanya dapat menggambarkan hubungan sebab-akibat pada mean (conditional mean) variabel respon (Koenker dan Hallock, 2001). Persamaan model regresi kuantil menunjukkan bentuk umum model regresi kuantil linier (Buhai, 2005).

Menurut Yu dan Moyeed (2001) regresi kuantil Bayesian prinsip dasar pemodelan dengan pendekatan Bayesian adalah mendapatkan distribusi posterior dari suatu parameter bila diketahui distribusi prior dan fungsi likelihood yang sesuai dengan kaidah Bayes. Apabila pendekatan ini digunakan untuk model regresi kuantil, maka distribusi posterior $\beta(\theta)$ yakni $\pi(\beta(\theta)|y)$ dapat dirumuskan sebagai berikut.

$$\pi(\beta(\theta)|y) \propto L(y|\beta(\theta))\pi(\beta(\theta))\pi(\beta(\theta))$$
 merupakan distribusi prior dari $\beta(\theta)$ dan $L(y|\beta(\theta))$ adalah fungsi likelihood data dengan asumsi residual berdistribusi *asimetris Laplace*.

MCMC adalah metode umum yang digunakan untuk mendapatkan suatu nilai (sampel) β dari distribusi peluang yang diketahui $\pi(\beta(\theta))$, lalu nilai yang didapatkan tersebut dikoreksi sehingga bisa lebih mendekati distribusi posterior yang dikehendaki $\pi(\beta|y)$. Sampel-sampel β didapatkan secara berurutan (sequentially), atau dengan kata lain distribusi peluang yang digunakan untuk menghasilkan sampel saat ini tergantung dari sampel yang dihasilkan sebelumnya, sehingga membentuk suatu rantai Markov (Gelman et al., 2004).

Bayes faktor pertama kali diperkenalkan oleh Harold Jeffreys pada tahun 1960 sebagai alternatif dari pengujian hipotesis frekuentis. Statistik ini sering digunakan untuk membandingkan beberapa model guna mendapatkan model terbaik (Lavine dan Schervish, 1999). Jika digunakan hipotesis $M_1 > M_2$, maka interpretasi Bayes faktor dapat dijelaskan seperti pada Tabel 2.1 (Kass dan Raftery, 1995).

Menurut Kass dan Raftery (1995) menyatakan bahwa perhitungan Bayes faktor cukup sulit untuk dilakukan. Namun, Chib dan Jeliakoz (2001) memberikan gagasan mereka untuk menghitung estimasi logaritma marginal likelihood, $\log m(y|M_1)$, untuk sembarang nilai β_l^* yang memiliki densitas tinggi.

RMSE (Root Mean Square Error) merupakan salah satu statistik yang sering digunakan untuk mengevaluasi kebaikan performa model atau estimator. Statistik ini mengukur selisih antara nilai yang diprediksi oleh suatu model/estimator dengan nilai sebenarnya, yang disebut juga sebagai error atau residual. Persamaan berikut menunjukkan formulasi RMSE jika diasumsikan terdapat sebanyak n error model (Chai dan Draxler, 2014).

Kanker payudara terjadi karena adanya kerusakan pada gen yang mengatur pertumbuhan dan differensiasi sehingga sel itu tumbuh dan berkembang biak tanpa dapat dikendalikan. Penyebaran kanker payudara terjadi melalui pembuluh getah bening dan tumbuh di kelenjar getah bening, sehingga kelenjar getah bening aksila ataupun supraklavikula membesar. Kemudian melalui pembuluh darah kanker menyebar ke organ lain seperti paru-paru, hati dan otak. Kanker payudara ditemukan di seluruh dunia, tahun 2003, insiden kanker payudara di Belanda 91 per 100.000 penduduk, Amerika 71,7 per 100.000 penduduk, Swiss 70 per 100.000 wanita, Australia 83,2 per 100.000 penduduk, Kanada 84,7, Indonesia 26 per 100.000 wanita pada tahun 2002 dan Jepang 16 per 100.000 penduduk. dalam penelitian ini, metode yang akan digunakan. Kanker payudara lebih sering dijumpai pada umur 40-49 tahun yaitu sebesar 30,35%. Sehingga analisis seperti inilah yang sangat diperlukan untuk mengetahui seberapa lama pasien yang menderita kanker paru dapat bertahan hidup. Dan seberapa besar resiko kegagalan dalam bertahan hidup selama waktu yang diperkirakan

2. Metode Penelitian

2.1 Sumber Data

Penelitian ini menggunakan data skunder yang diperoleh dari RSUP Dr. Kariadi Semarang mengenai rata-rata waktu hidup pasien kanker payudara peluang bertahan hidup, dan peluang kegagalan pasien untuk bertahan secara statistika dengan menggunakan metode dokumentasi pasien yang mengidap penyakit kanker payudara RSUP Dr. Kariadi Semarang yang dimulai pada tanggal 25 Maret 2017 sampai dengan 25 Mei 2017. Data kanker yang diambil merupakan data pasien kanker payudara yang dibatasi dengan sensor tipe I, dengan waktu penyensoran yang dimulai dari tanggal 1 Januari 2013 sampai dengan 31 Mei 2017 dan diperoleh sebanyak 50 data pasien.

Langkah-langkah Penelitian

Metode dan tahapan analisis yang akan digunakan untuk mencapai tujuan penelitian adalah sebagai berikut:

1. Melakukan eksplorasi data terhadap semua variabel respon dan prediktor yang digunakan dalam penelitian.
2. Mengidentifikasi variabel untuk variabel terikat yang sebagian datanya memiliki skala pengukuran diskrit dan sebagian yang lain berskala kontinu
3. Melakukan estimasi parameter Regresi Tobit Kuantil Bayesian (TKB).
4. Menentukan Fungsi Likelihood Model Regresi Tobit Kuantil Bayesian Terbaik.
5. Melakukan Perbandingan Performa Estimator Tobit Kuantil Bayesian dan Powel.
6. Melakukan pemodelan Regresi Tobit Kuantil Bayesian bagi pasien pengidap kanker payudara.
7. Melakukan Pengujian Estimasi Parameter Regresi Tobit.

8. Melakukan pengujian Uji Beda karakteristik (pasien bertahan hidup (survive) atau ,meninggal).
9. Melakukan pengujian Uji Linieritas Variabel Prediktor.
10. Interpretasi Hasil Estimasi Parameter Regresi Tobit Kuantil Bayesian.
11. Pemilihan Model Regresi TKB Terbaik
12. Kesimpulan hasil dan saran
13. Selesai

Hasil Penelitian

2.2 Deskripsi Data Penelitian

Statistik deskriptif dari data pasien kanker payudara dengan status pasien yang masih bertahan (survival) dan pasien yang sudah meninggal dan digunakan dalam penelitian ini ditampilkan dalam tabel sebagai berikut.

Statistik	Status	Rata-rata waktu hidup pasien skala (0-10)
Minimum	Hidup	0.00
	Meninggal	0.00
Maksimum	Hidup	7.581
	Meninggal	3.523
Rata-rata	Hidup	8.868
	Meninggal	4.341
Standar Deviasi	Hidup	5.378
	Meninggal	2.786

Berdasarkan tabel dapat diketahui bahwa pasien yang mengidap kanker payudara baik yang masih bertahan hidup ataupun meninggal memiliki persentase masing-masing 54.68% dan 36.61%. Hal ini menunjukkan bahwa rata-rata waktu hidup pasien pengidap kanker lebih besar dibandingkan rata-rata kematian setelah dirawat dan diberikan perlakuan (*kemoteraphy*) di rumah sakit.

Statistik	Status	X ₁	X ₂	X ₃
Minimum	Hidup	82.00	1.12	1.00
	Meninggal	100.00	0.69	2.00
Maksimum	Hidup	8000.00	92.66	18.00
	Meninggal	3400.00	91.52	19.00
Rata-rata	Hidup	1076.81	50.62	4.85
	Meninggal	1780.51	44.65	5.45
Standar Deviasi	Hidup	1888.94	19.73	1.61
	Meninggal	2184.92	19.02	1.95

Berdasarkan tabel dapat diketahui bahwa rata-rata dan standar deviasi usia pasien pengidap kanker payudara X₁ saat masuk studi pengobatan untuk bertahan (*survival*) lebih rendah daripada angka kematian. Padahal, jarak antar nilai minimum dan maksimum usia untuk bertahan hidup dibandingkan meninggal lebih lebar. Rata-rata dan standar deviasi keadaan awal pasien kanker payudara saat awal masuk ke studi *performance status* (skor *karnofsky*) X₂ dalam kondisi masih hidup (survive) lebih besar daripada dinyatakan meninggal atau dengan kondisi yang sangat buruk. Selain itu, rata-rata bulan diagnosis seorang penderita X₃ dinyatakan mengidap kanker payudara untuk sembuh atau bertahan hidup lebih sedikit daripada diagnosis untuk meninggal.

2.3 Fungsi Likelihood Model Regresi Tobit Kuantil Bayesian Terbaik

Bentuk umum persamaan fungsi asimetris *Laplace* adalah sebagai berikut.

$$f_{\theta}(u) = \theta(1 - \theta) \exp(-\rho_{\theta}(u)) \quad , 0 < \theta < 1$$

Jika $u = \varepsilon_i$ dan $\tau = 0$, maka $u = y_i - maks(x_i^t \beta, 0)$, maka bentuk persamaan berubah menjadi.

$$f_{\theta}(\varepsilon_i) = \theta(1 - \theta) \exp\{-\rho_{\theta}(y_i - maks(x_i^t \beta, 0))\}$$

Jika sudah diketahui fungsi asimetris *Laplace* yang bersesuaian dengan model tobit, maka bisa didapatkan fungsi likelihood $L(y|\beta(\theta))$ sebagai berikut.

$$\begin{aligned}
 L(y|\beta(\theta)) &= \prod_{i=1}^n f_{\theta}(\varepsilon_i) \\
 &= \prod_{i=1}^n \theta(1 - \theta) \exp\{-\rho_{\theta}(y_i - maks(x_i^t \beta, 0))\} \\
 &= \theta^n (1 - \theta)^n \exp\{-\rho_{\theta}(y_i - maks(x_i^t \beta, 0))\}
 \end{aligned}
 \tag{4.4}$$

Distribusi posterior $\pi(\beta(\theta)|y)$ dapat dirumuskan sebagai berikut.

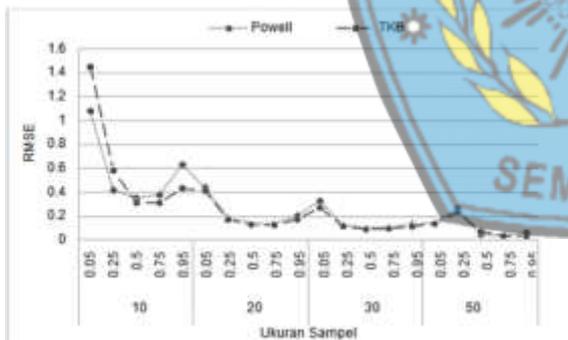
$$\pi(\beta(\theta)|y) \propto L(y|\beta(\theta))\pi(\beta(\theta)) \quad (4.5)$$

Estimator model TKB diperoleh dengan menghitung nilai mean dari distribusi posterior parameter .

2.4. Perbandingan Performa Estimator Tobit Kuantil Bayesian dan Powell

Banyaknya Prediktor	TKB					Powell				
	0.05	0.25	0.50	0.75	0.95	0.05	0.25	0.50	0.75	0.95
1	1.857	1.008	0.274	0.154	0.195	1.288	0.927	0.208	0.151	0.215
3	1.448	0.583	0.316	0.316	0.433	1.079	0.418	0.355	0.378	0.632
5	2.931	0.560	0.480	0.429	0.597	0.869	0.530	0.404	0.440	0.998
7	1.362	0.629	0.458	0.416	0.701	1.264	0.652	0.500	0.541	1.766

Tabel hanya menampilkan besarnya RMSE dari estimator intersep model tobit dengan jumlah prediktor yang berbeda. Berdasarkan tabel ini dapat dilihat bahwa secara umum performa estimator TKB lebih baik daripada Powell. Selain itu dapat pula diketahui bahwa semakin banyak prediktor yang digunakan, maka RMSE cenderung semakin besar. Atau dengan kata lain, performa estimator semakin menurun.



Gambar 4.1 RMSE Intersep (β_0) dari Estimator TKB dan Powell dengan Jumlah Sampel Berbeda ($p=3$)

2.5. Hasil Uji Beda Karakteristik Pasien Kanker Payudara

Variabel	Statistik uji <i>t</i>	P-value	Keterangan
Umur (X_1)	-11.959	0.000	Beda
Skor Karnofsky (X_2)	10.976	0.000	Beda
Bulan Diagnosa (X_3)	-11.683	0.000	Beda
Jenis Kanker (X_4)	8.234	0.000	Beda
Metode Pengobatan (X_5)	0.177	0.860	Tidak berbeda
Status	-5.714	0.000	Beda

Variabel	Statistik uji <i>chi-square</i>	P-value	Keterangan
Skor Karnofsky (X_2)	163.656	0.000	Beda
Status	73.186	0.000	Beda

Uji beda terhadap karakteristik pasien kanker payudara terhadap pasien yang masih bertahan hidup dan meninggal menggunakan uji *t* dan menggunakan uji independensi *chi-square*. Uji *t* digunakan untuk karakteristik yang bersifat kontinu, sedangkan uji *chi-square* untuk kategorik (diskrit). Hasil pengujian adalah sebagai berikut.

Berdasarkan tabel dapat diketahui bahwa hampir semua karakteristik/variabel antara status pasien yang masih bertahan hidup dan meninggal menunjukkan perbedaan, kecuali persentase metode pengobatan (X_5). Oleh karena itu, dapat dikatakan bahwa variabel wilayah dapat dijadikan sebagai variabel moderator dalam analisis lebih lanjut.

2.6. Hasil Pengujian Linieritas Variabel Prediktor

Hasil uji linieritas variabel prediktor kontinu terhadap variabel respon menggunakan uji RESET, sedangkan uji linieritas untuk prediktor dummy menggunakan uji korelasi *biserial point* disajikan dalam tabel berikut.

Variabel Prediktor	Status	Statistik Uji	P-value	Keterangan
Umur (X_1)	Hidup	7.077	0.008	Nonlinier
	Meninggal	48.131	0.000	Nonlinier
Skor Karnofsky (X_2)	Hidup	0.003	0.958	Linier
	Meninggal	3.683	0.055	Linier
Bulan Diagnosa (X_3)	Hidup	0.769	0.381	Linier
	Meninggal	0.000	0.639	Linier
Jenis Kanker (X_4)	Hidup	0.000	0.983	Linier
	Meninggal	0.644	0.422	Linier
Metode Pengobatan (X_5)	Hidup	3.558	0.059	Linier
	Meninggal	0.644	0.422	Linier
Status Pengobatan (X_{Status})	Hidup	6.201	0.013	Nonlinier
	Meninggal	24.397	0.000	Nonlinier

Dari tabel diatas, prediktor yang berhasil dilinierkan adalah variabel umur (X_1) dan variabel status (X_{Status}) melalui transformasi logaritma dengan nilai p-value masing-masing

adalah 0.317 dan 0.151. Hasil pengujian linieritas dengan transformasi selengkapnya dapat dilihat pada Lampiran 5. Variabel-variabel prediktor yang bersifat nonlinier terhadap respon tidak diikutsertakan dalam analisis lebih lanjut.

Variabel Prediktor	Kondisi	Korelasi Biserial	Statistik Uji t	P-value	Keterangan
(X_2) Skor Karnofsky	Hidup	-0.096	-3.824	0.000	Linier
	Meninggal	-0.022	-0.758	0.224	Nonlinier
(X_3) Bulan Diagnosa	Hidup	-0.178	-7.186	0.000	Linier
	Meninggal	-0.191	-6.847	0.000	Linier

Dalam tabel memperlihatkan bahwa variabel dummy untuk Skor Karnofsky X_2 terhadap status pasien yang masih bertahan (*survival*) yang menunjukkan hubungan yang linier dengan variabel respon. Di sisi lain, untuk status pasien yang meninggal, hasil uji korelasi Pearson untuk dummy Bulan diagnosa X_3 mendukung adanya hubungan linier dengan respon, sedangkan dummy Skor Karnofsky X_2 tidak. Karena variabel-variabel dummy ini pada dasarnya berasal dari satu variabel, maka keduanya tetap disertakan dalam pemodelan regresi Tobit Kuantil Bayesian.

2.7. Hasil Estimasi Parameter Regresi Tobit Kuantil Bayesian

Penelitian ini menggunakan data sekunder mengenai rata-rata waktu hidup pasien kanker payudara yakni peluang bertahan hidup pasien dan peluang kegagalan pasien (meninggal) yang terjadi secara alami tanpa dimanipulasi. Jumlah-jumlah sampel tersebut sangat sesuai untuk pemodelan regresi TKB, sebagaimana hasil simulasi yang telah dijabarkan pada subbab 4.2 sebelumnya. Estimasi parameter regresi Tobit Kuantil Bayesian dilakukan untuk lima titik kuantil yaitu 0.05, 0.25, 0.50, 0.75 dan 0.95. Adapun hasil estimasi ini disajikan dalam tabel.

Parameter	Estimator Tobit Kuantil Bayesian				
	Kuantil				
	0.05	0.25	0.50	0.75	0.95
β_0	-0.464	0.8988	-0.741	0.215	0.735
β_1	-2.633	0.568	0.400	0.862	3.056
β_2	1.288	-4.3081	0.460	0.594	0.775
β_3	-4.103	-2.2482	-0.034	0.801	-0.044
β_4	0.374	-0.3244	0.250	0.554	4.430
β_5	-2.479	-4.5692	-0.164	-0.054	-0.283
β_{status}	0.082	2.7235	1.860	2.129	8.068

Dapat diketahui bahwa estimator model regresi Tobit Kuantil Bayesian untuk pasien yang masih bertahan (*survival*) bervariasi antar

kuantil. Hal ini mengindikasikan bahwa penggunaan model regresi Tobit Kuantil Bayesian cukup tepat untuk memodelkan data tersebut. Besarnya estimator parameter untuk variabel Y (*time*), variabel X_1 (Umur), X_2 (Skor), X_3 (Bulan), X_4 (Kanker), X_5 (Metode Pengobatan), variabel Status. Variabel Y (*time*) adalah variabel *dependent* atau variabel terikat yang menunjukkan waktu hidup pasien kanker payudara (dalam hari). Variabel X_1 (Umur) adalah variabel *independent* yang menunjukkan umur pasien kanker payudara saat awal masuk studi (dalam tahun). Variabel X_2 (Skor) menunjukkan keadaan pasien kanker payudara saat awal masuk ke studi *performance status* (skor *karnofsky*). Variabel X_3 (Bulan) menunjukkan banyaknya bulan dari diagnosis awal (dalam bulan). Variabel X_4 (Kanker) menunjukkan klasifikasi kanker yang diderita dengan nilai satu (1) menunjukkan *adenocarcinoma*, dua (2) menunjukkan *skuamosa*, tiga (3) menunjukkan *small*, empat (4) menunjukkan besar. Variabel X_5 (Metode Pengobatan) menunjukkan metode yang digunakan pasien kanker payudara untuk berobat dengan nilai satu (1) menunjukkan kemoterapi dan dua (2) menunjukkan non kemoterapi.

Parameter	Estimator TKB Kuantil				
	0.05	0.25	0.50	0.75	0.95
β_0	-4.229	-0.043	0.624	-0.363	-1.021
β_1	-1.989	2.929	1.663	4.351	12.829
β_2	1.012	0.206	0.606	0.545	1.910
β_3	3.056	0.941	0.280	1.575	3.385
β_4	-5.158	-0.041	-0.062	-0.138	-0.388
β_5	2.668	0.481	0.101	1.818	4.854
β_{status}	-0.148	-0.068	-0.054	-0.213	-0.343

Berdasarkan tabel dapat diketahui bahwa perilaku estimator Tobit Kuantil Bayesian untuk pasien dengan status meninggal hampir mirip dengan pasien yang masih bertahan (*survival*) dalam artian bahwa estimator-estimator tersebut bervariasi antar kuantil. Estimator parameter untuk variabel umur (X_1) cenderung semakin meningkat seiring pertambahan nilai kuantil. Keadaan ini menunjukkan bahwa pengaruh variabel tersebut semakin besar untuk pengeluaran konsumsi susu rumah tangga di pedesaan yang semakin tinggi. Selain itu, estimator untuk variabel Skor Karnofsky (X_2), bulan diagnosa (X_3), jenis kanker (X_4), metode pengobatan (X_5) dan

status penyensoran (X_{Status}) menunjukkan bahwa pengaruh variabel-variabel tersebut berubah secara fluktuatif terhadap perubahan nilai kuantil.

Model Reduksi (M_i)	Kuantil Bayesian				
	0.05	0.25	0.50	0.75	0.95
(X_1) Umur	0.01 ^(*)	0.01 ^(*)	136.39 ^(**)	753.67 ^(**)	821.6 ^(**)
(X_2) Skor Karnofsky	0.01 ^(*)	0.01 ^(*)	14.29 ^(**)	192.49 ^(**)	49.21 ^(**)
(X_3) Bulan Diagnosa	0.01 ^(*)	0.01 ^(*)	-71.84 ^(*)	-26.33 ^(*)	-180.24 ^(*)
(X_4) Jenis kanker	0.01 ^(*)	0.01 ^(*)	222.32 ^(**)	307.48 ^(**)	521.86 ^(**)
(X_5) Metode Pengobatan	0.01 ^(*)	0.01 ^(*)	90.42 ^(**)	-52.03 ^(**)	144.32 ^(**)
(X_{Status}) Status penyensoran	0.01 ^(*)	0.01 ^(*)	209.27 ^(**)	830.28 ^(**)	1042.66 ^(**)

Legenda: Ketetapan pembatasan terhadap M_i , (-) negatif, (0) tidak ada, (+) positif, (*) kuat, (**) sangat kuat

Berdasarkan tabel dapat diketahui bahwa model terbaik kuantil 0.50, 0.75, dan 0.95 adalah model yang tidak mengikutsertakan prediktor bulan diagnosa (X_3). Adapun model terbaik kuantil 0.05 dan 0.25 merupakan model penuh. Hal yang menarik dari hasil ini adalah besarnya Bayes faktor untuk model pada kuantil 0.05 dan 0.25 yang menunjukkan nilai mendekati nol, atau dengan kata lain Bayes faktor tidak dapat memutuskan mana di antara model penuh dan model reduksi yang sesuai untuk memodelkan variabel respon. Namun demikian, dalam penelitian ini diputuskan bahwa apabila terdapat perbandingan dua model yang menghasilkan Bayes faktor mendekati nol, maka model yang digunakan adalah model penuh.

3. Kesimpulan

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan, dapat disimpulkan beberapa hal sebagai berikut

1. Fungsi likelihood pada persamaan berikut digunakan untuk memperoleh estimator model regresi TKB dengan pendekatan bayesian (teknik MCMC Metropolis-Hastings).

$$L(y|\beta(\theta)) = \theta(1 - \theta)\exp\{-\rho_\theta(y_i - \max(x_i^t \beta, 0))\}$$

2. Berdasarkan hasil simulasi, performa estimator regresi TKB lebih baik daripada estimator Powell ketika ukuran sampel

yang digunakan lebih dari 50 dan prediktor yang digunakan cukup banyak. Di sisi lain, kedua estimator tersebut ternyata tidak cukup baik untuk memodelkan kuantil bawah seperti kuantil 0.05. Selain itu, performa estimator TKB dan Powell lebih baik daripada estimator tobit standar untuk model yang memiliki error berdistribusi bukan normal.

3. Model terbaik regresi Tobit Kuantil Bayesian (TKB) bagi pasien pengidap kanker payudara yang masih bertahan (*survival*) ditunjukkan oleh persamaan-persamaan berikut.

$$\hat{y}_{(0.05)} = \begin{cases} -0.464 - 2.633X_1 + 1.288X_2 - 4.103X_3 + 0.374X_4 - 2.479X_5 - 0.082 \log(X_{Status})y' > 0 \\ 0 & y' > 0 \end{cases}$$

$$\hat{y}_{(0.25)} = \begin{cases} 0.899 + 0.568X_1 - 4.308X_2 - 2.248X_3 - 0.324X_4 - 14.569X_5 - 2.723 \log(X_{Status})y' > 0 \\ 0 & y' > 0 \end{cases}$$

$$\hat{y}_{(0.50)} = \begin{cases} -1.343 - 1.255X_1 + 2.552X_2 - 0.110X_3 - 0.356X_4 - 0.106X_5 + 2.986 \log(X_{Status})y' > 0 \\ 0 & y' > 0 \end{cases}$$

$$\hat{y}_{(0.75)} = \begin{cases} -0.634 - 0.520X_1 + 0.935X_2 + 0.616X_3 - 0.065X_4 - 0.081X_5 + 2.473 \log(X_{Status})y' > 0 \\ 0 & y' > 0 \end{cases}$$

$$\hat{y}_{(0.95)} = \begin{cases} -0.657 + 3.298X_1 + 4.753X_2 + 3.325X_3 - 0.453X_4 - 0.448X_5 + 11.116 \log(X_{Status})y' > 0 \\ 0 & y' > 0 \end{cases}$$

Adapun model regresi Tobit Kuantil Bayesian terbaik bagi pasien pengidap kanker payudara dengan status meninggal.

$$\hat{y}_{(0.05)} = \begin{cases} -4.229 - 1.989 \log(X_1) + 1.012X_2 + 3.056X_3 - 5.158X_4 + 2.668X_5 - 0.148 X_{Status}y' > 0 \\ 0 & y' > 0 \end{cases}$$

$$\hat{y}_{(0.25)} = \begin{cases} -0.043 + 2.929 \log(X_1) + 0.206X_2 + 0.941X_3 - 0.041X_4 + 0.481X_5 - 0.068 X_{Status}y' > 0 \\ 0 & y' > 0 \end{cases}$$

$$\hat{y}_{(0.50)} = \begin{cases} -2.473 + 1.501 \log(X_1) + 1.323X_2 + 4.289X_3 - 5.158X_4 + 0.199X_5 - 0.047 X_{Status}y' > 0 \\ 0 & y' > 0 \end{cases}$$

$$\hat{y}_{(0.75)} = \begin{cases} 0.422 + 5.597 \log(X_1) + 0.477X_2 + 1.420X_3 - 0.118X_4 - 0.198X_5 - 0.242X_{Status}y' > 0 \\ 0 & y' > 0 \end{cases}$$

$$\hat{y}_{(0.95)} = \begin{cases} 2.656 + 16.887 \log(X_1) + 1.355X_2 + 6.811X_3 - 0.308X_4 - 0.611X_5 - 0.405 X_{Status}y' > 0 \\ 0 & y' > 0 \end{cases}$$

Saran

Saran-saran yang dapat disampaikan berdasarkan hasil penelitian yang ada adalah sebagai berikut.

1. Dalam penelitian ini, model tobit yang digunakan bersifat linier sehingga prediktor yang bersifat nonlinier namun memiliki peranan penting, seperti variabel umur, yang tidak dapat dimasukkan dalam model. Oleh karena itu diperlukan pengembangan terhadap model regresi TKB nonlinier.
2. Dari hasil simulasi dan analisis yang telah dilakukan, ditemukan bahwa estimator regresi TKB tidak cukup baik untuk mengestimasi model untuk kuantil-kuantil bawah. Oleh karena itu disarankan untuk melakukan modifikasi dan kajian lebih lanjut terkait hal tersebut.

DAFTAR PUSTAKA

- Aditiawarman. 2005. *Hubungan Ketahanan Hidup 1 Tahun Penderita Kanker Payudara yang Dirawat di RSUP Dr. Kariadi Semarang dengan Faktor – Faktor yang Berhubungan*. Karya Akhir PPDS I: Fakultas Kedokteran Universitas Diponegoro.
- Agresti, A. (2002). *Categorical Data Analysis*. New York: John Wiley & Sons.
- Ahmed, F. E., Vos, P. W., dan Holbert, D.: *Modeling Survival in Colon Cancer A Methodological Review*. *Molecular Cancer*, 6, 15 (2007)
- Amemiya, T. 1985. *Advanced Econometrics*. Harvard University Press. Cambridge, Massachusetts. USA.
- Arlinda, P. 2002. Faktor risiko kanker payudara pasien wanita di Rumah Sakit Kanker Dharmais, Jakarta. *Indonesia Journal of Cancer*. 2002;6(3):105-111.
- Baehaqi, R. 2010. *Hubungan Jumlah Leukosit dan Skor Karnofsky pada Pasien Kanker Payudara*. Karya Tulis Ilmiah. Semarang: Fakultas Kedokteran Universitas Diponegoro.
- Bierens, H. J. 2004. *The Tobit Model*.
- Buhai, I. S. (2005). "Quantile Regression: Overview and Selected Applications". *Ad Astra*, Vol. 4, 1-17.
- Chai, T., dan Draxler, R. R. (2014). "Root Mean Square Error (RMSE) or Mean Absolute Error (MAE): Arguments Against Avoiding RMSE in The Literature". *Geoscientific Model Development*, Vol. 7, 1247-1250.
- Chen, C. (2005). *An Introduction to Quantile Regression and The QUANTREG Procedure*. Retrieved October 20th, 2014, from <http://www2.sas.com/proceedingd/sugi30/213-30.pdf>
- Chuansumrit A, T K, H S, et al.: *Dengue Hemorrhagic fever in children with underlying hematologic-oncologic diseases*. *Thai J Hematol Transf Med*, 13, (2003)
- Collet, D.: *Modelling Survival Data in Medical Reseach*. London: Chapman and Hall. (1994)
- Dajan, A, 1986, *Pengantar Metode Statistik II*, Penerbit LP3ES, Jakarta.
- Dinas Kesehatan Provinsi Jawa Tengah. *Profil Kesehatan Provinsi Jawa Tengah Tahun 2009*. Jawa Tengah (Indonesia): Dinas Kesehatan Provinsi Jawa Tengah: 2009.
- Gale, D & Charette, J. 1999. *Oncology Nurning Care Plans*. Translated by I Made Kariyasa. Edited by Monica Ester. Jakarta: EGC.
- Galton, Francis. 1886. *Regression Towards Mediocrity in Hereditary Stature*. *Journal of the Anthropological Institute* 15, pp 246-63
- Gelman, A., Carlin, J. B., Stern, H. S., dan Rubin, D. B. (2004). *Bayesian Data Analysis Second Edition*. Florida: Chapman and Hall.
- Glass, G., dan Hopkins, K. (1995). *Statistical Methods in Education and Psychology 3rd Edition*. Boston: Allyn and Bacon.
- Greene, W. H. 2008. *Econometrics Analysis, 6th Edition*. New Jersey: Prentice Hall
- Greene, William H. 2012. *Econometric Analysis. 7th Edition*. New Jersey: Prentice Hall.
- Gujarati, D. 2004. *Basic Econometrics. Fourth Edition. International Edition*. Singapore. McGraw-Hill
- Gujarati, D N., D C. . 2013. Edisi 5. Buku 1. *Dasar-Dasar Ekonometrika*. Jakarta: Salemba Empat
- Gujarati, D., D C. P. 2013. Edisi 5. Buku 2. *Dasar-Dasar Ekonometrika*. Jakarta: Salemba Empat.

- Hill, B.M.: The three parameter lognormal distribution and Bayesian analysis of a point-source epidemic. *Journal of the American Statistical Association* 58, 72-84 (1963)
- Hosmer, D.W., Lemeshow, S. and May, S.: *Applied Survival Analysis*. Wiley & Sons, Inc., Hoboken, New Jersey. (2008)
- Joreskog, K. G. 2002. Censored Variabels and Censored Regression, <http://aac.asm.org/cgi/reprint/50/1/62>
- Karnofsky DA, Burchenal JH. 1949. "The Clinical Evaluation of Chemotherapeutic Agents in Cancer." In: MacLeod CM (Ed), *Evaluation of Chemotherapeutic Agents*. Columbia Univ Press. Page 196.
- Kass, R. E., dan Raftery, A. E. (1995). "Bayes Factors". *Journal of The American Statistical Association*, Vol. 90, 773-795.
- Koenker, R., and Basset, G. (1978). "Regression Quantiles". *Econometrica*, Vol. 46, 33-50.
- Koenker, R., and Hallock, K. (2001). "Quantile Regressions". *Journal of Economic Perspectives*, Vol. 15, 143-156.
- Kutner, M. H., C. J. Nachtsheim J. Neter and W. Li. 2004. *Applied Linear Regression Models*. New York: McGraw-Hill/Irwin
- Latan, H. 2014. *Aplikasi Analisis Data Statistik untuk Ilmu Sosial Sains dengan IBM SPSS*. Bandung: Alfa Beta.
- Lavine, M., dan Schervish, M. J. (1999). "Bayes Factors: What They Are and What They Are Not". *The American Statistician*, Vol. 53, 119-122.
- Lawless, J.F. 1982. *Statistical Models and Methods for Lifetime Data* USA: John Wiley & Sons, Inc.
- Le, C. T.: *Applied Survival Analysis*. New York: John Wiley and Sons, Inc. (1997)
- Luwia, A. 2005. Hubungan jenis histologi dan ketahanan hidup 5 tahun penderita kanker payudara. *Maj Kedok Nusantara*. 2005;39(1):1-5.
- Osborne, J., dan Waters, E. (2002). "Four Assumptions Of Multiple Regression That Researchers Should Always Test". *Practical Assessment, Research & Evaluation*, Vol. 8 (1), 1-5.
- Promkes RSUD Tugurejo Semarang. 2015. *Layanan Onkologi Mengenal Kanker*. Jawa Tengah (Indonesia): RSUD Tugurejo Semarang: 2015.
- Robinson, C., T, S., and Schumaker, R. (2013). "Test of Moderatio Effects: Difference in Simple Slopes versus the Interaction Term". *Multiple Linear Regression Viewpoints*, Vol. 39 (16-24)
- Salim, T. 2007. Model Tobit Standar dan Model Heckit. [SKRIPSI]. Fakultas MIPA. UI: Depok. <URL: <http://lontar.ui.ac.id/file?file=digital/20180918-011-07-Model%20Tobit.pdf>>. Diakses pada: tanggal 16 April 2014
- Shukur, G., and Mantalos, P. (2004). "Size and Power of the RESET Test as Applied to Systems of Equations: A Bootstrap Approach". *Journal of Modern Applied Statistical Methods*, Vol. 3 (2), 1-22.
- Sukestiyarno. 2011. *Olah Data Penelitian Berbantuan SPSS*. Semarang: Universitas Negeri Semarang.
- Supartono & Suryanto, A. 2012. *Faktor – Faktor yang Mempengaruhi Ketahanan Hidup Satu Tahun Penderita Kanker Payudara Stadium Lanjut di RSUP Dr. Kariadi Semarang*. *Medical Hospitalia*, vol 1(1): 25 – 31.
- Sutjipto, M. 2000. Faktor – Faktor Risiko yang Berhubungan Ketahanan Hidup 1 Tahun Penderita Kanker Payudara yang Dirawat di RSUP Dr. Kariadi Semarang.
- Sutriston, A. 1992. Faktor – Faktor yang Mempengaruhi Ketahanan Hidup Satu Tahun Penderita Kanker Payudara Stadium Lanjut di RSUP Dr. Kariadi Semarang. *Medical Hospitalia*, vol 1(1): 25 – 31.
- Tobin, J. 1967. Tobin's Q Ratio As An Indicator of the valuation of the company. *Journal of Financial Economics*. Vol LIII, No.3: June, pp. 287 – 298.
- Widarjono, A. 2007. *Ekonometrika: Teori dan Aplikasi untuk Ekonomi dan Bisnis*. Edisi Kedua. Yogyakarta: Ekonisia Fakultas Ekonomi Universitas Islam Indonesia.
- Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Massachusetts: MIT Press.
- Yao, Y., and Lee, Y. (2010). *Another Look at Linear Programming for Feature Selection via Methods of Regularization*. Ohio: The Ohio State University.