



ISSN 0216-2393

GRADIEN

Vol. 5 No. 2 Juli 2009

JURNAL MIPA



FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS BENGKULU

Gradien	Vol.5	No. 2	Hal. 448-499	Bengkulu, Juli 2009	ISSN 0216-2393
---------	-------	-------	--------------	------------------------	-------------------



ISSN 0216-2393

GRADIEN

Vol. 5 No. 1 Januari 2009

JURNAL MIPA

DAFTAR ISI

Fisika

1. Pendugaan Struktur Bawah Permukaan Daerah Rawan Gerakan Tanah Di Desa Kampung Gandung Kabupaten Lebong Dengan Metode Geolistrik Tahanan Jenis Konfigurasi Wenner (*Suhendra*) 448-452
2. Analisis Karakteristik Pasang Surut Dengan Metode Admiralty Di Perairan Tapak Padri Bengkulu (*Supiyati*) 453-457
3. Analisa Kualitas Air Sumur Puskesmas Pembantu Desa Talang Pauh Kecamatan Pondok Kelapa Bengkulu Tengah Dengan Menggunakan Alat Elektrolisa Dan TDS 458-462
4. Pemodelan Rangkaian RC Dengan Injeksi Sinyal Listrik Pada Tanah Untuk Menentukan Nilai Kepadatan Tanah (*Suwardi*) 463-467
5. Sistem Kontrol Alat Listrik Dengan Memanfaatkan Fasilitas SMS Pada Mobile Phone Siemens Berbasis Mikrokontroller Atmel AT80S8535 (*Syamsul Bahri*) 468-471
6. Potensi Energi Panas Bumi Berdasarkan Data Gradien Suhu Bawah Permukaan Daerah Gunungapi Kaba Bengkulu (*Arif Ismul Hadi*) 472-475

Kimia

7. Daya Ekstraksi Pelarut Organo Phosphor Dalam Ekstraksi Cair-Cair Asam-Asam Karboksilat Pada Limbah Cair Agro Industri (*Agus MH Putranto*) 476-484
8. Penggunaan Biji Kelor (*Moringa Oleifera Lamk*) Sebagai Koagulan Alami Untuk Menurunkan Kandungan Zat Warna Berbahaya Dalam Limbah Cair Industri Kain Besurek (*Teja Dwi Sutanto*) 485-488

Matematika

9. Pendugaan Parameter Regresi Berdasarkan Sebaran T-Student (*Nurul Astuty Yensi.B*) 489-493

Biologi

10. Studi Populasi Harimau Sumatera (*Panthera Tigris Sumatrae*, Pocock 1829) Dan Hewan Mangsanya Di Air Ipuh Panjang-Air Berau Taman Nasional Kerinci Seblat Kabupaten Muko-Muko (*Jarulis*) 494-499



Pendugaan Parameter Regresi Berdasarkan Sebaran T-Student

Nurul Astuty Yensy. B

Jurusan Pendidikan MIPA, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Bengkulu, Indonesia

Diterima 10 Juli 2009; Disetujui 15 Juli 2009

Abstract - In this paper we have investigated, by mean of simulation, properties of estimates of regression parameters based on t distribution. The best estimate was obtained by comparing the mean square error resulted from fitting t distribution with various degrees of freedoms. We used simulated data generated from symmetric distributions including uniform, normal, logistic, laplace, Student-t and Tukey's Symetric lambda distribution. We Also used the model to analyze several real data. The results showed that, in general, the best estimate was given by the t distribution with 5 degrees of freedom. However, we have not studied the distributions of the estimates which is essential for inference.

Key Word: Student-t Distribution, Estimates of Regression Parameters

1. Pendahuluan

Regresi merupakan salah satu contoh pemodelan statistika yang paling sering digunakan dalam penelitian, baik yang bersifat percobaan maupun survei. Model yang biasa digunakan adalah:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} + \varepsilon_i,$$

dengan asumsi bahwa x_i bersifat tetap dan ε_i merupakan peubah acak normal yang saling bebas dengan nilai tengah 0 dan ragam σ^2 . Pendugaan parameter dilakukan dengan metode kuadrat terkecil (MKT), yang hasilnya sama dengan hasil pemodelan normal.

Komponen galat merupakan ciri membedakan model statistika dari model matematika. Sebagai peubah acak, galat memiliki sebaran dengan bentuk bermacam-macam, namun jika ukuran contoh cukup besar pemodelan normal mampu memberikan pendekatan dengan baik. Pada contoh berukuran kecil belum tentu demikian, karena di samping peka terhadap pencilan juga kurang efisien jika sebaran galat berekor panjang [1].

Pada kenyataannya sering dijumpai data yang secara alami tidak menyebar normal, misalnya data mengenai banyaknya hama/penyakit yang menyerang suatu luasan areal tanaman maupun yang mati akibat pemberian suatu dosis pestisida, data mengenai

pendapatan penduduk yang tinggal pada suatu daerah, maupun data mengenai produksi sporangium pada suatu taraf konsentrasi metalaxyl yang diberikan. Di samping itu untuk memperoleh data berukuran besar kadang-kadang juga dijumpai kendala, misalnya pada penelitian mengenai ternak-ternak besar.

Beranjak dari kenyataan tersebut, perlu dicari suatu pemodelan yang sekalipun pada contoh berukuran kecil bersifat tegar (*robust*) terhadap sebaran galat, khususnya sebaran setangkup. Sebaran t merupakan salah satu sebaran setangkup yang mempunyai panjang ekor bervariasi tergantung pada derajat bebasnya. Dengan demikian dimungkinkan ada sebaran t dengan derajat bebas tertentu yang cukup baik untuk memodelkan galat yang berasal dari sebaran berekor panjang maupun pendek.

Jika galat dimodelkan menyebar t peubah ganda, penduga parameter mudah diperoleh dan pada derajat bebas berapapun hasilnya selalu sama dengan penduga kuadrat terkecil [7]. Sebaiknya jika galat dimodelkan menyebar t peubah tunggal (sekali pun saling bebas) maka penduga parameter tidak dapat dinyatakan secara eksplisit, sehingga pendugaannya harus dilakukan secara iteratif.

Masalah komputasi merupakan salah satu kendala dalam mewujudkan gagasan pemodelan selain normal.

Akibatnya pemodelan normal terpaksa digunakan, walaupun belum tentu tepat. Barangkali hasilnya masih cukup bagus jika pemodelan yang dilakukan disertai dengan diagnosis, misalnya dengan membuang pencilan. Namun hal ini masih saja mengandung kelemahan, karena galat baku yang dihasilkan menjadi kecil (*underestimate*). Dengan perkembangan dunia komputasi yang semakin pesat dan ditemukannya teknik-teknik optimisasi yang canggih, masalah komputasi bukan menjadi kendala lagi.

Model linear dengan galat menyebar t peubah ganda telah dipelajari oleh Zellner [7] serta Lange, *et al* [2]. Oleh sebab itu pada penelitian ini akan dipelajari model linier dengan berbagai sebaran galat t peubah tunggal agar diketahui sifat-sifat statistiknya. Berdasarkan sifat-sifat yang dimiliki diharapkan dapat diperoleh derajat bebas tertentu yang mampu menghasilkan statistik lebih tegar. Hasil penelitian ini diharapkan dapat digunakan dalam praktek sebagai alternatif dari Pemodelan normal yang sudah lazim digunakan selama ini.

2. Metode Penelitian

Model - Pemodelan statistika atas dasar sebaran t -Student pada dasarnya menganggap bahwa galat dari model linear menyebar t -Student. Pada penelitian ini model linear yang digunakan adalah:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} + \varepsilon_i,$$

(dapat dituliskan $y_i = x_i' \beta + \varepsilon_i$), dengan asumsi bahwa x_i bersifat tetap dan ε_i merupakan peubah acak t berderajat bebas v yang saling bebas, dengan parameter lokasi 0 dan parameter skala σ .

Misalkan y adalah peubah acak t berderajat bebas v dengan parameter lokasi μ dan parameter skala σ , maka fungsi kepekatannya adalah:

$$f(y, \mu, \sigma, v) = \frac{\Gamma((v+1)/2)}{\Gamma(v/2)\Gamma(1/2)\sqrt{v}\sigma} \left[1 + \frac{(y-\mu)^2}{v\sigma^2} \right]^{-(v+1)/2} \quad (1)$$

jika y_i , $i = 1, 2, \dots, n$, masing-masing adalah peubah acak t berderajat bebas v yang saling bebas dengan parameter lokasi μ_i dan parameter σ_i , maka fungsi kepekatan peluang bersamanya adalah:

$$f(y_i, \mu_i, \sigma_i, v, n) = \prod_{i=1}^n \frac{\Gamma((v+1)/2)}{\Gamma(1/2)\Gamma(v/2)\sqrt{v}\sigma_i} \left[1 + \frac{(y_i - \mu_i)^2}{v\sigma_i^2} \right]^{-(v+1)/2} \quad (2)$$

Pada regresi linier dengan parameter skala homogen, $\mu_i = x_i' \beta$ dan $\sigma_i = \sigma$, sehingga

$$f(y_i, \mu_i, \sigma, v, n) = \prod_{i=1}^n \frac{\Gamma((v+1)/2)}{\Gamma(1/2)\Gamma(v/2)\sqrt{v}\sigma} \left[1 + \frac{(y_i - x_i' \beta)^2}{v\sigma^2} \right]^{-(v+1)/2} \quad (3)$$

Pendugaan Parameter - Pada model linier dengan galat menyebar t , pendugaan parameter yang paling mungkin dilakukan adalah dengan metode kemungkinan maksimum (MKM).

Logaritma persamaan (3) adalah

$$\begin{aligned} \ln f &= n \ln(\sigma) + n \ln(\Gamma(v+1)/2) - n \ln(\Gamma(1/2)) \\ &\quad - n \ln(\Gamma(v/2)) - n/2 \ln(v) \\ &\quad - (v+1)/2 \sum_{i=1}^n \ln(1 + (y_i - x_i' \beta)^2 / v\sigma^2) \end{aligned} \quad (4)$$

Misalkan g adalah vektor turunan pertama persamaan (4) terhadap β dan H matriks turunan keduanya, maka

$$g = \frac{v+1}{v\sigma^2} \sum_{i=1}^n \frac{(y_i - x_i' \beta) x_i}{1 + (y_i - x_i' \beta)^2 / v\sigma^2} \quad (5)$$

$$H = -\frac{v+1}{v\sigma^2} \sum_{i=1}^n \frac{[1 - (y_i - x_i' \beta)^2 / v\sigma^2] [x_i x_i']}{[1 + (y_i - x_i' \beta)^2 / v\sigma^2]^2} \quad (6)$$

Penyelesaian secara aljabar tidak memungkinkan untuk menyatakan β secara eksplisit, sehingga pendugaan parameter harus dilakukan secara iteratif. Dari bentuk turunan pertama tersebut tampak bahwa sekalipun parameter skala dimodelkan homogen, Pemodelan t melakukan regresi terbobot dengan pembobot

$$[1 + (y_i - x_i' \beta)^2 / v\sigma^2]^{-1}$$

yang besarnya berbanding terbalik dengan besarnya sisaan. Hal ini berakibat bahwa statistik yang dihasilkan lebih tegar terhadap pencilan dan keheterogenan ragam galat.

Sifat-Sifat Penduga Parameter - Karakteristik suatu penduga parameter dapat diketahui berdasarkan bentuk sebaran yang dimodelkan. Jika $E[gg'] = -E[H]$ dan b adalah statistik tak bias bagi β serta sebaran galat yang dimodelkan memenuhi syarat keteraturan maka $\text{Var}(b) \geq (-E(H))^{-1}$. Jika untuk contoh berukuran besar b menyebar normal asimtotik, maka asimtot bagi $\text{Var}(b)$ adalah $(-E[H])^{-1}$ [6]. Pada kondisi ini, besaran $(-E[H])^{-1}$ berfungsi sebagai had bawah ragam penduga

parameter Cramer-Rao, atau lazim disebut had bawah Cramer-Rao dan disingkat dengan HBCR [4]. Jika g dapat dinyatakan sebagai $A(\beta)(b-\beta)$, sedangkan $A(\beta)$ bebas dari nilai pengamatan, maka $\text{Var}(b) = (-E[H])^{-1}$. Berdasarkan kenyataan tersebut dapat dipastikan, bahwa HBCR dapat dicapai maka ragam bagi penduga kemungkinan maksimum sama dengan HBCR [6].

Pada pemodelan t , b tidak dapat dinyatakan secara eksplisit sehingga g tidak dapat dinyatakan sebagai $A(\beta)(b-\beta)$. Dengan demikian walaupun $(-H)^{-1}$ merupakan penduga tak bias bagi $(-E[H])^{-1}$, maka $(-H)^{-1}$ bukan penduga ragam yang tak bias. Alternatif lain untuk menduga ragam ialah melalui mekanisme *bootstrap*, walaupun memerlukan waktu pemrosesan agak lama. Pendugaan ragam melalui *bootstrap* ini pernah dilakukan oleh Lange, *et al.* [2], baik terhadap data pengamatan maupun sisaan.

Setyono telah melakukan kajian simulasi pemodelan t pada peubah tunggal pada berbagai derajat bebas, untuk galat yang menyebar normal, seragam, laplace, logistic, Cauchy, t pada $v = 2, 3, 4, 5, 6$ dan 10 , serta menyebar λ pada $\lambda = -0,224, -0,147, -0,089, -0,003$ dan $0,099$ [5]. Dari kajian tersebut diketahui bahwa penduga parameter mulai konstan sejak $v = 10$, sedangkan lf pada persamaan (4) mulai konstan sejak $v = 400$. Dengan demikian jika v juga merupakan parameter yang diduga, sedangkan pendugaannya tidak secara iteratif, maka cukup mencoba $v = 3, 4, \dots, 10$ dan ∞ .

Seperti halnya dengan penduga kuadrat terkecil, maka selama galat bernilai tengah nol dan peubah bebas bersifat tetap, maka penduga parameter yang dihasilkan oleh pemodelan t bersifat tak bias. Jika peubah bebas dimodelkan tidak tetap (acak), penduga parameter yang dihasilkan tetap tidak berbias, walaupun ragamnya besar. Dengan demikian sifat ketakbiasan ini hanya tergantung pada nilai harapan galat.

Lebih lanjut Setyono menyatakan bahwa kuadrat tengah galat bagi statistik hasil pemodelan t semakin kecil dengan semakin besarnya ukuran contoh [5]. Dengan demikian pemodelan t selalu menghasilkan

statistik yang konsisten. Statistik hasil pemodelan normal justru tidak konsisten jika galat yang dihadapi menyebar Cauchy. Hal ini terjadi karena statistik hasil pemodelan normal bagi nilai tengah populasi adalah rata-rata contoh, sedangkan pada peubah Cauchy sebaran bagi rata-rata contoh sama saja dengan sebaran bagi nilai-nilai pengamatan. Penjelasan lebih lengkap dapat dilihat misalnya pada [6].

Pada pemodelan t dengan $v < \infty$ nilai HBCR tidak pernah dicapai, sehingga tidak dapat diketahui secara pasti apakah suatu statistik bersifat terbaik atau tidak. Berdasarkan keefisienan nisbi statistik hasil pemodelan t terhadap statistik hasil pemodelan normal diketahui bahwa pemodelan normal lebih efisien hanya untuk memodelkan galat yang menyebar normal atau seragam, sedangkan untuk memodelkan galat dari sebaran lainnya selalu ada pemodelan t yang lebih efisien. Jika galat pada model regresi dianggap mempunyai sebaran yang berekor pendek, pada hal kenyataannya berekor panjang, maka keefisienan nisbinya relatif rendah. Sebaliknya jika galat pada model regresi dianggap mempunyai sebaran berekor panjang, padahal kenyataannya berekor pendek, maka keefisienan nisbinya tidak terlalu rendah. Jika pada setiap sebaran galat dan ukuran contoh, keefisienan nisbi tersebut diberi peringkat dari yang paling efisien hingga paling kurang efisien, hasilnya dapat dilihat pada Tabel 1.

Tabel 1. Frekuensi Suatu Derajat Bebas Menduduki Peringkat Pertama sampai Terakhir Berdasarkan Keefisienan Nisbi

		Peringkat																			
		n=10					n=15					n=30					Total				
v		1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
2	0	1	1	12	2	0	2	1	11	2	0	1	1	12	6	0	4	3	3	3	3
3	7	1	0	4	4	7	1	0	3	5	8	0	0	4	4	22	2	0	7	13	1
4	2	7	4	3	0	2	7	3	4	0	2	9	2	3	0	6	23	9	10	0	0
5	2	4	10	0	0	1	5	10	8	0	2	4	10	0	0	5	13	30	0	0	0
6	3	4	1	8	0	5	2	1	8	0	2	3	3	8	0	10	9	5	24	0	0

Tampak bahwa peringkat keefisienan tidak tergantung pada ukuran contoh. Berdasarkan peringkat tersebut dapat dikatakan bahwa $v=5$ relatif paling tegar terhadap sebaran galat, karena selalu menduduki peringkat ke 1, 2 atau 3.

3. Hasil Dan Pembahasan

Sebagai penerapan pemodelan t digunakan data Stack Loss [3] dan data Spora (Daryono, 1994). Pada umumnya keunggulan metode-metode tegar, Andrew misalnya adalah mampu memberikan hasil tegar pada kondisi data megandung pencilan (misal data Stack Loss), dalam arti mendekati hasil pemodelan normal setelah pencilan dibuang (Tabel 2). Sebaliknya pada data tanpa pencilan tidak dijamin mampu memberikan hasil yang mendekati hasil pemodean normal.

Tabel 2. Dugaan Koefisien Regresi untuk Data Stack Loss dari Berbagai Metode

Metode	b_0	b_1	b_2	b_3	If
Normal (N)	-39.92	0.72	1.30	-0.15	-
(Std)	11.90	0.13	0.37	0.16	52.51
N-Pencilan	-42.45	0.96	0.56	-0.11	*)
Huber	-41.00	0.83	0.91	-0.13	-
Andrews	-37.20	0.82	0.52	-0.07	-
$t, v=3$	-39.96	0.86	0.69	-0.11	-
	7.23	0.12	0.28	0.10	51.21
$t, v=5$	-39.92	0.84	0.88	-0.13	-
	9.36	0.14	0.35	0.12	51.97

*) tanpa obs no. 4 dan 21. Galat baku dihitung melalui $(-H)$

Statistik hasil pemodelan t bersifat lebih "Moderat", karena pada kondisi tanpa pencilan hasilnya tidak jauh berbeda dengan hasil pemodelan normal, sedangkan dalam kondisi ada pencilan hasilnya tidak jauh berbeda dengan hasil metode-metode yang selama ini dikenal tegar. Pada data Stack Loss ini, derajat bebas penghasil If tertinggi adalah 3.

Berbeda dengan metode tegar pada umumnya, pemodelan t ini tidak dimaksudkan untuk menggantikan pemodelan normal yang lazim digunakan, melainkan memasukkan pemodelan normal sebagai salah satu bagiannya. Dengan menganggap derajat bebas sebagai salah satu parameter yang diduga, dimungkinkan diperoleh hasil seperti pemodelan normal.

Pemodelan t dapat digunakan sebagai alternatif diagnosis bagi pemodelan normal [2]. Data Spora pada Tabel 3 merupakan teladan yang baik untuk hal ini. Pada tabel tersebut dibandingkan hasil pemodelan normal, pemodelan normal yang disertai dengan membuang pencilan, dan pemodelan t pada v yang If maksimumnya tertinggi (untuk data ini pada $v = 3$).

Hasil pemodelan t lebih mendekati hasil pemodelan normal tanpa pencilan. Pada pemodelan normal, pencilan pada suatu pengamatan juga berpengaruh terhadap nilai-nilai sisaan pada pengamatan yang lain, sedangkan pada pemodelan t hanya berpengaruh terhadap nilai sisaan pada pengamatan itu sendiri.

4. Kesimpulan

Dengan memodelkan gaia saling bebas dengan ragam identik, pemodelan normal memberikan bobot sama kepada semua pengamatan, sedangkan pemodelan t memberikan bobot yang sama yang besarnya berbanding terbalik dengan nilai sisaan. Hal ini berakibat bahwa statistik yang dihasilkan oleh pemodelan t lebih tegar terhadap pencilan dan keheterogenan ragam galat. Keterangan tersebut semakin kuat jika derajat bebas semakin kecil.

Tabel 3. Dugaan Koefisien Regresi dan Nilai-Nilai Sisaan Hasil Pemodelan Normal dan t untuk Data Spora

X	Y	Sisaan Hasil Pemodelan		
		Normal (N)	$t, v = 3$	N-pencilan
0.0	904	616.28	773.77	dibuang
0.5	112	-172.11	-16.65	38.22
1.0	56	-224.49	-71.08	-16.92
5.0	55	-196.57	-59.46	-11.04
25.0	9	-97.94	-42.39	-22.63
50.0	1	74.84	28.45	12.38
b_0		287.72 (181.64)	130.23 (120.64)	74.64
b_1		-7.23 (7.93)	-3.15 (4.69)	-1.72

X: Konsentrasi Metalaxyl (ppm),

Y: Produksi Sporangium pada 8 Hari Setelah Inkubasi

Pada semua sebaran galat setangkup yang bernilai tengah nol, statistik hasil pemodelan t selalu tak bias dan konsisten. Statistik hasil pemodelan normal tidak konsisten jika galat yang dihadapi menyebar Cauchy. Jika galat dianggap mempunyai sebaran berekor pendek padahal kenyataannya berekor panjang, keefesienan nisbinya relatif rendah. Sebaliknya jika dianggap berekor panjang padahal kenyataannya berekor pendek, keefesienan nisbinya tidak terlalu rendah. Pemodelan t dengan $v = 5$ relatif paling tegar terhadap berbagai sebaran galat setangkup.

Penduga parameter hasil pemodelan t relatif konstan sejak $v = 10$, sedangkan $1f$ mulai konstan sejak $v=400$. Oleh sebab itu dimungkinkan memilih statistik terbaik dengan hanya mencoba $v=3, 4, \dots, 10$, dan ∞ . Menduga v berdasarkan data melalui mekanisme iterasi seperti halnya menduga β tidak disarankan, karena selain rumit juga dimungkinkan diperolehnya $v < 3$ atau bahkan tidak bulat.

Pada pemodelan t , hanya ada ragam penduga parameter tidak pernah dicapai sehingga $(-H)^{-1}$ bukan penduga ragam yang tak bias. Oleh sebab itu sebaran $(-H)^{-1}$ maupun penduga parameternya perlu dikaji, agar dapat digunakan untuk keperluan inferensia. Ada kecurigaan bahwa penduga parameter menyebar setangkep, bahkan mungkin normal.

Dengan diproduksinya komputer mikro berkecepatan tinggi dimungkinkan membuat paket program pemodelan t . Dalam penelitian ini telah dihasilkan program sederhana regresi berganda pemodelan t dalam bahasa Turbo Pascal.

Daftar Pustaka

- [1] Krasker, W. S and R. E. Welsch. 1992. Efficient Bounded-Influence Regression Estimation. *Journal of the American Statistical Association*. 77 (375): 595-604.
- [2] Lange, K. L., R.J.A. Little, and M.G. Taylor. 1989. Robust Statistical Modelling Using the t Distribution. *Journal of the American Statistical Association*. 84 (408):881-896.
- [3] Montgomery, D. C. and E.A. Peck. 1992. *Introduction to Linear Regression Analysis*. John Wiley & Sons. New York.
- [4] Nasoetion, A. H. dan A. Rambe. 1983. *Teori Statistika*. Bratara Karya Aksara. Jakarta.
- [5] Setyono. 1996. *Pemodelan Statistik Atas Dasar Sebaran t -student*. Tesis Magister Sains. Program Pascasarjana IPB, Bogor. Tidak dipublikasikan.
- [6] Stuart, A. and J.K. Ord. 1991. *Kendall's Advanced Theory of Statistics*. Great Britain for Edward Arnold. London.
- [7] Zellner, A. 1986. Bayesian and Non-Bayesian Analysis of the Regression Model With Multivariate Student- t Error Terms. *Journal of the American Statistical Association*. 71 (354): 400-405.