

Perbandingan Metode *Stepwise*, *Best Subset Regression*, dan Fraksi dalam Pemilihan Model Regresi Berganda Terbaik

HERLINA HANUM

Jurusan Matematika, Universitas Sriwijaya, Sumatera Selatan, Indonesia

INTISARI: Pemilihan model regresi terbaik adalah penentuan peubah bebas mana yang akan dimasukkan dalam model regresi sehingga model tersebut dapat menjelaskan perilaku peubah tak bebas dengan baik. Metode yang umum digunakan antara lain *Stepwise* dan *Best Subset Regression*. Metode Fraksi lebih diarahkan untuk penentuan model terbaik dengan kecenderungan multikolinear yang minimum. Dalam tulisan ini dikaji kelebihan dan kekurangan Metode Fraksi dibandingkan dengan *Stepwise* dan *Best Subset Regression*. Karena perbandingan ini difokuskan pada pemilihan model terbaik pada data yang memiliki masalah multikolinear maka contoh masalah yang diambil mewakili data dengan tingkat multikolinearitas yang tinggi. Hasil analisis terhadap kedua data tersebut dengan ketiga metode menunjukkan bahwa ketiga metode tersebut sama-sama dapat memberikan model terbaik. Kelebihan Metode Fraksi adalah mampu memberikan model terbaik yang memiliki kecenderungan kolinearitas yang lebih rendah dari 2 metode lain. Sementara keharusan untuk mengambil fraksi merupakan kekurangan Metode Fraksi.

KATA KUNCI: model regresi, metode *stepwise*, *best subset regression*, fraksi

April 2011

1 PENDAHULUAN

Model regresi terbaik adalah model yang dapat menjelaskan perilaku peubah tak bebas dengan sebaik-baiknya dengan memilih peubah-peubah bebas dari sekian banyak peubah bebas yang tersedia dalam data. Untuk menentukan peubah bebas mana yang akan dimasukkan ke dalam model regresi, menurut Draper^[1], ada dua kriteria yang saling bertentangan yaitu

1. Agar persamaannya bermanfaat untuk peramalan, biasanya ingin dimasukkan sebanyak mungkin peubah sehingga diperoleh nilai ramalan yang andal
2. Karena untuk memperoleh informasi dari banyak peubah serta pemantauannya seringkali diperlukan biaya yang tinggi, maka diinginkan persamaan regresi yang mencakup sesedikit mungkin peubah. Kompromi diantara kedua kriteria itulah yang disebut pemilihan model regresi terbaik.

Masalah yang sering muncul dalam regresi berganda adalah adanya hubungan linear antar peubah bebas. Kondisi ini disebut masalah multikolinear. Adanya masalah multikolinear ditunjukkan oleh nilai *Variance Inflation Factors (VIF)* yang tinggi. Jika ada masalah multikolinear maka kesimpulan yang didapat dari hasil pengujian untuk model regresi maupun untuk masing-masing peubah yang ada dalam model, seringkali tidak

tepat. Oleh sebab itu, masalah multikolinear harus dihindari.

Untuk pemilihan model regresi terbaik ada beberapa metode yang biasa digunakan. Dua diantaranya *Stepwise* dan *All Possible Regression (Best Subset Regression)*. Sementara Metode Fraksi belum dikenal karena baru ditemukan.

Stepwise dan *Best Subset Regression* memulai pemilihan dengan model paling sederhana yaitu model dengan satu peubah. Selanjutnya disusupkan peubah lain satu per satu sampai didapat model yang memenuhi kriteria terbaik. Kriteria didasarkan pada penambahan R^2 dan R^2_{adj} , pengurangan S^2 , atau kedekatan nilai *C-p Mallow* dengan jumlah peubah dalam model. Metode *Stepwise* memilih peubah berdasarkan korelasi parsial terbesar dengan peubah yang sudah ada dalam model. Dalam metode ini peubah yang sudah masuk dalam model dapat saja dikeluarkan lagi, sehingga langkah yang diperlukan menjadi banyak. Sementara *Best Subset Regression* memilih peubah berdasarkan penambahan nilai R^2 tertinggi dari model terpilih sebelumnya.

Metode Fraksi menggunakan hasil analisis korelasi sebagai dasar pemilihan peubah yang akan dimasukkan dalam model. Peubah-peubah dikelompokkan berdasarkan kedekatannya secara linear. Suatu kelompok alias berisi peubah-peubah yang saling berkorelasi kuat. Model terbaik dibentuk dari peubah-peubah yang mewakili masing-masing kelompok alias. Tetapi dengan ketentuan bahwa peubah yang dipilih adalah

fraksi dari jumlah peubah bebas^[2].

Tujuan penulisan ini adalah untuk mengkaji kelebihan dan kekurangan Metode Fraksi dibandingkan dengan *Stepwise* dan *Best Subset Regression*. Contoh masalah yang diambil mewakili data dengan tingkat multikolinearitas tinggi.

2 KAJIAN PUSTAKA

Dalam kajian pustaka ini disajikan kriteria yang digunakan dalam pemilihan regresi ganda terbaik. Berbagai kriteria tersebut adalah R^2 , R_{adj}^2 , S^2 , $C-p$ Mallow dan VIF

2.1 Koefisien Determinasi (R^2)

Koefisien determinasi adalah ukuran bagian ragam peubah terikat yang dapat dijelaskan secara bersama-sama oleh peubah bebas yang ada didalam model. Nilai koefisien determinasi

$$R^2 = \frac{RSS_p}{SStotal} = 1 - \frac{SSE_p}{SStotal}$$

dengan RSS_p = jumlah kuadrat regresi, $SStotal$ = jumlah kuadrat total, SSE = jumlah kuadrat galat dan p = jumlah peubah bebas dalam model. R^2 akan terus bertambah seiring bertambahnya peubah bebas yang dimasukkan dalam model. Peubah yang potensial ditambahkan dalam model adalah yang memberi penambahan nilai R^2 yang cukup berarti^[3].

2.2 R_{adj}^2

R_{adj}^2 adalah penyesuaian dari R^2 . Nilai R_{adj}^2 ditentukan dengan rumus:

$$R^2 = 1 - \left(\frac{n-1}{n-p} \right) \frac{SSE_p}{SStotal} \quad \text{dengan } n = \text{jumlah amatan}$$

Nilai R_{adj}^2 hanya akan naik jika nilai $(n-p)SSE_p$ turun, karena $(n-1)SStotal$ tetap. Model yang baik memiliki R_{adj}^2 yang besar^[3].

2.3 Cp-Mallow

Nilai dugaan yang didapat dari persamaan regresi berdasarkan sebagian peubah bebas pada umumnya bias. Untuk menilai kabaikan model digunakan *means square error (MSE)* dengan varian dan biasnya. C.L. Mallow menyarankan statistik

$$C - p = \frac{SSE_p}{\hat{\sigma}^2} - (n - 2p)$$

Penyimpangan $C - p$ dari p digunakan sebagai ukuran bias. Model terbaik berdasarkan $C - p$ adalah model yang memiliki nilai $C - p$ terdekat dengan jumlah peubah dalam model^[3].

2.4 Variance Inflation Factor (VIF) dalam Regresi

Variance inflation factor (VIF) atau faktor pengaruh varian digunakan untuk mendeteksi apakah suatu peubah bebas memiliki hubungan linear kuat dengan peubah bebas lainnya (adanya multikolinear antar peubah bebas). VIF menunjukkan berapa besar ragam dugaan koefisien regresi ke- i meningkat dari yang seharusnya jika R_i^2 sama dengan nol: artinya adalah keadaan peubah bebas ke- i ortogonal terhadap peubah-peubah bebas lain dalam analisis. VIF memberikan indikasi yang beralasan dan intuitif dari pengaruh multikolinearitas pada ragam koefisien regresi ke- i ^[4].

$VIF = 1$ menunjukkan tidak ada hubungan; $VIF > 1$, sebaliknya. Nilai VIF terbesar diantara semua peubah bebas sering digunakan sebagai indikator multikolinearitas terburuk. Montgomery^[5] menyatakan bahwa jika VIF lebih besar dari 5-10, maka pendugaan koefisien regresi tidak baik.

3 METODE PENGAJIAN

Tulisan ini dibuat berdasarkan hasil analisis data menggunakan metode *Stepwise*, *Best Subset Regression*, dan Metode Fraksi. Dalam Kajian ini digunakan 2 data contoh yang masing-masing memiliki 8 dan 6 peubah bebas. Kedua data tersebut memiliki korelasi tinggi antar peubah bebasnya. Agar terlihat jelas perbandingan dari hasil ketiga metode tersebut, penyajian dilakukan untuk masing-masing data. Langkah pertama adalah menyajikan model penuh untuk data contoh. Kemudian dipilih model terbaik dengan menggunakan masing-masing metode. Kajian difokuskan pada bentuk dan sifat model terbaik yang dihasilkan dari ketiga metode tersebut. Kajian sifat model terbaik difokuskan pada nilai R_{adj}^2 dan nilai VIF masing-masing peubah. Dari model terbaik masing-masing dikaji kelebihan dan kekurangan model yang dihasilkan. Sebagai penutup disajikan hasil perbandingan secara umum untuk kedua data contoh tersebut.

4 PEMBAHASAN

Dalam bagian pembahasan ini disajikan hasil analisis terhadap 2 data contoh. Data pertama mewakili data yang juga memiliki kecenderungan multikolinear tapi korelasi antar peubah bebasnya tidak sekuat data kedua. Data tersebut merupakan hasil simulasi menggunakan sebaran normal di Minitab^[6]. Data kedua yang digunakan adalah Data Bears yang diambil dari Minitab^[6]. Data ini digunakan untuk mewakili pemilihan model terbaik pada data yang peubah-peubah

bebasnya saling berkorelasi kuat. Berikut ini disajikan hasil analisis untuk kedua data tersebut.

4.1 Analisis untuk Data dengan 8 Peubah Bebas

Data ini merupakan hasil simulasi menggunakan sebaran normal pada Minitab^[6]. Pada data ini peubah terikat Y hanya berkorelasi nyata dengan peubah bebas X_1, X_2, X_3 , dan X_5 yaitu sebesar 0.972, 0.863, 0.913, dan -0.499. Peubah X_1, X_2 , dan X_3 saling berkorelasi kuat. Sementara X_5 berkorelasi nyata dengan X_3 tetapi tidak cukup kuat. X_6 dan X_7 saling berkorelasi kuat. Peubah X_4 dan X_8 tidak berkorelasi dengan peubah bebas lainnya. Model regresi dengan semua peubah bebas menghasilkan model

$$Y = -34.5 + 36.1X_1 - 10.0X_2 + 25.4X_3 + 2.12X_4 + 59X_5 + 3.11X_6 - 3.07X_7 - 0.061X_8$$

TABEL 1: Analisis regresi model penuh untuk data 1

Predictor	Coef	SE Coef	T	P	VIF
Constant	-34.47	45.68	-0.75	0.466	
X_1	36.122	5.945	6.08	0.000	29.4
X_2	-10.033	4.576	2.19	0.051	18.8
X_3	25.405	9.541	2.66	0.022	5.9
X_4	2.122	3.599	0.59	0.567	2.3
X_5	58.5	389.4	0.15	0.883	1.6
X_6	3.109	1.608	1.93	0.079	13.4
X_7	-3.068	1.401	-2.19	0.051	11.0
X_8	-0.0609	0.2006	-0.30	0.767	1.4
$S = 5.296 \quad R - S_q = 98.8\% \quad R - S_{q(adj)} = 98.0\%$					

Pada model ini terjadi multikolinear pada peubah-peubah X_1, X_2, X_3, X_6 , dan X_7 .

Pemilihan Model Terbaik dengan Metode Stepwise untuk Data 1

Langkah selanjutnya adalah menentukan model terbaik dengan metode *Stepwise*. Tiga langkah metode *Stepwise* menghasilkan

Berdasarkan nilai $C - p$ model terbaik yang didapat dari metode *Stepwise* ada pada langkah 3. Model ini mengandung peubah bebas X_1, X_2 , dan X_3 dengan persamaan

$$Y = -17.9 + 38.9X_1 - 13.9X_2 + 30.3X_3$$

Ditinjau dari nilai $R - Sq(adj) = 98.0\%$ model ini sama dengan model penuh, tetapi dalam model ini terjadi masalah multikolinear. Jadi sebenarnya model ini tidak cukup baik. Model terbaik kedua memuat

TABEL 2: Pemilihan model terbaik dengan metode Stepwise untuk data 1

Step	1	2	3
Constant	-47.82	-27.05	-17.92
X_1	32.7	47.0	38.9
P-Value	0.000	0.000	0.000
X_2		-14.6	-13.9
P-Value		0.003	0.000
X_3			30.3
P-Value			0.002
S	9.00	7.05	5.31
R-Sq	94.46	96.79	98.29
R-Sq(adj)	94.15	96.41	97.97
C-p	36.0	16.2	4.1

TABEL 3: Analisis terhadap model $Y = -17.9 + 38.9X_1 - 13.9X_2 + 30.3X_3$

Predictor	Coef	SE Coef	T	P	VIF
Constant	-17.922	6.063	-2.96	0.009	
X_1	38.903	3.904	9.97	0.000	12.6
X_3	30.334	8.109	3.74	0.002	4.2
X_2	-13.925	3.135	-4.44	0.000	8.8
$S = 5.311 \quad R - S_q = 98.3\% \quad R - S_{q(adj)} = 98.0\%$					

peubah bebas X_1 dan X_2 . Karena X_1 dan X_2 berkorelasi sangat kuat dengan koefisien korelasi sebesar 0.941, dapat dipastikan bahwa model inipun memiliki masalah multikolinear.

Pemilihan Model Terbaik dengan Metode Best Subset untuk Data 1

Pemilihan model terbaik dengan Metode *Best Subset Regression* menghasilkan:

Berdasarkan nilai $C - p$ ada dua model yang dapat diambil sebagai model terbaik. Kedua model tersebut adalah model pada langkah 3 dengan 3 peubah dan pada langkah 4 dengan 4 peubah. Tetapi jika ditinjau dari penambahan nilai R^2_{adj} lebih cenderung dipilih pada langkah 3 yaitu dengan peubah bebas X_1, X_2 , dan X_3 . Model ini sama dengan model terbaik pilihan Metode *Stepwise* pada langkah 3. Selanjutnya, model terbaik kedua pilihan kedua metode tersebut juga sama yaitu model dengan peubah bebas X_1 dan X_2 .

Pemilihan Model Terbaik dengan Metode Fraksi untuk Data 1

Peubah X_1, X_2, X_3 saling berkorelasi kuat jadi hanya boleh diambil salah satu saja. Jika jika diambil fraksi 1/2 maka harus dipilih 4 peubah bebas. Pilih X_1

TABEL 4: Analisis terhadap model $Y = -17.9 + 38.9X_1 - 13.9X_2 + 30.3X_3$

Vars	R-Sq	-Sq(adj)	C-p	S	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈
1	94.5	94.2	36.0	9.0043	X							
2	96.8	96.4	16.2	7.0541	X	X						
3	98.3	98.0	4.1	5.3107	X	X	X					
4	98.4	98.0	5.1	5.3152	X	X	X			X		
5	98.8	98.4	3.4	4.7683	X	X	X		X	X		

karena berkorelasi paling kuat dengan Y . Begitu juga dengan X_6 dan X_7 , dipilih X_6 karena berkorelasi dengan Y lebih kuat daripada X_7 . Dua peubah bebas lainnya dapat dipilih dari X_4, X_5 , dan X_8 . Pilih X_4 dan X_5 yang memiliki korelasi lebih kuat dengan Y . Jadi model yang dipilih mengandung peubah-peubah X_1, X_5, X_6 , dan X_8 .

$$Y = 32.6 + 31.1X_1 + 3.01X_4 - 737X_5 + 0.180X_6$$

TABEL 5: Analisis data 1 dengan peubah X_1, X_4, X_5 , dan X_6

Predictor	Coef	SE Coef	T	P	VIF
Constant	32.55	73.89	0.44	0.666	
X_1	31.080	2.225	13.97	0.000	1.4
X_4	3.011	4.657	0.65	0.528	1.3
X_5	-737.0	619.5	-1.19	0.253	1.4
X_6	0.1802	0.8923	0.20	0.843	1.4

$S = 9.130$ $R - S_q = 95.3\%$ $R - S_{q(adj)} = 94.0\%$

Walaupun masih ada beberapa peubah bebas yang tidak nyata tetapi model ini memiliki $R^2_{adj} = 94.0\%$ dan tidak ada masalah kolinearitas lagi. Jika model ini dianggap kurang memuaskan karena ada peubah yang tidak nyata maka dapat digunakan fraksi lebih kecil yaitu 1/4 atau 2 peubah bebas. Metode Fraksi akan memilih peubah X_1 yang berkorelasi paling kuat dengan Y dan peubah bebas X_5 yang juga berkorelasi nyata dengan Y tetapi tidak dengan X_1 . Analisis terhadap model dengan peubah bebas X_1 dan X_5 adalah

$$Y = 49.3 + 31.3X_1 - 846X_5$$

Model ini memiliki nilai yang sedikit lebih besar daripada model pada fraksi 1/2 dan hanya kurang 4% dari model penuh. Dengan nilai VIF sebesar 1,2 untuk kedua peubah bebas dan $R^2_{adj} = 94.5\%$, model ini sangat baik untuk menjelaskan Y .

Untuk data ini, model yang didapat dari Metode Fraksi lebih baik daripada model terbaik dari *Stepwise* dan *Best Subset Regression* terutama jika pertimbangan utamanya adalah model tanpa kolinearitas.

4.2 Analisis untuk Data 2

Data Bears adalah data mengenai ukuran badan dan umur beruang. Peubah tak bebasnya adalah berat (*Weight*) dengan peubah-peubah bebas umur (*Age*), Panjang Kepala (*Head L.*), Lebar Kepala (*Head W.*), Lingkar Leher (*Neck G.*), Panjang badan (*Length*), dan lingkar dada (*Chest G.*). Hasil analisis korelasi menunjukkan bahwa semua peubah bebas dalam data ini saling berkorelasi kuat dengan nilai koefisien korelasi 0,669 sampai 0,940. Begitu juga terhadap peubah terikat *Weight*, koefisien korelasi tertinggi dengan peubah *Chest G.* yaitu 0,966. Model regresi penuh yang melibatkan semua peubah bebas

$$\begin{aligned} Weight = & -237 + 0.422 Age - 3.43 Head L. \\ & + 0.33 Head W. + 4.72 Neck G. \\ & + 0.066 Length + 9.64 Chest G. \end{aligned}$$

TABEL 6: Analisis regresi model penuh untuk data 2

Predictor	Coef	SE Coef	T	P	VIF
Constant	-236.77	26.97	-8.78	0.000	
Age	0.4216	0.1398	3.02	0.003	2.3
Head.L	-3.429	4.301	-0.80	0.428	7.9
Head.W	0.329	3.979	0.08	0.934	3.2
Neck.G	4.717	1.824	2.59	0.012	10.2
Length	0.0659	0.8723	0.08	0.940	8.3
Chest.G	9.640	1.085	8.89	0.000	9.9

dengan $S = 28.15, R^2_{adj} = 94.5\%$ dan nyata dalam uji F , model penuh ini sangat baik. Tetapi ada beberapa peubah bebas yang tidak nyata dan memiliki nilai VIF yang besar. Jadi model ini memiliki masalah kolinearitas.

Pemilihan Model Terbaik dengan Metode *Stepwise* untuk Data 2

Model regresi terbaik yang didapat dengan metode *Stepwise* adalah

Nilai R^2_{adj} model pada langkah 2 dan 3 hampir sama. Tetapi berdasarkan nilai $C - p$, model terbaik adalah pada langkah 3 yaitu model yang memuat peubah bebas *Chest G.*, *Age*, dan *Neck G.*. Berikut ini hasil

TABEL 7: Analisis regresi model penuh untuk data 2

Step	1	2	3
Constant	-274.7	-247.4	-255.1
Chest.G	12.88	11.52	9.33
P-Value	0.000	0.000	0.000
Age		0.49	0.42
P-Value		0.001	0.003
Neck.G			4.3
P-Value			0.008
S	30.9	28.9	27.8
R-Sq	93.48	94.37	94.85
R-Sq(adj)	93.40	94.23	94.66
C-p	18.4	7.1	1.9

analisis terhadap model pada langkah 3 tersebut.

$$Weight = -255 + 0.416 Age + 4.27 Neck G. + 9.33 Chest G.$$

TABEL 8: Analisis data 2 dengan peubah Chest G., Age, dan Neck G.

Predictor	Coef	SE Coef	T	P	VIF
Constant	-255.07	14.83	-17.19	0.000	
Age	0.4162	0.1356	3.07	0.003	2.3
Neck.G	4.273	1.572	2.72	0.008	7.8
Chest.G	9.3321	0.9479	9.85	0.000	7.8
$S = 27.77 \quad R - S_q = 94.9\% \quad R - S_q(adj) = 94.7\%$					

Model ini memiliki kecendrungan kolinearitas yang relatif tinggi terutama pada Chest G. dan Neck G.. Hal ini disebabkan korelasi yang tinggi antara kedua peubah. Jika digunakan model terbaik pada langkah 2, maka didapat model

$$Weight = -247 + 0.491 Age + 11.5 Chest G.$$

dengan $S = 28.86, R - S_q = 94.4\%, R - S_q(adj) = 94.2\%$, semua peubah bebas nyata dan nilai $VIP = 2, 2$ untuk kedua peubah.

Bila dipertimbangkan masalah kolinearitas maka model inilah sebenarnya yang terbaik karena memiliki kecenderungan kolinearitas yang jauh lebih kecil dari model dengan 3 peubah bebas. Tetapi karena kecenderungan untuk menetapkan model yang memiliki semua peubah bebas yang nyata, orang sering kali memilih model pada langkah 3 sebagai model terbaik.

Pemilihan Model Terbaik dengan Metode Best Subset Regression untuk Data 2

Pemilihan model terbaik menggunakan metode Best Subset sampai langkah ke -3(karena tidak ada lagi penambahan R^2) disajikan pada Tabel 9. Model terbaik

pada langkah 2 dan 3 sama persis dengan model terbaik pilihan metode Stepwise

Pemilihan Model Terbaik dengan Metode Fraksi untuk Data 2

Berdasarkan nilai VIF yang ada pada regresi penuh, Metode Fraksi memilih untuk menghilangkan 3 peubah dengan nilai VIF terbesar yaitu Neck G., Length, Chest G.. Dari peubah yang tersisa didapat model regresi

$$Weight = -360 + 1.02 Age + 30.0 Head L. + 17.8 Head W.$$

Pada model ini nilai VIF menjadi jauh lebih kecil, semua peubah bebas nyata, tetapi nilai R^2_{adj} juga berkurang jauh walaupun hasil uji F tetap nyata. Model ini sudah cukup baik untuk menjelaskan Weight terlebih lagi jika fokusnya adalah menghindari masalah multikolinear.

Jika model tersebut dianggap kurang memuaskan berdasarkan nilai R^2_{adj} dipilih 3 peubah bebas lain yang dapat menjelaskan Weight dengan lebih baik, tetapi kurang berkorelasi sesamanya dan dapat mewakili peubah bebas yang tidak dimasukkan dalam model. Peubah bebas Chest G. dapat mewakili Neck G., Head L. dapat mewakili Length, lalu diambil Age karena korelasinya lebih rendah terhadap semua peubah bebas.

$$Weight = -247 + 0.491 Age - 0.05 Head L. + 11.5 Chest G.$$

Model ini memiliki $R^2_{adj}=94,2\%$ yang hampir sama dengan model penuh. Tetapi ada kecenderungan multikolinearitas pada Head L. dan Chest G. ($VIP = 4, 5$ dan $5, 2$) sehingga Head L. tidak nyata.

Dari dua model tersebut dapat dijelaskan bahwa kecendrungan kolinearitas makin kuat jika ada. Peubah Chest G. berkorelasi kuat dengan Length dan Head L. menyebabkan Length dan Head L. tidak nyata. Karena Metode Fraksi membatasi pemilihan model terbaik dengan 1/2 dari jumlah peubah bebas, artinya model yang terpilih harus memiliki 3 peubah bebas, maka kedua model inilah yang ditawarkan. Dengan fokus pemilihan model terbaik tanpa masalah multikolinear, Metode Fraksi lebih memilih model pertama yaitu dengan peubah bebas Age, Head L., dan Head W.. Tetapi jika diizinkan untuk memilih 2 peubah bebas saja maka Metode Fraksi akan memilih peubah pertama Chest G. karena berkorelasi paling kuat dengan Weight. Peubah kedua dipilih Age karena kurang berkorelasi dengan Chest G.. Dengan demikian model terbaik dengan 2 peubah bebas yang dihasilkan ketiga metode sama.

TABEL 9: Model terbaik berdasarkan metode *Best Subset Regression* untuk data 2

Vars	R-Sq	R-Sq(adj)	C-p	S	A	HL	HW	NL	L	CG
1	93.5	93.4	18.4	30.872						X
2	94.4	94.2	7.1	28.862	X					X
3	94.9	94.7	1.9	27.774	X			X		X

TABEL 10: Analisis data 2 dengan peubah *Chest G.*, *Age*, dan *Neck G.*

Predictor	Coef	SE Coef	T	P	VIF
Constant	-360.18	45.73	-7.88	0.000	
Age	1.0206	0.2478	4.12	0.000	2.1
Head.L	30.031	4.623	6.50	0.000	2.6
Head.W	17.758	6.591	2.69	0.009	2.5

$S = 52.44$ $R - S_q = 81.7\%$ $R - S_{q(adj)} = 81.0\%$

5 KESIMPULAN

Pemilihan model regresi terbaik dengan metode *Stepwise* dan *Best Subset Regression* tidak mempertimbangkan masalah multikolinear. Akibatnya model terbaik pilihan kedua metode tersebut masih mengandung masalah multikolinear. Sementara Metode Fraksi lebih fokus untuk menghindari masalah multikolinear, sehingga menghasilkan model terbaik tanpa ada masalah multikolinear. Tetapi keharusan untuk memuat fraksi dari jumlah peubah bebas dalam data menjadi penghambat untuk lebih bebas menentukan model terbaik.

DAFTAR PUSTAKA

- [1] Draper, N., and H. Smith, 1992, *Analisis Regresi Terapan edisi kedua* (Terjemahan oleh Bambang Sumantri), Gramedia, Jakarta.
- [2] Hanum, H., 2009, Penentuan model regresi berganda terbaik menggunakan Metode Fraksi, *Penelitian Fundamental*, Universitas Sriwijaya
- [3] Karim, M.E., Selecting of the best regression equation by sorting out variables, <http://www.angelfire.com/ab5/get5/selreg.pdf>, diunduh 15 Februari 2011.
- [4] O'Brien, R.M., 2007, A Caution Regarding Rules of Thumb for Variance Inflation Factors, *Quality & Quantity* 41:673-690. prentice
- [5] Montgomery, D.C. and E.A. Peck, 1982, *Introduction to Linear Regression Analysis*, John Wiley & Sons, New York.
- [6] Minitab, 2000, Minitab Version 13 : Data _____