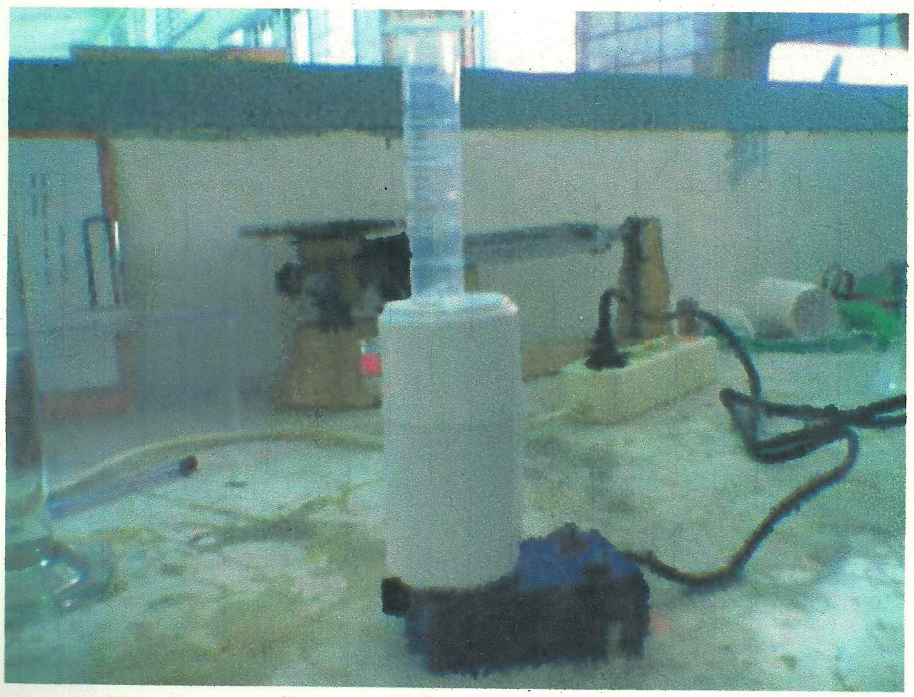
ISSN 0216-2393

GRADIEN

Vol. 6 No. 1 Januari 2010

JURNAL MIPA



FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS BENGKULU

Gradien	Vol. 6	No. 1	Hal. 500-559	Bengkulu, Januari 2010	ISSN 0216-2393

ISSN 0216-2393

GRADIEN

Vol. 6 No. 1 Januari 2010

JURNAL MIPA

DAFTAR ISI

ISSN 0216-2393

Jurnal Gradien Vol.6 No.1 Januari 2010 : 537-541

Pendugaan Model Regresi Dengan Metode Kuadrat Terkecil Parsial

Nurul Astuty Yensy. B

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

Diterima 25 November 2009; Disetujui 07 Desember 2009

Abstrak - Penelitian ini bertujuan untuk menentukan model pendugaan persamaan regresi dan menduga ragam $\hat{y}$ yang dihasilkan Metode Kuadrat Terkecil Parsial (MKTP). Data yang digunakan yaitu hasil pengukuran konsentrasi lemak 36 contoh ikan trout (*rainbow trout*) yang dipublikasikan dalam majalah *Technometrics* (1995). Data konsentrasi lemak membentuk vektor kolom y dan daya serap kepekatan lemak diukur pada 9 panjang gelombang membentuk matriks data X. Model yang digunakan berbentuk: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_9 x_9$, dengan y = konsentrasi lemak ikan trout (%) dan x_i = daya serap pada panjang gelombang ke-i. Dugaan koefisien regresinya dilihat pada saat iterasi dimana nilai PRESS minimal, sedangkan penilaian kesesuaian model dilihat melalui selang kepercayaan hasil dari metode *Jacknife* dan *Bootstrap*. Hasil penelitian menunjukkan koefisien regresi yang dihasilkan dengan metode *Bootstrap* lebih baik dari pada metode *Jacknife*, tetapi waktu yang dibutuhkan lebih lama, dimana metode *Bootstrap* menggunakan replikasi sebanyak B = 2000 dan metode *Jacknife* n = 36. Nilai B harus dicari terlebih dahulu sedangkan untuk *Jacknife* replikasinya sudah pasti sama dengan banyaknya pengamatan. Pendugaan ragam $\hat{y}$ diperoleh melalui persentil *Bootstrap* 2,5% dan 97,5% menghasilkan selang kepercayaan bagi $\hat{y}$ yang menyempit di sekitar nilai tengah y, makin jauh dari nilai tengah y selang makin lebar.

Kata Kunci: **lemak, MKTP , *Jacknife* dan *Bootstrap***

1. Pendahuluan

Dalam bidang kimia pengukuran seringkali didasarkan pada pengamatan dari spektra peubah ganda untuk mengukur absorban pada beberapa panjang gelombang. Ukuran absorban ini dapat digunakan untuk menentukan komposisi campuran kimia [9]. Penggunaan data empirik dan pengetahuan untuk menentukan bagaimana menduga informasi pada peubah respon **y** yang tidak diketahui berdasarkan informasi pada peubah penjelas **X** yang tersedia melalui suatu fungsi matematik ini disebut proses *kalibrasi [8]*.

Aplikasi kalibrasi peubah ganda umumnya dalam bidang spektrofotometri untuk memprediksi peubah-peubah khemometrik dari data spektral [6] ,[5], [7], [8]. Secara statistik hal ini dapat diselesaikan dengan regresi linier ganda. Namun demikian karena jumlah panjang gelombang bisa mencapai ratusan seringkali melebihi jumlah contoh (pengamatan) maka metode-metode regresi tradisional akan menemui masalah kolinieritas. Banyak alasan kalibrasi komersial menggunakan teknik memilih peubah untuk membatasi jumlah panjang gelombang tetapi cara ini mungkin akan menyebabkan hilangnya informasi dalam banyak hal. (Helland, 1998).

Beberapa metode yang dapat mengatasi masalah kolinier ini diantaranya metode komponen utama, metode *ridge* dan metode kuadrat terkecil parsial (MKTP) (Geladi dan Kowalski, 1996; Martens dan Naes, 1999; Young, 2004). Simulasi-simulasi cenderung menunjukkan bahwa MKTP *dalam komputasi lebih efesien dibandingkan dengan* metode komponen utama maupun metode ridge [2], [4], [11].. Namun demikian pada penelitian sebelumnya hanya didapatkan nilai dugaan dari peubah respon **y** saja [8], sedangkan untuk menilai baik tidaknya hasil dugaan belum ada formula untuk mendapatkan penduga ragam. Untuk mengatasi hal ini beberapa peneliti menyarankan metode pendugaan nonparametrik (Metode *Jacknife* dan *Bootstrap*). Penelitian ini bertujuan untuk menentukan model pendugaan persamaan regresi dan menduga ragam $\hat{y}$ yang dihasilkan MKTP.

2. Metode Penelitian

Data yang digunakan dalam penelitian ini berasal dari data yang dipublikasikan dalam majalah Technometrics [9]. Data ini berisi hasil pengukuran konsentrasi lemak 36 contoh ikan trout (*rainbow trout*). Konsentrasi lemak diukur melalui 2 tahap; pertama berdasarkan analisis laboratorium yang baku menghasilkan data konsentrasi, membentuk vektor kolom **y**; ke-dua berdasarkan instrumen NIR. Pengukuran dengan *instrumen* NIR berdasarkan daya serap kepekatan lemak yang diukur pada 9 panjang gelombang membentuk matriks data **X**.

Model yang didasarkan pada data di atas berbentuk:

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_9 x_9 + \varepsilon$

y = konsentrasi lemak ikan trout (%)

x_i = daya serap pada panjang gelombang ke-i.

Perhitungan dengan algoritma kuadrat terkecil parsial ini harus dilakukan dengan iterasi sejumlah banyaknya peubah **x**. Dugaan koefisien regresinya dilihat pada saat iterasi dimana nilai PRESS minimal [2], [11]. Penilaian kesesuaian model dilihat melalui selang kepercayaan hasil dari metode *Jacknife* dan *Bootstrap*.

a. Algoritma MKTP

Dalam algoritma ini dilakukan 2 tahapan:

1. Langkah kalibrasi: untuk mendapatkan vektor pembobot **w**, vektor skor **t**, vektor muatan **p** dan koefisien regresi internal q.
2. Langkah pendugaan: untuk mendapatkan koefisien regresi eksternal **b**, y_{duga}.

Langkah kalibrasi:

1. Melakukan pembakuan terhadap matriks data **X** dan **y**, dengan nilai awal a = 1, selanjutnya dilakukan langkah 2-7:
2. Penentuan penduga vektor pembobot **w**:
   $\mathbf{w}_a = \mathbf{x}_{a-1}`\mathbf{y}_{a-1} / ||\mathbf{X}_{a-1}`\mathbf{y}_{a-1}||$
3. *Penentuan penduga skor* ***t***
   $\mathbf{t}_a = \mathbf{X}_{a-1}\mathbf{w}_a$
4. Penentuan penduga vektor muatan **p**:
   $\mathbf{p}_a = \mathbf{X}_{a-1}`\mathbf{t}_a/\mathbf{t}_a`\mathbf{t}_a$
5. Penentuan penduga koefisien regresi internal q
   $\mathbf{q}_a = \mathbf{y}_{a-1}`\mathbf{t}_a/\mathbf{t}_a`\mathbf{t}_a$
6. Penentuan sisaan **X** dan y
   $\mathbf{E}_a = \mathbf{X}_{a-1} - \mathbf{t}_a\mathbf{p}_a`$
   $\mathbf{f}_a = \mathbf{y}_{a-1} - \mathbf{t}_a\mathbf{q}_a$
7. Membentuk **X** dan y yang baru dengan:
   $X_a = E_a$
   $y_a = f_a$
   dan a = a + 1

Sesudah langkah ke-7 kembali ke langkah 2, proses ini dilanjutkan sampai 9 iterasi, sesuai dengan banyaknya peubah x dalam model.

b. Langkah Pendugaan:

1. Perhitungan **b** untuk A buah faktor:
   $\mathbf{b}_A = \mathbf{W}_A(\mathbf{P}_A`\mathbf{W}_A)^{-1}\mathbf{q}_A$
   pada iterasi ke-A nilai W_A, P_A *dan* q_A *adalah:*
   $\mathbf{W}_A = (\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_A)$
   $\mathbf{q}_A = (\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_A)`$
   $\mathbf{P}_A = (\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_A)`$

1. Dugaan y dapat dihitung dengan 2 cara:

2.1 $\hat{y} = Xb$

2.2 $\hat{y} = \sum_{a=1}^{A} t_a q_a$

2. Perhitungan nilai KTA dan R^2 pada masing-masing iterasi.

c. Validasi Silang

Prosedur:

i = 1, 2, ..., 36

1. Menghapus pengamatan ke-i
2. Melakukan langkah kalibrasi algoritma MKTP terhadap 35 pengamatan sisanya, diperoleh untuk 9 iterasi:

   b (*i*), y (*i*)

3. Menghitung PRESS $\sum_{i=1}^{36}(y_i - \hat{y}_{(i)})^2$ untuk 9 iterasi dimana y_i = pengamatan ke–i dan $\hat{y}_{(i)}$ = dugaan y_i bila pengamatan ke-i dihapus.

4. Menentukan pada iterasi ke berapa PRESS minimal.
5. ~~Menghitung bias dari koefisien~~ regresi:
   Bias = 35 ($b_{(.)}$ – b)
   dimana: $b_{(.)} = \sum_{i=1}^{36} b_{(i)} / 36$
6. Menghitung *pseudovalues*:
   $J_i = 36\,b - 35\,b_{(i)}$
   b adalah dugaan β didasarkan pada 36 pengamatan.
   $b_{(i)}$ adalah dugaan β bila pengamatan ke-i dihapus.
7. Menghitung penduga *Jackknife* bagi β:
   $J_{(b)} = 1/36\ \Sigma\ J_i$
8. Menghitung penduga simpangan baku koefisien regresi:
   $$\hat{se}_{jack} = \{\sum_{i=1}^{36} (J_i - J_{(b)})^2 / (35)(36)\}^{1/2}$$
9. Menentukan selang kepercayaan 95% bagi β:
   $J_{(b)} \pm t_{35}(0{,}975)\ \hat{se}_{jack}$

d. Algoritma *Bootstrap*

1. Memilih secara acak sebanyak B contoh *Bootstrap* yang saling bebas $x^{*1}, x^{*2}, \ldots, x^{*B}$, masing-masing terdiri dari 36 data.
2. Menghitung dugaan parameter β pada masing-masing contoh *Bootstrap* ke r:
   $b_{(r)}$, dengan r = 1, 2, ..., B.
3. Menduga simpangan baku parameter β dengan simpangan baku contoh *Bootstrap*:
   $$\hat{se}_B = \{\sum_{r=1}^{B} [b_{(r)} - b_{(.)}]^2 / (B-1)\}^{1/2}$$, dimana
   $$b(.) = \sum_{r=1}^{B} b_{(r)} / B$$

Dipilih nilai B = 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000.

4. Menghitung bias pada masing-masing replikasi:
   $bias_B = \mathbf{b_{(.)} - b}$
5. Membuat grafik Bias vs. Replikasi B.
6. Menentukan pada replikasi berapa bias konvergen.
7. Pada replikasi tersebut nilai se_B dibaca.
8. Menentukan selang kepercayaan 95% bagi β.
   $B \pm \Phi^{-1}(0{,}975)\ \hat{se}_B$

3. Hasil dan Pembahasan

Pendugaan dengan algoritma MKTP untuk berbagai iterasi menghasilkan nilai PRESS, KTA dan R^2 yang dapat dilihat pada Tabel 1 berikut:

Tabel 1. Nilai PRESS, KTA dan R^2 pada berbagai iterasi

Iterasi	PRESS	KTA	R^2
1	0,43944	0,011126	0,610584
2	0,23217	0,005234	0,816801
3	**0,07953**	**0,001597**	**0,944106**
4	0,07522	0,001449	0,949272
5	0,07420	0,001350	0,952736
6	0,08009	0,001205	0,957814
7	0,07970	0,001180	0,958693
8	0,08296	0,001166	0,959196
9	0,08379	0,001154	0,959614

PRESS minimal dicapai pada iterasi ke-5 dengan nilai R^2 = 95,27% dan KTA 0,0013. Tetapi bila diperhatikan pada iterasi ke-3, nilai R^2 sudah baik yaitu 94,41%, selisih nilai PRESS antara iterasi ke-3 dan ke-5 relatif tidak terlalu besar (hanya sekitar 0,005).

Dugaan Persamaan regresi:

$\hat{y} = 1{,}71\ x_1 - 0{,}15\ x_2 - 2{,}2\ x_3 - 1{,}69\ x_4 - 1{,}70\ x_5 + 1{,}43\ x_6 + 1{,}81\ x_7 + 1{,}58\ x_8 - 0{,}02\ x_9$

Dengan menggunakan transformasi (Montgomery & Peck, 1992), maka:

$b_j^{(o)} = b_j\,(s_y / s_j)^{1/2}$,   j = 1,2, ..., k

$s_j = \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2$ dan $s_y = \sum_{i=1}^{n} (y_i - \bar{y})^2$

$b_0 = \bar{y} - \sum_{j=1}^{k} b_j \bar{x}_j$

diperoleh dugaan persamaan regresi dalam peubah asal:

$\hat{y}^{(o)} = 39{,}06 + 52{,}19\ x^{(o)}_1 - 5{,}11\ x^{(o)}_2 - 81{,}41\ x^{(o)}_3 - 62{,}29\ x^{(o)}_4 - 68{,}57\ x^{(o)}_5 + 59{,}31\ x^{(o)}_6 + 71{,}13\ x^{(o)}_7 + 61{,}62\ x^{(o)}_8 - 1{,}60\ x^{(o)}_9$

dimana $y^{(o)}$, $x^{(o)}_i$ (i = 1,2, ..., 9) merupakan peubah-peubah yang belum dibakukan dengan KTA = 0,001597 dan R^2 = 94,41%.

Hasil Pendugaan Ragam dengan Metode *Jacknife* dan *Bootstrap*

Perhitungan dengan metode *Jacknife* menghasilkan bias dan simpangan baku seperti Tabel 2 berikut:

Tabel 2. Bias dan Simpangan Baku Koefisien Regresi.

Penduga	Bias	se_{jack}
β_1	-0,0536	0,4282
β_2	-0,0349	0,1386
β_3	-0,0009	0,3662
β_4	0,0061	0,2900
β_5	0,0473	0,3266
β_6	0,0233	0,2184
β_7	0,0201	0,2573
β_8	0,0206	0,2283
β_9	-0,0348	0,2809

Sedangkan bias dari koefisien regresi β_1, β_2, ..., β_9 dengan metode *bootstrap* konvergen pada replikasi B = 2000 seperti Tabel 3 berikut:

Tabel 3. Bias dan Simpangan Baku Koefisie Regresi untuk B = 2000

Penduga	Bias	se_{jack}
β_1	-0,0265	0,3989
β_2	-0,0271	0,1257
β_3	-0,0141	0,3534
β_4	-0,0062	0,2741
β_5	0,0460	0,3259
β_6	0,0113	0,2174
β_7	0,0128	0,2484
β_8	0,0138	0,2244
β_9	-0,0154	0,2755

Perhitungan baik dengan metode *Jacknife* maupun *Bootstrap* menghasilkan peubah x_2 dan x_9 yang tidak nyata karena selang kepercayaan bagi β pada taraf nyata 5% maupun 1%, 0 terletak di dalam selang β_2 dan β_9. Ini mengindikasikan peubah-peubah x_2 dan x_9 dapat diabaikan dalam model.

Dengan demikian metode *Jacknife* dan *Bootstrap* keduanya menyimpulkan koefisien-koefisien β_1, β_3, β_4, ..., β_8 nyata, sedangkan β_2 dan β_9 tidak nyata. Tabel 4 menunjukkan bahwa selang kepercayaan 95% bagi koefisien regresi hasil metode *Bootstrap* lebih sempit dibandingkan selang hasil metode *Jacknife*. Ini menunjukkan bahwa hasil yang diperoleh melalui metode *Boostrap* lebih baik dari pada metode *Jacknife*.

Tabel 4. Lebar Selang Kepercayaan 95% bagi β

Penduga	*Jacknife*	*Bootstrap*
β_1	1,7396	1,5637
β_2	0,5631	0,4927
β_3	1,4877	1,3853
β_4	1,1786	1,0921
β_5	1,3268	1,2775
β_6	0,8872	0,8522
β_7	1,0454	0,9737
β_8	0,9275	0,8796
β_9	1,1413	1,0800

Selanjutnya pendugaan ragam $\hat{y}$ digunakan persentil *Bootstrap* 2,5% dan 97,5% menghasilkan selang kepercayaan bagi $\hat{y}$ yang menyempit di sekitar nilai tengah y. Nilai-nilai $\hat{y}$ terletak di dalam batas-batas selang kepercayaan 95%. Makin jauh dari nilai tengah keragaman y makin besar.

4. Kesimpulan dan Saran

Evaluasi kesesuaian model yang dilakukan dengan metode *Jacknife* dan metode *Bootstrap* keduanya menghasilkan kesimpulan yang sama yaitu koefisien regresi β_2 dan β_9 tidak nyata yang berarti peubah x_2 dan x_9 dapat diabaikan. Selang kepercayaan 95% bagi β yang dihasilkan metode *Bootstrap* lebih baik dari metode *Jacknife* (lebih sempit), tetapi perhitungan dengan metode *Bootstrap* membutuhkan waktu yang lebih lama dibandingkan metode *Jacknife* karena metode *Bootstrap* menggunakan replikasi sebanyak B = 2000 yang jauh lebih besar dibandingkan replikasi *Jacknife* n = 36. Nilai B harus dicari terlebih dahulu sedangkan untuk *Jacknife* replikasinya sudah pasti sama dengan banyaknya pengamatan. Pendugaan ragam $\hat{y}$ diperoleh melalui persentil *Bootstrap* 2,5% dan 97,5% menghasilkan selang kepercayaan bagi $\hat{y}$ yang menyempit di sekitar nilai tengah y, makin jauh dari nilai tengah y selang makin lebar.

Bagi peneliti lebih lanjut disarankan melakukan evaluasi MKTP berdasarkan kriteria pemilihan model lainnya seperti pencilan, plot sisaan, dan *leverage* karena dalam penelitian ini evaluasi terhadap penduga model hanya berdasarkan nilai KTA, PRESS dan R^2.

Daftar Pustaka

[1] Efron, B & Tibshirani, R.J. 1994. *An Introduction to the Bootstrap*. Chapman & Hall, New York.

[2] Geladi, P & Kowalski, B. 1996. Partial Least Square: a Tutorial, *Analytica Chimica Acta.*, 185:1-17.

[3] Helland, I.S. 1998. On the Structure of partial Least Square. *Commun. Statistic Simulation Comput.*, 17:581-607.

[4] Marten, H & Naes, T. 1999. *Multivariate Calibration.* Jhon Wiley & Sons. Chichester. England.

[5] Martens, H & Jensen, S.A. 1993. Partial Least Squares Regression: A New two-stage NIR calibration Method. In:*Progress in Cereal Chemistry and Technology 5a* (Proceeding, 7th world Cereal and Bread Congress, Prague, June. 1992, J.Holas, J.Kratochvit, eds) Elsevier Publ., Amsterdam, 607-647.

[6] Martens, H & Wold, H. 1993. The Multivariate Calibration Problem in Chemistry Solved by PLS Method. *Proc. Conf.matrix Pencils* (A. Ruhe, B. Kagstom, eds), March 1992, Lecture Notes in Mathematics, Springer Verlag, Heidelberg., 286-293.

[7] Martens, H. 1995. *Multivariate Calibration*. Dr. Techn. Thesis. Technical University of Norway. Trondheim.

[8] Naes, T & Martens, H. 1995. Comparison of Prediction Methods for Multicolinear Data. *Commun. Statistic Simulation Comput.* 14 :545-576.

[9] Naes, T. 1995. Multivariate Calibration when the Error Covariance Matrix is Structured. *Technometrics*. 27:301-311.

[10] Wold, H. 1993. *Soft Modeling, The basic Design and Some Extentions, in System Under Indirect Observartion, I-II,* K.G Joreskom and H. Wold, eds., Nort-Holland. Amsterdam.

[11] Young, P.Y. 2004. A Reformulation of the Partial Least Square Regression Algorithm. *Siam J. Sci. Stat. Comput.*, 15:25-230.