Statistika, Vol. 6 No. 1, 1 – 6
Mei 2006

Improvisasi Penaksir Model Linier

MULYANA
Jurusan Statistika FMIPA Unpad

1. Pendahuluan

Dalam model linier dengan asumsi kekeliruannya berdistribusi identik independen dengan rata-rata 0 dan varians konstan sama dengan σ^2, menggabungkan antara penaksir aktual (nilai ramalan) dengan nilai rata-rata merupakan segi (*aspect*) penting dalam analisis regresi terapan. Misalkan sebuah pabrik farmasi membuat obat dengan formulasi baru dan ingin menelaah daya sembuhnya jika dibandingkan dengan formulasi lama, yang tingkat (lama) kesembuhannya dipengaruhi oleh beberapa variabel pada pasien. Dalam hal ini biasanya yang ditelaah pihak produsen adalah rata-rata tingkat kesembuhan, sedangkan pasien nilai aktualnya, sehingga persoalannya bagaimana menggabungkan kedua telaahan itu secara statistika ?

2. Teori

Perhatikan model linier

$$\underline{Y} = X\underline{\beta} + \underline{e} \qquad (1)$$

dengan

$\underline{Y}$, vektor variabel respon (variabel tidak bebas) berukuran nx1

X, matriks variabel *explanatory* (variabel bebas) berukuran nxm , m < n, m > 2 , dengan rank penuh

$\underline{\beta}$, vektor parameter model berukuran mx1

$\underline{e}$, vektor kekeliruan model berukuran nx1, dengan asumsi $E(\underline{e}) = 0$, dan

$E\left(\underline{E}\underline{E}'\right) = \sigma^2 I$, I matriks identitas berukuran nxn

Dengan menggunakan metode kuadrat terkecil, penaksir untuk $\underline{\beta}$ adalah

$$\hat{\underline{\beta}} = (X'X)^{-1} X'\underline{Y} \qquad (2)$$

yang merupakan statistik tak-bias dan bervarians minimum, sehingga penaksir aktual untuk $\underline{Y}$ adalah $\hat{\underline{Y}} = X\hat{\underline{\beta}}$.

Karena $E(\underline{Y}) = X\underline{\beta}$, maka berdasarkan sifat kelinieran, penaksir untuk rata-rata $\underline{Y}$, $E(\underline{Y})$ adalah $E(\hat{\underline{Y}}) = X\hat{\underline{\beta}}$, sehingga dari paparan tersebut tersurat bahwa $X\hat{\underline{\beta}}$ memiliki peran dua penaksir, yaitu sebagai penaksir nilai aktual dan nilai rata-rata untuk $\underline{Y}$. Persoalannya bagaimana menyajikan statistik $X\hat{\underline{\beta}}$ jika diinginkan perannya lebih dominan sebagai penaksir nilai aktual dari pada sebagai nilai rata-rata atau sebaliknya? Berdasarkan teori Statistika-Matematis, untuk keperluan tersebut diperlukan formulasi dari jumlah kuadrat kekeliruan model, agar bisa dibangun fungsi target beserta fungsi kegagalan (*loss function*) dan fungsi resikonya (*risk function*).

Jumlah kuadrat kekeliruan model adalah

$$\underline{e}'\underline{e} = \left(\underline{Y} - X\underline{\beta}\right)'\left(\underline{Y} - X\underline{\beta}\right) \qquad (3)$$

yang jika dijabarkan akan diperoleh persamaan

$$\underline{e}'\underline{e} = \left(\underline{Y} - X\hat{\underline{\beta}}\right)'\left(\underline{Y} - X\hat{\underline{\beta}}\right) + \left(X\underline{\beta} - X\hat{\underline{\beta}}\right)'\left(X\underline{\beta} - X\hat{\underline{\beta}}\right) \tag{4}$$

Pada Persamaan (4) tersurat, fungsi kegagalan untuk $X\hat{\underline{\beta}}$ jika digunakan sebagai penaksir $X\underline{\beta}$ dibangun atas kombinasi linier yang diboboti dengan persamaan

$$f\left(X\underline{\beta}, X\hat{\underline{\beta}}\right) = c\left(\underline{Y} - X\hat{\underline{\beta}}\right)'\left(\underline{Y} - X\hat{\underline{\beta}}\right) + (1-c)\left(X\underline{\beta} - X\hat{\underline{\beta}}\right)'\left(X\underline{\beta} - X\hat{\underline{\beta}}\right) \tag{5}$$

c: skalar nonstokastik, 0 < c < 1

Karena suku pertama pada Persamaan (4) merupakan jumlah kuadrat penaksir nilai aktual $\underline{Y}$ dan suku keduanya jumlah kuadrat residu $E(\underline{Y})$, sehingga fungsi target untuk $\underline{Y}$ dapat dibangun berdasarkan persamaan

$$\underline{T} = (1-\lambda)\underline{Y} + \lambda E(\underline{Y}) \tag{6}$$

λ : skalar nonstokastik, 0 < λ < 1

Dapat ditunjukan bahwa $\hat{\underline{T}} = (1-\lambda)\hat{\underline{Y}} + \lambda\, E(\hat{\underline{Y}}) = X\hat{\underline{\beta}}$ dan $E(\underline{T}) = X\underline{\beta}$ yang berarti $E(\hat{\underline{T}}) = X\hat{\underline{\beta}}$, sehingga $\underline{T}$ identik dengan $\underline{Y}$. Fungsi kegagalan untuk $\hat{\underline{T}}$ jika digunakan sebagai penaksir $\underline{T}$ sama dengan

$$\begin{aligned} f\left(\underline{T}, \hat{\underline{T}}\right) &= \left(\underline{T} - X\hat{\underline{\beta}}\right)'\left(\underline{T} - X\hat{\underline{\beta}}\right) \\ &= (1-\lambda)^2\left(\underline{Y} - X\hat{\underline{\beta}}\right)'\left(\underline{Y} - X\hat{\underline{\beta}}\right) + \lambda^2\left(X\underline{\beta} - X\hat{\underline{\beta}}\right)'\left(X\underline{\beta} - X\hat{\underline{\beta}}\right) \\ &\quad + 2\lambda(1-\lambda)\left(\underline{Y} - X\hat{\underline{\beta}}\right)'\left(X\underline{\beta} - X\hat{\underline{\beta}}\right) \end{aligned} \tag{7}$$

Pada Persamaan (7) tersurat, dua suku pertamanya identik dengan Persamaan (5) dan suku ketiganya merupakan kovarian yang diboboti antara residu nilai aktual dengan residu rata-rata hitung $\underline{Y}$. Sehingga Persamaan (7) merupakan pengembangan sederhana (*simple extention*) dari Persamaan (5), yang berarti Persamaan (7) merupakan fungsi kegagalan untuk $X\hat{\underline{\beta}}$ (jika digunakan sebagai penaksir $X\underline{\beta}$) yang sebaiknya digunakan, dengan fungsi resiko sama dengan

$$\begin{aligned} E\left(f\left(\underline{T}, \hat{\underline{T}}\right)\right) &= (1-\lambda)^2 E\left(\underline{Y} - X\hat{\underline{\beta}}\right)'\left(\underline{Y} - X\hat{\underline{\beta}}\right) + \lambda^2 E\left(X\underline{\beta} - X\hat{\underline{\beta}}\right)'\left(X\underline{\beta} - X\hat{\underline{\beta}}\right) \\ &\quad + 2\lambda(1-\lambda) E\left(\underline{Y} - X\hat{\underline{\beta}}\right)'\left(\underline{Y} - X\hat{\underline{\beta}}\right) \end{aligned}$$

yang sama dengan total dari penaksir rata-rata jumlah kuadrat kekeliruan (*total predictive mean square error*). Dari paparan ini disimpulkan bahwa formulasi untuk menggabungkan antara penaksir nilai aktual dengan rata-rata hitungnya harus mengikuti Persamaan (6).

Shalabh (1999) mengemukakan, menggunakan fungsi resiko di bawah fungsi kegagalan dengan Persamaan (7) dapat digunakan dua bentuk penaksir untuk $\underline{\beta}$, yaitu penaksir kuadrat terkecil seperti pada Persamaan (2), dan penaksir berdasarkan aturan Stein (*Stein-rule estimator*), yang persamaannya

$$\underline{\hat{\beta}}_S = \left(1-\left(\frac{a}{n-m+2}\right)\frac{\underline{Y}'H^C\underline{Y}}{\underline{Y}'H\underline{Y}}\right)\underline{\hat{\beta}} \qquad (8)$$

a: a > 0, skalar karakaterisasi penaksir

$H = X(X'X)^{-1}X'$, $H^C = I - H$, I matriks identitas

Dapat ditunjukan $\underline{\hat{\beta}}_S$ bukan penaksir takbias, dan akan merupakan penaksir takbias jika $\underline{Y}'H^C\underline{Y} = 0$, yaitu jika $\underline{\hat{Y}} = \underline{Y}$ (model regresi sangat cocok sebagai model ramalan), sehingga dalam penggunaannya harus dikombinasi linierkan dengan penaksir kuadrat terkecil, dengan persamaan

$$\underline{b} = (1-w)\underline{\hat{\beta}} + w\underline{\hat{\beta}}_S \qquad (9)$$

0<w<1, skalar nonstokastik

Jadi dalam hal ini penaksir untuk $X\underline{\beta}$, bisa digunakan

$$\underline{p} = X\underline{\hat{\beta}} = X(X'X)^{-1}X'\underline{Y} = H\underline{Y} \qquad (9)$$

atau

$$\underline{p}_S = X\underline{\hat{\beta}}_S = X\left\{1-\left(\frac{a}{n-m+2}\right)\frac{\underline{Y}'H^C\underline{Y}}{\underline{Y}'H\underline{Y}}\right\}\underline{\hat{\beta}} = H\underline{Y} - \left(\frac{a}{n-m+2}\right)\frac{\underline{Y}'H^C\underline{Y}}{\underline{Y}'H\underline{Y}}H\underline{Y} \qquad (10)$$

atau kombinasi liniernya

$$\begin{aligned}\underline{P} &= (1-w)\underline{p} + w\underline{p}_S \\ &= (1-w)H\underline{Y} + w\left\{H\underline{Y} - \left(\frac{a}{n-m+2}\right)\frac{\underline{Y}H^C\underline{Y}}{\underline{Y}H\underline{Y}}H\underline{Y}\right\} \\ &= H\underline{Y} - w\left(\frac{a}{n-m+2}\right)\frac{\underline{Y}H^C\underline{Y}}{\underline{Y}H\underline{Y}}H\underline{Y}\end{aligned} \qquad (11)$$

yang formulasinya setara dengan Persamaan (10).

Jika $\underline{p}$ dan $\underline{P}$ digunakan sebagai penaksir untuk fungsi target $\underline{T}$, maka

$$\begin{aligned}E(\underline{T}-\underline{p}) &= E\{(1-\lambda)\underline{Y} + \lambda E\underline{Y} - H\underline{Y}\} = (1-\lambda)E\underline{Y} + \lambda E\{E(\underline{Y})\} - HE\underline{Y} \\ &= (1-\lambda)E\underline{Y} + \lambda E(\underline{Y}) - X(X'X)^{-1}X'X\underline{\beta} = E\underline{Y} - X\underline{\beta} = X\underline{\beta} - X\underline{\beta} = \underline{0}\end{aligned} \qquad (12)$$

dan

$$\begin{aligned}E(\underline{T}-\underline{P}) &= E\left\{(1-\lambda)\underline{Y} + \lambda E\underline{Y} - H\underline{Y} + w\left(\frac{a}{n-m+2}\right)\frac{\underline{Y}'H^C\underline{Y}}{\underline{Y}'H\underline{Y}}H\underline{Y}\right\} \\ &= E\{(1-\lambda)\underline{Y} + \lambda E\underline{Y} - H\underline{Y}\} + Ew\left(\frac{a}{n-m+2}\right)\frac{\underline{Y}'H^C\underline{Y}}{\underline{Y}'H\underline{Y}}H\underline{Y} \\ &= w\left(\frac{a}{n-m+2}\right)\frac{\underline{Y}'H^C\underline{Y}}{\underline{Y}'H\underline{Y}}HE\underline{Y} = w\left(\frac{a}{n-m+2}\right)\frac{\underline{Y}'H^C\underline{Y}}{\underline{Y}'H\underline{Y}}X\underline{\beta}\end{aligned} \qquad (13)$$

karena $0 < w\left(\frac{a}{n-m+2}\right)\frac{\underline{Y}H^{C}\underline{Y}}{\underline{Y}H\underline{Y}} < 1$, maka $E(\underline{T}-\underline{p}) < E(\underline{T}-\underline{P})$

Fungsi resiko jika $\underline{p}$ dan $\underline{P}$ digunakan sebagai penaksir untuk fungsi target $\underline{T}$, masing-masing sama dengan

$$R(\underline{p}) = E(\underline{T}-\underline{p})'(\underline{T}-\underline{p}) = \left\{(1-\lambda)^2 n - (1-2\lambda)m\right\}\sigma^2 \tag{14}$$

$$R(\underline{P}) = E(\underline{T}-\underline{P})'(\underline{T}-\underline{P}) = \left\{(1-\lambda)^2 n - (1-2\lambda)m\right\}\sigma^2 - wa\frac{n-m}{n-m+2}E\left(\frac{1}{\underline{Y}'H\underline{Y}}\right)\left\{2\lambda(m-2) - wa\right\}\sigma^4 \tag{15}$$

Dari Persamaan (12), (13), (14) dan (15) dapat disimpulkan

1.jika $\lambda = 0$, maka

$\underline{T} = \underline{Y}$, $R(\underline{p}) = (n-m)\sigma^2$, $R(\underline{P}) = (n-m)\sigma^2 + w^2a^2\frac{n-m}{n-m+2}E\left(\frac{1}{\underline{Y}'H\underline{Y}}\right)\sigma^4$, sehingga $R(\underline{p}) \le R(\underline{P})$. Hal ini berarti $\underline{p}$ superior dari $\underline{P}$ jika $\underline{p}$ digunakan sebagai penaksir nilai aktual $\underline{Y}$, dan karena $\underline{p} = X\hat{\underline{\beta}}$, maka jika $X\hat{\underline{\beta}}$ digunakan sebagai penaksir nilai aktual $\underline{Y}$ maka nilai resikonya sama dengan $(n-m)\hat{\sigma}^2$, $\hat{\sigma}^2$ varians residu

2.jika $0 < \lambda < 1$, dan jika $wa < 2\lambda(m-2)$ atau $a < \frac{2\lambda(m-2)}{w}, m > 2$ maka $wa\left(\frac{n-m}{n-m+2}\right)E\left(\frac{1}{\underline{Y}'H\underline{Y}}\right)\left\{2\lambda(m-2) - wa\right\} > 0$

sehingga $R(\underline{p}) \ge R(\underline{P})$. Hal ini berarti $\underline{P}$ superior dari $\underline{p}$ jika $\underline{P}$ digunakan sebagai penaksir rata-rata $\underline{Y}$, yang berarti $X\hat{\underline{\beta}}$ tidak sepenuhnya dapat dijadikan penaksir aktual $\underline{Y}$ karena terkombinasi dengan sebagai penaksir rata-rata

Dari paparan tersebut, disimpulkan jika $X\hat{\underline{\beta}}$ digunakan sebagai penaksir nilai aktual $\underline{Y}$, maka resikonya $(n-m)\hat{\sigma}^2$, $\hat{\sigma}^2$ varians residu, dan nilai ini akan cukup kecil jika model regresi cocok digunakan sebagai model ramalan, dan untuk penaksir rata-rata $\underline{Y}$, $E(\underline{Y})$, sebaiknya digunakan statistik

$$\underline{P} = X\hat{\underline{\beta}} - w\left(\frac{a}{n-m+2}\right)\frac{\underline{Y}H^{C}\underline{Y}}{\underline{Y}H\underline{Y}}X\hat{\underline{\beta}}$$

karena nilai resikonya lebih kecil dari $(n-m)\hat{\sigma}^2$, jika $a < \frac{2\lambda(m-2)}{w}, m > 2$.

3. Terapan

Untuk menggunakan teori ini diperlukan dua kelompok sampel yang identik, dengan sampel kedua merupakan sampel lanjutan dari sampel pertama, misalnya untuk kasus pabrik farmasi seperti yang dikemukan pada pendahuluan, sampel pertama adalah tingkat penyembuhan obat dengan formulasi lama, dan yang kedua dengan formulasi baru.

Jika model linier sampel pertama disajikan seperti pada Persamaan (1), maka untuk sampel kedua oleh

$$\underline{Y}_f = X_f \underline{\beta} + \underline{e}_f \tag{16}$$

dengan $\underline{Y}_t$ vektor berukuran kx1, X_f matriks berukuran kxm dengan rank penuh, $m < k < n$.

Sudah dikemukakan pada teori, pada sampel pertama penaksir untuk $X\underline{\beta}$ adalah

$$\underline{p} = X(X'X)^{-1}X'\underline{Y} \text{ atau } \underline{P} = (1-w)\underline{p} - w\left(\frac{a}{n-m+2}\right)\frac{\underline{Y}'H^C\underline{Y}}{\underline{Y}H\underline{Y}}\underline{p}$$

dengan $a < \frac{2\lambda(m-2)}{w}, m > 2$, $H = X(X'X)^{-1}X'$, $H^C = I - H$, I matriks identitas berukuran mxm

dan fungsi targetnya

$$\underline{T} = (1-\lambda)\underline{Y} + \lambda E(\underline{Y}) = \underline{Y} - \lambda\{\underline{Y} - E(\underline{Y})\}$$

sehingga pada sampel kedua penaksir untuk $X_f\underline{\beta}$ dapat digunakan

$$\underline{p}_f = X_f\left(X_f{}'X_f\right)^{-1}X_f{}'\underline{Y}_f \text{ atau } \underline{P}_f = (1-w)\underline{p}_f - w\left(\frac{a}{k-m+2}\right)\frac{\underline{Y}_f H_f{}^C \underline{Y}_f}{\underline{Y}_f H_f \underline{Y}_f}\underline{p}_f$$

dengan $a < \frac{2\lambda(m-2)}{w}, m > 2$, $H_f = X_f\left(X_f{}'X_f\right)^{-1}X_f{}'$, $H_f{}^C = I_f - H_f$, I_f matriks identitas berukuran kxk.

dan fungsi targetnya

$$\underline{T}_f = (1-\lambda)\underline{Y}_f + \lambda E(\underline{Y}_f) = \underline{Y}_f - \lambda\{\underline{Y}_f - E(\underline{Y}_f)\}$$

Pada sampel pertama, jika $\underline{p}$ sebagai penaksir $\underline{T}$ (atau $X\hat{\underline{\beta}}$ sebagai nilai ramalan $\underline{Y}$, $\hat{\underline{Y}} = X\hat{\underline{\beta}}$) maka nilai resikonya sama dengan $(n-m)\hat{\sigma}^2$, $\hat{\sigma}^2$ varians residu sampel pertama, dan jika $\underline{P}$ sebagai penaksir $\underline{T}$ (atau $\underline{P}$ sebagai penaksir rata-rata $\underline{Y}$, $E(\hat{\underline{Y}}) = \underline{P}$) maka nilai resikonya lebih kecil dari $(n-m)\hat{\sigma}^2$. Analog untuk sampel kedua, jika $\underline{p}_f$ sebagai penaksir $\underline{T}_f$ (atau $X_f\hat{\underline{\beta}}$ sebagai nilai ramalan $\underline{Y}_f$, $\hat{\underline{Y}}_f = X_f\hat{\underline{\beta}}$) maka nilai resikonya sama dengan $(k-m)\hat{\sigma}_f^2$, $\hat{\sigma}_f^2$ varians residu model sampel kedua, dan jika $\underline{P}_f$ sebagai penaksir $\underline{T}_f$ (atau $\underline{P}_f$ sebagai penaksir rata-rata $\underline{Y}_f$, $E(\hat{\underline{Y}}_f) = \underline{P}_f$) maka nilai resikonya lebih kecil dari $(k-m)\hat{\sigma}_f^2$. Sehingga jika hasil penaksiran pada sampel pertama dan kedua digabungkan, maka penaksir nilai aktual respon sama dengan

$$(1-\lambda)\underline{p}_f + \lambda\underline{p}\ ,$$

dan penaksir rata-ratanya sama dengan

$$(1-\lambda)\underline{P}_f + \lambda\underline{P}.$$

4. Daftar Pustaka

Searle, S. R., 1971, Linear Models, John Wiley & Sons, New York.

Drafer, N. & Smith, H., 1981, *Applied Regression Analysis*, second edition, John Wiley & Sons, New York.

Shalabh, 1999, *Improving The Prediction in Linear Regression Models*, Journal of Statistical Research, Vol. 3 No. 1 pp 33 – 39, Bangladesh.

Graybill, F. A. , 1961, *An Introduction to Linear Statistical Models*, McGraw-Hill Book Co. Inc., New York

Berger, J. O., 1985, *Statistical Decision Theory and Bayesian Analysis*, second edition, Springer-Verlag, New York.

Ohtani, K., 1998, *The Excact Risk of Weighted Average Estimator of OLS and Stein-rule Estimator in Regression under Balanced Loss*, Statistics & Decisions, Vol. 16, pp 35–45.

Hogg, R. V. & Craig, A. T., 1978, *Introduction to Mathematical Statistics*, fourth edition, Macmillan Pub. Co. Inc., New York.