Prosiding Seminar Nasional Penelitian, Pendidikan dan Penerapan MIPA,
Fakultas MIPA, Universitas Negeri Yogyakarta, 14 Mei 2011

ANALISIS PEUBAH PREDIKTOR YANG MEMUAT KESALAHAN PENGUKURAN DENGAN REGRESI ORTOGONAL

Kismiantini

Jurusan Pendidikan Matematika, FMIPA Universitas Negeri Yogyakarta
Email: kismi@uny.ac.id

Abstrak

Analisis regresi linear sederhana adalah suatu analisis yang digunakan untuk mengetahui hubungan antara satu peubah prediktor dan satu peubah respons. Pada model regresi linear sederhana, peubah prediktor dianggap tetap (tidak memiliki distribusi) sedangkan peubah respons mengikuti distribusi normal. Bila peubah prediktor memuat kesalahan pengukuran sehingga memuat galat yang memiliki distribusi maka model regresi linear sederhana tidak tepat digunakan. Pada makalah ini akan mengkaji alternatif regresi yang mampu mengatasi permasalahan peubah prediktor yang memuat kesalahan pengukuran dengan menggunakan regresi ortogonal.

Kata kunci: peubah prediktor, kesalahan pengukuran, regresi ortogonal

PENDAHULUAN

Analisis regresi adalah suatu analisis yang digunakan untuk mengetahui hubungan antara satu atau lebih peubah prediktor dengan peubah respons. Model regresi linear sederhana adalah model yang paling sederhana untuk menjelaskan hubungan antara satu peubah prediktor (X) dan satu peubah respons (Y). Pada model regresi linear klasik, peubah prediktor diasumsikan diketahui (*fixed*) sehingga diperoleh tanpa adanya galat sedangkan peubah respons diasumsikan berdistribusi normal

Apabila peubah prediktor memuat kesalahan pengukuran (galat) maka model regresi linear sederhana kurang tepat digunakan. Alternatif regresi yang dapat digunakan untuk mengatasi permasalahan ini adalah regresi ortogonal. Pada model regresi ortogonal, baik peubah prediktor maupun peubah respons adalah peubah acak.

PEMBAHASAN

Regresi ortogonal digunakan untuk mengetahui hubungan antara peubah prediktor (X) dan peuah respons (Y) bila pada peubah prediktor tersebut mengandung kesalahan pengukuran (galat). Pada analisis regresi ortogonal, kedua peubah tersebut merupakan peubah kontinu. Regresi ortogonal ini sering digunakan pada data-data yang diperoleh dari hasil pengukuran seperti di suatu laboratorium.

Misalkan seorang peneliti yang bekerja di suatu laboratorium ingin mengetahui apakah suatu teknik baru yang relatif lebih murah untuk pengujian kadar glukose memberikan hasil pengukuran yang sama atau berbeda dengan teknik standar. Dalam hal ini, kadar glukose yang dihasilkan oleh teknik baru (Y) dimungkinkan terjadinya kesalahan pengukuran. Peubah prediktor adalah kadar glukose yang dihasilkan oleh teknik standar (X). Teknik baru ini akan digunakan sebagai pengganti teknik standar apabila menghasilkan pengukuran kadar glukose yang sama dengan teknik standar, sehingga regresi ortogonal lebih cocok digunakan daripada regresi linear sederhana.

Model Regresi Linear Sederhana

Model regresi linear sederhana adalah model regresi yang digunakan untuk mengetahui hubungan antara satu peubah prediktor dan peubah respons, dengan peubah prediktor diasumsikan tetap (*fixed*). Model regresi linear sederhana dinyatakan sebagai berikut:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad \varepsilon_i \overset{iid}{\sim} N\left(0, \sigma_\varepsilon^2\right) \tag{1}$$

dengan Y_i adalah peubah respons pada pengamatan ke-*i*, β_0 dan β_1 adalah parameter regresi, X_i adalah peubah prediktor pada pengamatan ke-*i*, ε_i adalah galat (kesalahan pengukuran) pada pengamatan ke-*i*.

Estimator β_0 dan β_1 dapat diperoleh dengan metode maksimum likelihood (Bain & Engelhardt, 1992: 508). Berikut adalah fungsi likelihood:

$$L = L(\beta_0, \beta_1) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(Y_i - \beta_0 - \beta_1 X_i)^2\right] \tag{2}$$

Selanjutnya dilogaritmanaturalkan sehingga diperoleh fungsi log-likelihood sebagai berikut

$$\ln L = -\frac{n}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i)^2 \tag{3}$$

Lalu diturunkan terhadap β_0 dan β_1 dan disamadenganakan nol maka diperoleh persamaan maksimum likelihood berikut:

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} Y_i \Rightarrow \hat{\beta}_0 = \frac{\sum_{i=1}^{n} Y_i}{n} - \hat{\beta}_1 \frac{\sum_{i=1}^{n} X_i}{n} = \overline{Y} - \hat{\beta}_1 \overline{X} \tag{4}$$

$$\hat{\beta}_0 \sum_{i=1}^{n} X_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} X_i Y_i \Rightarrow \hat{\beta}_1 = \frac{\sum_{i=1}^{n} X_i Y_i - \frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n}}{\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}} = \frac{S_{XY}}{S_{XX}} \tag{5}$$

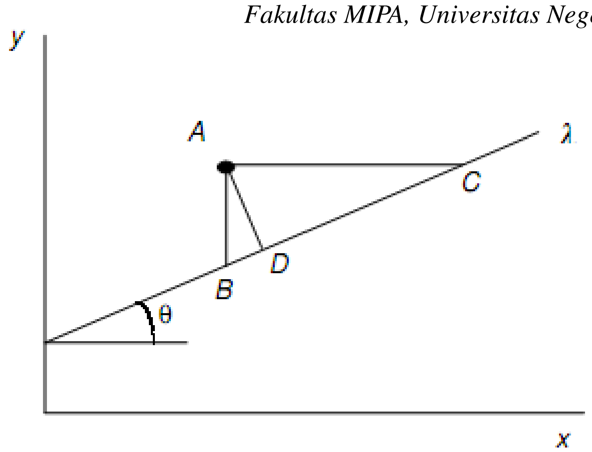
Estimasi Parameter Regresi dengan *Orthogonal Least Squares*

Estimator bagi β_0 dan β_1 pada model regresi linear sederhana (1) dapat diperoleh dengan metode *orthogonal least squares* (Dessanaike & Wang, 2003: 22-23) yaitu meminimumkan jarak antara pengamatan dengan garis dugaan. Dari Gambar 1 berikut diperoleh bahwa

$$\beta_1 = tg(\theta) \tag{6}$$

dan jarak kuadrat antara pengamatan $A(X_i, Y_i)$ dan garis dugaan adalah

$$AD^2 = [\cos(\theta)(Y_i - \beta_0 - \beta_1 X_i)]^2 = \frac{(Y_i - \beta_0 - \beta_1 X_i)^2}{1 + \beta_1^2} \tag{7}$$



Gambar 1. Ilustrasi dari estimasi parameter dengan metode *orthogonal least squares*

Selanjutnya digunakan metode *orthogonal least squares* untuk meminimumkan $L=\sum_{i=1}^{n}\frac{(Y_i-\beta_0-\beta_1X_i)^2}{1+\beta_1^2}$, lalu diturunkan secara parsial terhadap β_0 dan β_1 turunan tersebut dan disamadengankan nol yaitu

$$\frac{\partial L}{\partial \beta_0}=\sum_{i=1}^{n}\frac{-2(Y_i-\beta_0-\beta_1X_i)}{1+\beta_1^2}=0 \qquad (8)$$

$$\frac{\partial L}{\partial \beta_1}=\sum_{i=1}^{n}\frac{-2(1+\beta_1^2)(Y_i-\beta_0-\beta_1X_i)X_i-2\beta_1(Y_i-\beta_0-\beta_1X_i)^2}{(1+\beta_1^2)^2}=0 \qquad (9)$$

Dari persamaan (4) diperoleh

$$\hat{\beta}_0=\overline{Y}-\hat{\beta}_1\overline{X} \qquad (10)$$

dengan $\overline{X}$ menyatakan rata-rata sampel *X* dan $\overline{Y}$ adalah rata-rata sampel *Y*. Selanjutnya dari persamaan (9) dan (10) diperoleh

$$\hat{\beta}_1^2S_{XY}-\hat{\beta}_1\left(S_{YY}^2-S_{XX}^2\right)-S_{XY}=0 \qquad (11)$$

dengan S_{XX}^2 dan S_{YY}^2 berturut-turut adalah variansi sampel dari *X* dan *Y*, sedangkan S_{XY} adalah kovariansi sampel *X* dan *Y*. Sehingga solusi dari persamaan (11) untuk $\hat{\beta}_1$ adalah

$$\hat{\beta}_1=\frac{\left(S_{YY}^2-S_{XX}^2\right)\pm\sqrt{\left(S_{YY}^2-S_{XX}^2\right)^2-4S_{XY}^2}}{2S_{XY}} \qquad (12)$$

Namun demikian pembilang pada persamaan (12) harus positif sehingga estimator β_1 adalah

$$\hat{\beta}_1=\frac{\left(S_{YY}^2-S_{XX}^2\right)+\sqrt{\left(S_{YY}^2-S_{XX}^2\right)^2-4S_{XY}^2}}{2S_{XY}} \qquad (13)$$

dengan

$$\overline{X}=\frac{\sum_{i=1}^{n}X_i}{n},\ \overline{Y}=\frac{\sum_{i=1}^{n}Y_i}{n},\ S_{XX}=\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2,\ S_{YY}=\sum_{i=1}^{n}\left(Y_i-\overline{Y}\right)^2;\ S_{XY}=\sum_{i=1}^{n}\left(X_i-\overline{X}\right)\left(Y_i-\overline{Y}\right).$$

Model Regresi Ortogonal

Misal *X* dan *Y* memuat komponen galat δ dan ε, dengan komponen galat tersebut berasal dari kesalahan pengukuran, sehingga model regresi ortogonal (Leng *et al.*, 2007: 1-2) adalah

$$\begin{aligned} X &= \xi + \delta, \quad \delta \sim N\left(0, \sigma_\delta^2\right) \\ Y &= \eta + \varepsilon, \quad \varepsilon \sim N\left(0, \sigma_\varepsilon^2\right) \end{aligned} \qquad (14)$$

$\eta = \beta_0 + \beta_1 \xi$ dengan ξ, δ dan ε saling bebas.

dan diasumsikan bahwa *X* dan *Y* mengikuti distribusi normal bivariat,

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} \mu \\ \beta_0 + \beta_1 \mu \end{pmatrix}, \begin{pmatrix} \tau^2 + \sigma_\delta^2 & \beta_1 \tau^2 \\ \beta_1 \tau^2 & \beta_1^2 \tau^2 + \sigma_\varepsilon^2 \end{pmatrix}\right) \qquad (15)$$

Sehingga,

$$\begin{aligned} E(X) &= E(\xi) + E(\delta) = \mu \\ E(Y) &= E(\eta) + E(\varepsilon) = \beta_0 + \beta_1 \mu \\ Var(X) &= Var(\xi) + Var(\delta) = \tau^2 + \sigma_\delta^2 \\ Var(Y) &= Var(\eta) + Var(\varepsilon) = Var(\beta_1 \xi) + Var(\varepsilon) = \beta_1^2 \tau^2 + \sigma_\varepsilon^2 \\ Cov(X, Y) &= Cov(\xi + \delta, \beta_0 + \beta_1 \xi + \varepsilon) = \beta_1 \tau^2 \end{aligned} \qquad (16)$$

Selanjutnya estimator β_1 dapat diperoleh dengan metode maksimum likelihood. Namun demikian tergantung oleh rasio dua variansi galat yaitu $\lambda = \sigma_\varepsilon^2 / \sigma_\delta^2$. Sehingga estimator maksimum likelihood dari β_1 (Leng *et al.*, 2007: 1-2) adalah

$$\hat{\beta}_1 = \frac{S_{YY} - \lambda S_{XX} + \sqrt{\left(S_{YY} - \lambda S_{XX}\right)^2 + 4\lambda S_{XY}^2}}{2S_{XY}} \qquad (17)$$

Penurunan selengkapnya dapat dilihat di Fuller (1987: 13-16). Bila λ = 1, maka persamaan (17) sama dengan persamaan (13). Apabila σ_ε^2 dan σ_δ^2 tidak diketahui maka dapat diestimasi dengan ragam sampel dari *Y* dan *X* (Carroll & Ruppert, 1994: 7).

Ilustrasi Regresi Ortogonal

Sebuah perusahaan peralatan medis ingin menentukan apakah alat pemonitor tekanan darah yang baru mereka ciptakan setara dengan alat pemonitor tekanan darah yang telah beredar di pasaran. Dari sampel acak 60 orang yang diukur tekanan darah sistolik dengan menggunakan dua alat pemonitor tersebut. Peubah respons adalah tekanan darah sistolik (dalam mmHg) yang diperoleh dari alat pemonitor baru, sedangkan peubah prediktor adalah tekanan darah sistolik yang diperoleh dari alat pemonitor yang beredar di pasaran. Berdasarkan studi yang dilakukan sebelumnya, perusahaan mengetahui bahwa rasio variansi galat adalah 0,90. Berikut datanya:

Tabel 1. Data tekanan darah sistolik

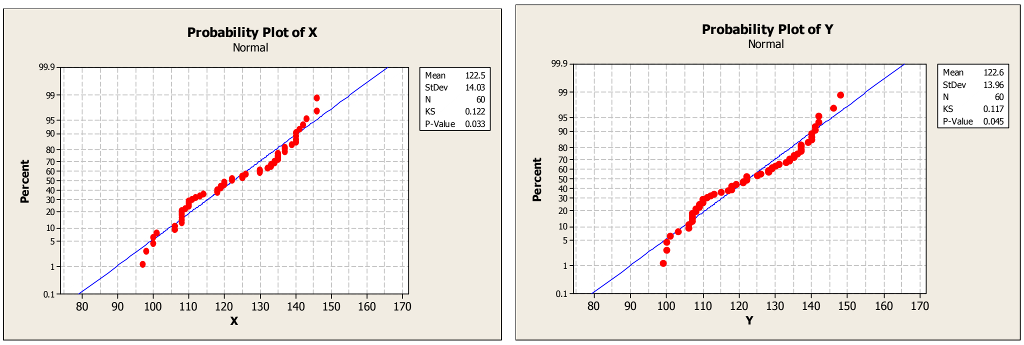
No	Baru	Pasaran	No	Baru	Pasaran	No	Baru	Pasaran
1	100	100	21	107	108	41	99	97
2	122	120	22	113	113	42	130	130
3	129	132	23	128	130	43	134	134
4	136	139	24	142	142	44	142	143
5	110	110	25	109	109	45	115	114
6	111	110	26	103	100	46	107	108

7	137	137	27	122	122	47	119	119
8	134	133	28	133	134	48	135	135
9	141	140	29	140	140	49	141	141
10	112	112	30	107	111	50	118	118
11	110	110	31	106	108	51	107	106
12	121	120	32	128	130	52	122	122
13	131	133	33	125	125	53	137	137
14	140	140	34	140	140	54	146	146
15	118	119	35	118	118	55	100	98
16	108	108	36	108	108	56	106	106
17	129	126	37	121	118	57	126	125
18	137	135	38	117	120	58	137	137
19	135	135	39	139	135	59	148	146
20	109	108	40	101	101	60	110	110

Data diambil dari worksheet Minitab 16 (BLOODPRESSURE.MTW)

Analisis data menggunakan Minitab 16 yang digunakan untuk

1. menyelidiki apakah peubah prediktor (*X*) dan peubah respons (*Y*) masing-masing berdistribusi normal atau tidak dengan uji kolmogorov smirnov
2. menyelidiki pemenuhan asumsi-asumsi pada regresi ortogonal
3. mengestimasi parameter regresi dan inferensi parameter pada model regresi ortogonal
4. membuat plot *X* dan *Y* beserta garis regresi



(a) (b)

Gambar 2. Plot peluang normal

Berdasarkan Gambar 2 dilakukan pengujian normalitas dengan menggunakan uji Kolmogorov-Smirnov, berikut langkah pengujiannya.

Hipotesis

H_0 : Data berdistribusi normal

H_1 : Data tidak berdistribusi normal

Taraf nyata : α = 0,01

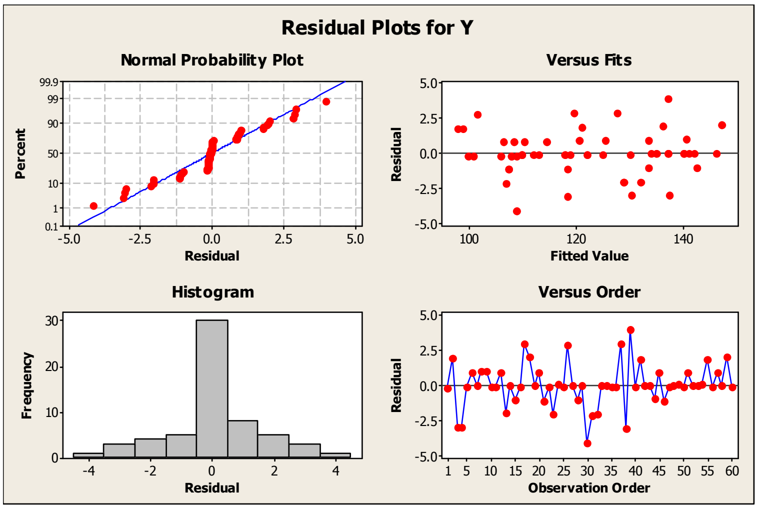
Statistik Uji : $D_n = \sup_x \left| F(x_n) - F(x) \right|$

Kriteria keputusan: H_0 ditolak jika p-value > 0,01

Hitungan:

Peubah	*X*	*Y*
Kolmogorov-Smirnov	0,122	0,117
p-value	0,033	0,045
Kesimpulan	H_0 diterima	H_0 diterima

Kesimpulan:
Karena *p-value* untuk masing-masing peubah kurang dari 0,05 maka dapat disimpulkan bahwa kedua peubah mengikuti distribusi normal.



Gambar 3. Plot residual bagi peubah respons

Dari Gambar 3, asumsi galat berdistribusi normal terpenuhi yang ditunjukkan oleh titik-titik residual pada gambar *normal probability plot* mengikuti garis diagonal, asumsi galat memiliki ragam yang konstan terpenuhi yang ditunjukkan dengan oleh titik-titik residual pada gambar *versus fits* yang tidak membentuk pola tertentu, dan asumsi galat saling bebas juga terpenuhi yang ditunjukkan oleh titik-titik sisaan pada gambar *versus order* yang acak. Semua asumsi dalam model regresi ortogonal terpenuhi maka dapat dilakukan inferensi terhadap parameter regresinya.

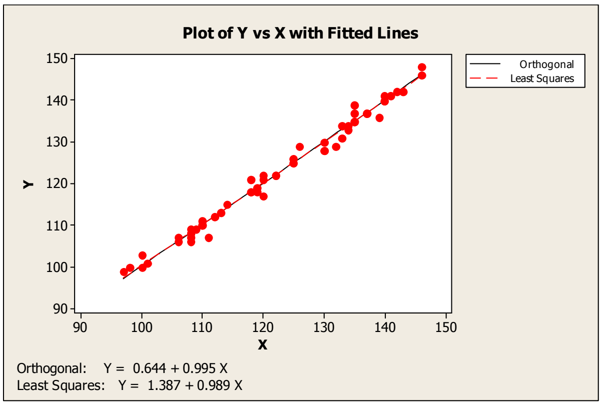
Tabel 2. Ringkasan output Minitab 16

Model	Koefisien	*Standard Error*	Z	*p-value*	Selang Kepercayaan 99% bagi β_i	*s*
Regresi Ortogonal $\lambda = 0{,}90$	$b_0 = 0{,}644$ $b_1 = 0{,}995$	$s\{b_0\} = 1{,}745$ $s\{b_1\} = 0{,}014$	0,369 70,346	0,712 0,000	(-3,850; 5,138) (0,959; 1,032)	1,131
	Variansi Galat untuk $Y = \hat{\sigma}_{\varepsilon}^2 = 1{,}079$; Variansi Galat untuk $X = \hat{\sigma}_{\delta}^2 = 1{,}198$					
Regresi Linear Sederhana	$b_0 = 1{,}387$ $b_1 = 0{,}989$	$s\{b_0\} = 1{,}734$ $s\{b_1\} = 0{,}014$	0,800 70,350	0,427 0,000	(-4,633; 4,700) (0,961; 1,037)	1,516

Berdasarkan Tabel 2, pada model regresi ortogonal diperoleh

1. persamaan regresi ortogonal: $\hat{Y} = 0{,}644 + 0{,}995X$
2. nilai intersep mendekati 0 dan *slope* mendekati 1 sehingga kedua alat pemonitor memberikan hasil pengukuran tekanan darah sistolik yang sama.
3. 0 termuat dalam selang kepercayaan bagi intersep dan 1 termuat dalam selang kepercayaan bagi *slope*, sehingga tidak ada bukti yang menyatakan bahwa kedua alat pemonitor memberikan hasil pengukuran yang berbeda.
4. Bahwa nilai *s* pada model regresi ortogonal lebih kecil dari model regresi linear sederhana sehingga model regresi ortogonal merupakan model regresi yang lebih baik daripada model regresi linear sederhana

Pada model regresi linear sederhana diperoleh kesimpulan yang sama dengan model regresi linear sederhana, namun model regresi linear sederhana tidak tepat digunakan karena peubah prediktor memuat kesalahan pengukuran (*measurement error*).



Gambar 4. Plot data tekanan darah sistolik

Dari Gambar 4 dapat dilihat bahwa persamaan regresi ortogonal cocok untuk data tersebut. Titik-titik cukup dekat pada garis regresi ortogonal. Garis regresi *least square* dekat dengan garis regresi ortogonal pada data ini.

KESIMPULAN

Model regresi ortogonal lebih tepat digunakan apabila peubah prediktor memuat kesalahan pengukuran. Apabila rasio dua variansi galat adalah 1 maka nilai $\hat{\beta}_1$ pada model regresi ortogonal akan samadengan nilai $\hat{\beta}_1$ pada model regresi linear sederhana.

DAFTAR PUSTAKA

Bain, L.J. & Engelhardt, M. 1992. *Introduction to probability and mathematical statistics*. 2nd edition. California: Duxbury press.

Carrol, R.J. & Ruppert, D. 1994. The use and misuse of orthogonal regression estimation in linear errors-in-variables models. www.stat.tamu.edu/ftp/pub/rjcarroll/orthogonal.ps [Diakses tanggal 10 April 2011].

Dissanaike, G. & Wang, S. 2003. A critical examination of orthogonal regression. *Social Science Research Network*-id407560. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=407560 [Diakses tanggal 1 April 2011]

Fuller, W.A. 1987. *Measurement error models*. New York: John Wiley & Sons.

Leng, L., Zhang, T., Keinman, L. & Zhu, W. 2007. Ordinary least square regression, orthogonal regression, geometric mean regression and their applications in aerosol science. *Journal of Physics* 78(2007): 1-5.