

MENGATASI HETEROGENITAS VARIANS METODA KUADRAT TERKECIL TERBOBOTI

Oleh

Ani Minarni

(Jurusan Matematika, FMIPA – Universitas Negeri Medan)

ABSTRAK

Jika varians galat diketahui atau dapat diduga dengan baik, maka WLS efektif untuk mengatasi masalah pendugaan parameter β_1, β_2 yang tidak dapat dilakukan oleh OLS akibat tidak dipenuhinya asumsi homogenitas varians. Tetapi jika varians tidak diketahui ataupun tidak dapat diduga maka WLS tidak bisa digunakan, sebagai gantinya lakukan transformasi terhadap variabel independen.

Kata kunci : Galat, heterogenitas, terboboti, varians.

I. PENDAHULUAN

Pandang fungsi regresi linier sederhana (untuk satu variabel independen) berikut:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \varepsilon_{ij}$$

$$(i=1,2, \dots, m ; j=1,2, \dots, n)$$

dengan β_0 adalah *intersep* dan β_1 menyatakan koefisien regresi.

Untuk data yang didapatkan dari sampel maka fungsi regresi dugaannya adalah:

Dengan asumsi galat percobaan menyebar normal dengan rata-rata nol dan varians σ^2 (homogen).

$$\hat{y}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 x_{ij}$$

Parameter β_0 dan β_1 kita duga dari data sampel dan dugaan yang baik adalah yang memiliki sifat BLUE (Best Linier Unbiased Estimator), agar fungsi regresi dugaan yang kita dapatkan sah (valid) untuk digunakan.

Suatu metoda pendugaan parameter regresi linier yang memiliki sifat *blue* adalah Metoda Kuadrat Terkecil Biasa (*Ordinary Least Square Methods*). Metoda ini akan baik digunakan jika galat memiliki rata-rata nol dan varians homogen. Untuk banyak kasus seringkali homogenitas varians tak dipenuhi. Bagaimanakah mengatasi kasus seperti itu?

Model regresi linier sebagai suatu fungsi linier mengasumsikan bahwa galat percobaan dalam fungsi regresi populasi memiliki rata-rata nol dan varians homogen (homoskeditas). Jika asumsi ini tak dipenuhi maka pendugaan parameter melalui metoda kuadrat terkecil biasa (OLS) tak bisa digunakan karena metoda ini sensitif, artinya sifat BLUE menjadi tak sepenuhnya benar. Karena itu, jika asumsi homoskeditas tak dipenuhi maka yang terbaik adalah menggunakan WLS (*Weighted Least Square*) yaitu suatu metoda penaksiran/pendugaan β_0 dan β_1 dengan memanfaatkan pembobot (w_i).

II. TINJAUAN PUSTAKA

A. Pendugaan parameter dengan Metoda Kuadrat Terkecil Biasa (Ordinary Least Square Methods)

Model linier umum, jika ditulis dalam notasi matriks adalah:

$$\underline{y} = \underline{x}\underline{\beta} + \underline{\varepsilon} \dots \dots \dots (1)$$

Penduga kuadrat terkecil biasa untuk persamaan (1) ini adalah:

$$\hat{\underline{\beta}} = (\underline{x}'\underline{x})^{-1} \underline{x}'\underline{y} \dots \dots \dots (2)$$

Persamaan (2) valid jika model memenuhi asumsi-asumsi berikut :

1. $E(\underline{\varepsilon}) = 0$, rata-rata galat sama dengan nol.
2. $Var(\underline{\varepsilon}) = \sigma^2 I_n$, varians galat homogen dengan I_n matriks identitas berukuran n.
3. $\underline{\varepsilon}_i \underline{\varepsilon}_j = 0, \forall i \neq j$.

Jika varians homogen tak dipenuhi, misalnya

$$Var(\underline{\varepsilon}) = \underline{V} \dots \dots \dots (3)$$

dengan

$$\underline{V} = diag(\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2) \dots \dots \dots (4)$$

suatu matriks varians-kovarians yang merupakan pendekatan yang paling umum bagi penduga kuadrat terkecil. Asumsi, galat tak berkorelasi dengan variansnya atau galat bersifat independen. Maka penduga β terbaik adalah penduga kuadrat terkecil umum (Generalized Least Square) yaitu:

$$\beta^* = (x' \underline{v}^{-1} x)^{-1} x' \underline{v}^{-1} y \dots \dots \dots (5)$$

Sifat yang dimiliki penduga ini adalah:

1. Tak bias, yaitu

$$E(\underline{\beta}^*) = \underline{\beta}$$

2. $\underline{\beta}^*$ suatu penduga kemungkinan maksimum dibawah kondisi $\varepsilon_i \sim N(0, \underline{V})$.

3. Penduga Varians minimum

Jika asumsi normalitas tak dipenuhi, maka menurut Gauss-Markoff, $\underline{\beta}^*$ memiliki varians minimum.

Jadi penduga pada persamaan (5) memiliki sifat sebagai BLUE sama dengan penduga kuadrat terkecil biasa. Dan walaupun tergantung pada matriks \underline{V} ; $\underline{\beta}^*$ meminimumkan beberapa jenis metoda kuadrat semisal:

$$JK_{res, \underline{v}} = (y - x\underline{\beta}^*) \underline{v}^{-1} (y - x\underline{\beta}^*)^{-1} \dots \dots \dots (6)$$

B. Pendugaan Parameter dengan Metoda Kuadrat Terkecil Terboboti (Weighted Least Square Methods)

Bila modelnya bersifat ε_i (galat) independen tetapi variansnya heterogen yaitu varians

ke-i adalah $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$ sehingga matriks Varkoffnya adalah $\underline{V} = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2)$ maka $\underline{\beta}^* = (x' \underline{v}^{-1} x)^{-1} x' y$ dengan sifat:

$$JK_{res(\text{weighted})} = \sum w_i (y_i - \hat{y}_i)^2$$

dengan $w_i = 1/\sigma_i^2$.

Hal-hal yang Perlu Mendapat Perhatian

Saat menggunakan WLS (weighted Least Square) harus dimiliki parameter σ_i^2 yang ternyata sulit untuk didapat sebagaimana sulitnya diperoleh parameter lain, misalnya parameter γ pada distribusi (sebaran) Weibull. Tetapi kalau data sampel yang dipunyai adalah hasil pengamatan (observasi) dengan replikasi pada setiap kombinasi variabel independen, maka varians sampel dapat diduga sehingga pembobot yang akan digunakan pada tiap pengamatan bisa diperoleh. Hanya, jika tak ada replikasi, σ_i^2 tak bisa diduga. Hal lain, pendugaan standar deviasi σ_i^2 dengan WLS harus hati-hati karena jika tidak tepat maka sifat-sifat penduga β dari WLS akan lebih buruk daripada yang didapat melalui OLS. Kemudian, pembobot (w_i) akan baik digunakan jika ukuran sampel minimal 9. Jika kurang dari itu sebaiknya WLS tak digunakan.

Matriks V dugaan, berpengaruh kuat pada pengujian hipotesis β , perhitungan C_p (coefficient of prediction) dan PRESS (Prediction Sum of Square), oleh sebab itu V harus diduga dengan benar dan teliti.

PRESS dan C_p diperlukan jika ada beberapa variabel independen (x) yang diperkirakan mempengaruhi y_{ij} , tapi diinginkan hanya x yang signifikan mempengaruhi y_{ij} . Jadi PRESS dan C_p dipakai pada Regresi Multipel untuk menentukan lebih lanjut model regresi yang cocok.

III. PENGUJIAN HIPOTESIS

Pasangan hipotesis yang akan diuji adalah:

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0$$

Statistik uji yang digunakan adalah :

$$T_{hit} = \beta_j^* / \sigma_j^* = \beta_j^* / \sqrt{c_{jj}}$$

dimana c_{jj} = unsur diagonal ke-j matriks $(x' \underline{v}^{-1} x)^{-1}$ dan β_j^* = koefisien regresi terboboti.

Kriteria uji : tolak H_0 jika $t_{hit} > t_{tabel} = t_{(n-2), \alpha}$.

Jika \underline{v} diketahui maka kriteria uji adalah :

Tolak H_0 jika z_{hit} berada diluar daerah $-z_{\alpha/2} < z < z_{\alpha/2}$, α = tingkat signifikansi.

Pada WLS : $JKT_{terboboti} = JKR_{terboboti} + JKG_{terboboti}$.

$$y' \underline{v}^{-1} y = \beta^* x' \underline{v}^{-1} y + (y - x\beta^*)' \underline{v}^{-1} (y - x\beta^*)$$

Jika tak ada asumsi varians diketahui, tetapi diasumsikan $\text{diag } \underline{v}$ diketahui maka

$\text{Var}(\epsilon) = \underline{v}\sigma^2$, dimana σ tidak diketahui, tetapi tidak mengubah pembobot.

Pembobotnya tetap \underline{v}^{-1} , dan $\sigma^2 = JKG_{terboboti}$

$$= s^2 = \sum w_i (y_i - \hat{y}_i)^2 / (n - p)$$

dengan p adalah banyaknya parameter yang akan diduga.

IV. HASIL DAN PEMBAHASAN

A. Data Efisiensi Transfer

Di bawah ini adalah data mengenai efisiensi transfer dari suatu alat penyemprot cat. Ada dua variabel independen yang digunakan untuk merespon efisiensi alat tersebut, yaitu kecepatan udara (X_1) dan voltase (X_2). Empat kombinasi dari faktor-faktor tersebut dipilih dalam sebuah rancangan percobaan dan masing-masing dengan sepuluh replikasi. Kecepatan udara dalam feet per menit dan voltase dalam kilovolt.

Tabel 1. Efisiensi Transfer

		Voltase (X_2)			
		50		70	
Kecepatan Udara (X_1)	60	87,5	88,2	77,4	68,1
		88,1	87,3	70,7	65,3
		89,5	89,2	67	61
		86,2	85,9	71,7	81,7
		90	87	79,2	60,3
	80	82,5	81,3	61,2	50,7
		81,6	80,7	67,2	52,3
		77,4	79,3	55,9	68,6
		81,5	82	52	69,5
		79,7	79,2	63,5	70,1

Tabel 2. Penduga Varians dan Pembobotnya

Voltase	Kecepatan Udara	Penduga Varians (σ_i^2)	Pembobot
50	60	1,8899	0,5291
50	120	2,5018	0,3997
70	60	54,3204	0,0184
70	120	60,6933	0,0165

dimana $\sigma_i^2 = 1/9 \sum (y_i - \bar{y})^2$ untuk setiap kombinasi perlakuan.

Dari penduga varians terlihat bahwa asumsi homogenitas varians tidak terpenuhi. Karena itu, regresi terbobotlah yang digunakan untuk pendugaan parameter dan pengujian hipotesisnya, dimana pembobotnya adalah $1/(\sigma_i^2)$.

B. Pendugaan Parameter β

Matriks Varkov dengan pembobot $w_i = 1/(\sigma_i^2)$ adalah

$$(x'v^{-1}x)^{-1} = \begin{bmatrix} \sum w_i & \sum w_i x_{1i} & \sum w_i z_{1i} \\ \text{simetris} & \sum w_i x_{1i}^2 & \sum w_i x_{1i} x_{2i} \\ & & \sum w_i x_{2i}^2 \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} 9,637329 & 827,9550 & 4888426 \\ & 79643,13 & 42014,09 \\ \text{simetris} & & 24930,58 \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} 19,9777 & -0,0093446 & -0,37597 \\ & 0,00011751 & -0,000014791 \\ \text{simetris} & & 0,00743730 \end{bmatrix}$$

Unsur $\underline{x}'v^{-1}\underline{y}$ adalah:

$$\underline{x}'v^{-1}\underline{y} = \begin{bmatrix} 809,87009 \\ 68507,4018 \\ 40953,7332 \end{bmatrix}$$

$$\beta_0^* = 141,5612,$$

$$S\beta_0^* = 4,4696$$

$$\beta_1^* = -0,1239,$$

$$S\beta_1^* = 0,0108$$

$$\beta_2^* = -0,9242,$$

$$S\beta_2^* = 0,0862$$

Jadi penduga β adalah $\beta^* = \begin{bmatrix} \beta_0^* \\ \beta_1^* \\ \beta_2^* \end{bmatrix}$

$$\hat{y}_{ij} = \underline{x}\beta^*$$

C. Pengujian Hipotesis

$$H_0 : \underline{\beta} = 0$$

$$t = (141,5612)/4,4696 = 31,67 \text{ untuk } H_0 : \beta_0 = 0$$

$$t = (-0,1239)/(0,0108) = -11,47 \text{ untuk } H_0 : \beta_1 = 0$$

$$t = (-0,9242)/(0,0862) = -10,72 \text{ untuk } H_0 : \beta_2 = 0$$

Hasilnya sebagai berikut:

Semua koefisien regresi berbeda secara signifikan dengan nol. Jadi, kecepatan udara (X_1) dan voltase (X_2) memberi pengaruh yang signifikan terhadap efisiensi transfer alat penyemprot cat. Dengan kata lain, voltase, mungkin juga kecepatan udara dapat dioptimumkan untuk memperoleh efisiensi transfer yang terbaik, karena $\hat{y}_i = x\underline{\beta}^*$, untuk $\underline{\beta}^*$ yang diperoleh mengandung arti setiap penurunan kecepatan udara sebesar 0,1239 dan penurunan voltase sebesar 0,9242 satuan akan mengakibatkan transfer alat penyemprot cat semakin efisien

IV. KESIMPULAN

Jika varians galat diketahui atau dapat diduga dengan baik, maka WLS efektif untuk mengatasi masalah pendugaan parameter β_1, β_2 yang tidak dapat dilakukan oleh OLS akibat tidak dipenuhinya asumsi homogenitas varians. Tetapi jika varians tidak diketahui ataupun tidak dapat diduga maka WLS tidak bisa digunakan, sebagai gantinya lakukan transformasi terhadap variabel independen.

DAFTAR PUSTAKA

Myer, R.H. (1990). *Classical and Modern Regression with Applications (2nd ed.)*. Boston: PWS Kent Publishing Company.

Ryan, T. P. (1997). *Modern Regression Methods*. Canada: Wiley Interscience Publication.