DISAIN EKSPERIMEN DENGAN RESPON TAK NORMAL

Oleh :
Arrival Rince Putri

Jurusan Matematika FMIPA Universitas Andalas

Abstrak

Masalah respon tak normal dan variansi tak konstan sering ditemui pada penelitian-penelitian analisis regresi. Untuk mengatasi masalah tersebut biasanya peneliti akan menemukan cara pendekatan klasik yaitu metode kudrat tekecil atau *Ordinary Least Square* (OLS) dimana dilakukan transformasi terhadap data asal untuk memodelkan secara empirik respon yang tak normal. Tetapi kenyataannya transformasi mempunyai banyak kelemahan. Satu metode lain dicadangkan yaitu *Generalized Linear Models* (GLM) yang menggunakan fungsi tertentu untuk merangkaikan penaksiran bagi respon dengan variabel bebas. Penelitian ini menemukan bahwa metode GLM lebih baik daripada metode OLS dalam mengatasi masalah respon tak normal dan variansi tak konstan.

Kata Kunci : respon tak normal, variansi tak konstan, OLS, GLM

I. PENDAHULUAN

Disain eksperimen telah banyak digunakan dalam berbagai bidang penelitian, baik di bidang teknik, sains, proses peningkatan mutu pada industri, maupun dalam bidang penelitian sosial. Penelitian yang dilakukan kebanyakan untuk variabel respon normal di mana galat diasumsikan berdistribusi normal independen dengan rata-rata nol dan variansi konstan.

Tetapi dalam beberapa tahun belakangan ini bidang-bidang penelitian sudah semakin berkembang. Para peneliti semakin banyak menemukan masalah dalam proses analisis penelitiannya terutama dalam menemukan model regresi terbaik. Masalah yang sering dihadapi sekarang adalah masalah respon tak normal. Tidaklah cukup beralasan jika diasumsikan data bilangan atau data diskrit akan berdistribusi normal sementara rata-rata dari data tersebut haruslah terbatas pada selang tertentu. Misalnya, rata-rata dari suatu respon yang berupa data bilangan adalah bernilai tak negatif, tetapi bila menggunakan penduga linear $x_i'\beta$ pada model linear klasik kadang-kadang akan menghasilkan nilai negatif, begitu juga dengan selang keyakinan untuk rata-rata responnya. Jika dari beberapa observasi didapati variansi dari data berubah-ubah terhadap rata-rata maka tentulah tidak sesuai untuk mengasumsikan variansi data konstan untuk observasi tersebut.

II. PERUMUSAN MASALAH

Dalam bidang industri, permasalahan yang sering dihadapi peneliti adalah bagaimana cara untuk menghemat biaya dan waktu dalam proses produksi. Salah satu cara yang biasa dilakukan adalah membuat disain eksperimen terhadap produk yang akan diproduksi, tapi eksperimen yang dibuat tanpa dilakukan pengulangan terhadap observasi. Eksperimen tanpa pengulangan akan menghasilkan galat tak normal dan variansi tak konstan dari data produk yang sedang diobservasi (Lewis, Montgomery dan Myers, 2001). Produk yang seperti ini sebaiknya tidak dilanjutkan dalam proses produksi karena tidak akan menghasilkan produk yang berkualitas baik.

Untuk mengatasi masalah tersebut biasanya peneliti akan menggunakan cara pendekatan klasik, yaitu dengan cara transformasi terhadap data asal (sampel

produk) untuk memodelkan secara empirik respon yang tak normal. Transformasi yang digunakan adalah metode kuadrat terkecil atau *Ordinary Least Square* (OLS). Transformasi dilakukan supaya asumsi kenormalan, variansi konstan dan model regresi yang terbaik dapat ditentukan. Tetapi pada kenyataannya setelah dilakukan transformasi, ketiga tujuan dari transformasi tersebut belum terjamin akan terwujud. Ditambah lagi dugaan respon yang diperoleh dari transformasi perlu diubah kembali kepada bentuk asal yang kebiasaannya tidak memberikan dugaan rata-rata yang teliti. Untuk itu metode OLS tidakl cukup baik dalam menghadapi masalah respon tak normal.

Satu metode lain dicadangkan yaitu menggunakan *Generalized Linear Models* (GLM) dimana tidak perlu melakukan transformasi terhadap data tetapi menggunakan fungsi tertentu yang merangkaikan penaksiran bagi variabel respon dengan variabel bebas.

Penelitian ini dilakukan untuk disain eksperimen yang mempunyai k faktor dimana setiap faktor terdiri dari dua arah. Disain eksperimen seperti ini disebut juga dengan disain faktorial 2^k. Sebagai contoh disain faktorial 2^2, berarti pada eksperimen ini melibatkan dua buah faktor, misal A dan B, dan setiap faktor terdiri dari dua arah, misal tinggi dan rendah. Sehingga pada eksperimen ini mempunyai 4 kombinasi perlakuan dan menghasilkan 4 buah respon. Pada penelitian ini dibatasi untuk disain faktorial 2^3.

Selang keyakinan terhadap metode ini ditentukan sebagai kriteria dalam menentukan metode terbaik. Selang keyakinan yang lebih pendek menghasilkan dugaan yang lebih baik. Data kasus diperlukan untuk membandingkan dan menentukan metode yang terbaik.

III. TINJAUAN PUSTAKA

a. *Generalized Linear Models* (GLM)

GLM terdiri dari tiga komponen utama, yaitu :

1. Variabel respon $Y_1, Y_2, \ldots, Y_n$ yang independen satu sama lain dan termasuk satu satu dari distribusi kelompok eksponen. Dalam hal ini variansi respon tergantung pada rata-rata μ_i dan mempunyai bentuk :

$$Var(Y_i) = \phi V(\mu_i) \qquad (1)$$

ϕ kadangkala dinyatakan dalam σ^2 yang dikenal sebagai parameter dispersi (bagian variansi yang tidak regantung pada rata-rata). $V(\mu_i)$ merupakan fungsi variansi. Parameter dispersi ϕ nilainya diketahui atau diduga, dalam distribusi Poisson $\phi = 1$.

2. Satu set parameter β_j dan variabel independen x_j, untuk $j = 1,2,\ldots,p$ yang dihubungkan oleh sebuah penduga linear η, dimana $\eta = \sum_{j=1}^{p} x_j \beta_j$
3. Fungsi penghubung g yang menghubungkan rata-rata dari Y, yaitu μ_i dengan x_i dalam bentuk hubungan : $g(\mu_i) = x_i'\beta$

Dalam GLM, variabel respon diasumsikan akan mempunyai distribusi dari kelompok eksponen. Oleh karena itu fungsi kepekatan peluang untuk variabel respon y kontinu atau diskret adalah :

$$f(y;\theta,\phi) = eksp(\frac{y\theta - b(\theta)}{a(\theta)} + c(y,\phi) \tag{2}$$

Untuk a, b dan c ditentukan berdasarkan distribusi tertentu. Jika ϕ diketahui, maka persamaan di atas merupakan model kelompok eksponen dengan parameter kanonik θ. Contoh fungsi untuk a dan c, yaitu : $a(\phi) = \frac{\phi}{w}$, $c(y, \frac{\phi}{w})$

dengan w adalah pembobot yang diketahui dan boleh berbeda untuk setiap observasi, sedangkan ϕ kadangkala dinyatakan dalam σ^2 yang dikenal sebagai parameter dispersi.

Log fungsi likelihood untuk persamaan (2) di atas adalah :

$$l(\theta,\phi;y) = \log f_y(y;\theta,\phi) \tag{3}$$

Dengan mensubstitusi persamaan (2) ke dalam persamaan (3), diperoleh :

$$l(\theta,\phi;y) = \frac{y\theta - b(\theta)}{a(\theta)} + c(y,\phi) \tag{4}$$

Turunan pertama dan kedua terhadap θ menghasilkan :

$$\frac{\partial l}{\partial \theta} = \frac{y - b'(\theta)}{a(\phi)} \tag{5}$$

$$\frac{\partial^2 l}{\partial \theta^2} = \frac{-b''(\theta)}{a(\phi)} \tag{6}$$

Untuk mendapatkan rata-rata dan variansi y digunakan dua bentuk hubungan berikut :

$$E(\frac{\partial l}{\partial \theta}) = 0 \tag{7}$$

$$E(\frac{\partial^2 l}{\partial \theta}) + E(\frac{\partial l}{\partial \theta})^2 = 0 \tag{8}$$

Dengan menggantikan persamaan (5) ke dalam persamaan (7), diperoleh :

$$E(\frac{\partial l}{\partial \theta}) = E(\frac{y - b'(\theta)}{a(\phi)}) = 0$$

$$\frac{E(y) - b'(\theta)}{a(\phi)} = 0$$

$$E(y) = b'(\theta)$$

Kemudian gantikan persamaan (6) dan (5) ke dalam persamaan (8) :

$$E(\frac{-b''(\theta)}{a(\phi)}) + E(\frac{y - b'(\theta)}{a(\phi)})^2 = 0$$

$$\frac{-b''(\theta)}{a(\phi)} + \frac{Var(y)}{a^2(\phi)} = 0$$

$$Var(y) = b''(\theta)a(\phi) = \frac{b''(\theta)\phi}{w} \quad ; \quad a(\phi) = \frac{\phi}{w}$$

Dapat dilihat bahwa variansi y merupakan perkalian dua buah fungsi. $b''(\theta)$ bergantung kepada parameter kanonik (merupakan fungsi untuk rata-rata) yang dikenali juga sebagai fungsi variansi, sedangkan $a(\phi)$ hanya bergantung pada ¢. Jika μ adalah rata-rata untuk y dan variansi merupakan fungsi dari rata-rata, maka :

$$Var(y) = \frac{V(\mu)\theta}{w} \quad ; \quad b''(\theta) = V(\mu)$$

Distribusi dugaan untuk variabel respon y dalam GLM ini biasanya dinyatakan dalam parameter rata-rata μ dan parameter dispersi ¢ atau parameter kanonik θ.

b. Fungsi Penghubung dan Penduga Linear

Rata-rata μ_i untuk respon data pengamatan ke-i dihubungkan dengan penduga linear oleh sebuah fungsi penghubung g yang bersifat monotonik dan mempunyai turunan, yaitu : $g(\mu_i) = x_i'\beta$, dimana x_i berupa vektor variabel bebas yang nilainya ditetapkan, sedangkan β adalah parameter yang akan diduga nilainya.

Adapun bentuk penghubung kanonik dan model untuk distribusi Poisson adalah (Myers & Mongomery, 1997 dan Hamada & Nelder, 1997) :

Parameter kanonik : $\theta = \ln \mu$

Fungsi penghubung : $\ln \mu = x_i'\beta$

Model : $\mu = e^{x_i'\beta}$

GLM menggunakan model penduga linear $\eta = x_i'\beta$. Untuk mengetahui rumusan model dari penduga linear dan jenis kovariatnya dapat dilihat dalam McCullagh dan Nelder (1989).

IV. TUJUAN PENELITIAN

Tujuan dari penelitian ini adalah :

1. Menguraikan penggunaan OLS dan GLAM dalam menangani masalah respon tak normal dan variansi tak konstan.
2. Menerapkan penggunaan metode OLS dan GLM pada disain faktorial 2^5, dan membandingkan kedua metode tersebut.

V. KONTRIBUSI PENELITIAN

Penelitian ini diharapkan dapat bermanfaat bagi industri sebagai pihak yang sering menghadapi masalah dalam membuat disain eksperimen yang dapat menghemat biaya dan waktu dalam proses produksi untuk menghasilkan produk yang bermutu.

VI. METODE PENELITIAN

a. Metode Penaksiran Parameter

Misalkan variabel tak bebas Y_i, untuk I = 1,2,....n mempunyai ketiga komponen utama dalam GLM. Akan ditentukan penaksir untuk parameter β dengan metode penaksir kemungkinan maksimum.

Berikut ni contoh sederhana yang menggunakan dua buah parameter β. Misalkan respon Y_i adalah variabel randim berdistribusi Poisson, sehingga variansi bagi Y_i merupakan fungsi dari rata-rata Y_i :

$$Var(Y_i) = E(Y_i)$$

dengan hubungan antara Y_i dengan x_i adalah berupa hubungan linear :

$$E(Y_i) = \mu_i = \beta_1 + \beta_2 x_i$$

dimana :

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \text{ dan } x_i' = \begin{bmatrix} 1 \\ x_i \end{bmatrix} \text{ untuk i = 1,2,...,n}$$

Hubungan Y_i dengan x_i di atas bisa juga ditulis dengan menggunakan fungsi penghubung $g(\mu_i) = \mu_i = x_i'\beta = \eta_i$ sehingga $\partial\mu_i / \partial\eta_i = 1$. Kemudian pembobot w_{ii} digunakan untuk menghitung variansi Y_i yang berubah terhadap rata-ratanya, dimana w_{ii} sebagai berikut :

$$w_{ii} = \frac{1}{Var(Y_i)} = \frac{1}{\beta_1 + \beta_2 x_i}$$

Jika digunakan $b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$ sebagai penaksir bagi β dan dibuat transformasi untuk variabel respon $z = \hat{\eta} + (y - \hat{\mu})(\frac{d\eta}{d\mu})$, sehingga diperoleh :

$$z_i = b_1 + b_2 x_i + (y_i - b_1 - b_2 x_i) = y_i$$

Dikenali bentuk hubungan $X'WX$ sebagai matriks informasi I :

$$I = \begin{bmatrix} \sum_{i=1}^{n} \frac{1}{b_1 + b_2 x_i} & \sum_{i=1}^{n} \frac{x_i}{b_1 + b_2 x_i} \\ \sum_{i=1}^{n} \frac{x_i}{b_1 + b_2 x_i} & \sum_{i=1}^{n} \frac{x_i^2}{b_1 + b_2 x_i} \end{bmatrix}$$

dan

$$X'Wz = \begin{bmatrix} \sum_{i=1}^{n} \frac{y_i}{b_1 + b_2 x_i} \\ \sum_{i=1}^{n} \frac{x_i y_i}{b_1 + b_2 x_i} \end{bmatrix}$$

dimana W merupakan matrik diagonal berukuran n x n dengan unsur-unsurnya adalah :

$$w_{ii} = \frac{1}{\mathrm{var}(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2$$

Penaksir kemungkinan maksimum bagi β diperoleh secara iterasi dari persamaan berikut :

$$(X'WX)^{(m-1)} b^{(m)} = X'Wz^{(m-1)}$$

dengan (m-1) menyatakan iterasi pada titik $b^{(m-1)}$. Prosedur-prosedur di atas dapat diperluas untuk beberapa kovariat x_i dan beberapa β.

b. Data Yang Digunakan

Data yang digunakan pada penelitian ini adalah data sekunder yang diperlihatkan pada Tabel 6.1.

Tabel 6.1 Data Pengamatan

No	X_1	X_2	X_3	Y
1	-	-	-	40
2	+	-	-	48
3	-	+	-	54
4	+	+	-	56
5	-	-	+	44
6	+	-	+	84
7	-	+	+	42
8	+	+	+	92

Sumber : John Peter W. M. 1990. Statistical Methods In Engineering And Quality Assurance. USA : John Wiley & Sons.

dimana : X_1 : suhu

X_2 : tekanan

X_3 : waktu

(-) : rendah

(+) : tinggi

VII. JADWAL PELAKSANAAN

Pelaksanaan penelitian dilaksanakan selama 5 bulan, bersama dengan anggota peneliti yang lain.

VIII. PERSONALIA PENELITI

1. Arrival Rince Putri, S.Si, MT, sebagai peneliti utama
2. Ir. Yudiantri Asdi, MSc, sebagai pembimbing
3. Sinta Asna Mery, sebagai anggota

IX. HASIL PENELTIAN

a. Pendugaan Selang Keyakinan

Untuk menghitung selang keyakinan pada GLM, metode delta digunakan untuk menghitung variansi. $Var(\hat{\mu}(x_i'\beta))$ dimana $\hat{\mu}(x_i'\beta)$ adalah nilai ramalan. Myers dan Montgomery (1997) membuktikan bahwa dengan menggunakan metode delta akan dihasilkan variansi bagi $\hat{\mu}(x_i'\beta)$ pada x_o seperti berikut :

$$Var[\hat{\mu}(x_o'\beta)] = d_o'[Var(\hat{\beta})]d_o \tag{9}$$

dengan

$$d_o = \frac{\partial \mu(x_o'\hat{\beta})}{\partial \beta}, \quad \text{dan} \quad d_o' = Var(y_o)x_o' \tag{10}$$

Menurut McCullagh & Nelder (1989) dan Myers (1990) :

$$Var(\hat{\beta}) = \frac{(X'VX)^{-1}}{[r(\phi)]^2} \tag{11}$$

Gantikan persamaan (10) dan (11) ke dalam persamaan (9), sehingga diperoleh bentuk berikut :

$$Var[\hat{\mu}(x_o'\beta)] = [Var(y_o)]^2 x_o' \frac{(X'VX)^{-1}}{[r(\phi)]^2} x_o \tag{12}$$

Secara asimtot $\dfrac{\hat{\mu}(x_o'\beta) - \mu(x_o'\beta)}{Var(y_o)\sqrt{x_o' \dfrac{(X'VX)^{-1}}{[r(\phi)]^2} x_o}} \sim N(0,1)$

Maka selang keyakinan 100 (1-α)% untuk rata-rata $\mu(x_o'\beta)$ adalah :

$$\mu(x_o'\beta) \pm z_{\alpha/2}\, Var(y_o) \sqrt{x_o' \frac{(X'VX)^{-1}}{[r(\phi)]^2} x_o} \qquad (13)$$

Untuk kasus Poisson diperoleh $\mu(x_i'\beta) = e^{x_i'\beta}$, sehingga selang keyakinan 100 (1-α)% untuk rata-rata $\mu(x_o'\beta)$ adalah :

$$e^{x_i'\beta} \pm z_{\alpha/2} e^{x_i'\beta} \sqrt{x_i'(X'VX)^{-1} x_i} \qquad (14)$$

b. Analisis Data Contoh

Metode OLS dan GLM kemudian diaplikasikan kepada data contoh di atas. Data diolah dengan menggunakan *software* S-Plus 2000. Akan ditentukan metode mana yang lebih baik dalam mengatasi masalah respon tak normal, yaitu dengan melihat panjang selang keyakinan yang dihasilkan dari kedua metode tersebut. Metode yang menghasilkan selang keyakinan yang terpendek adalah metode yang lebih baik dalam mengatasi masalah dalam penelitian.

Adapun hasil pengolahan data diperlihatkan pada tabel di bawah ini.

Tabel 9.1 Selang Keyakinan 95% OLS dan GLM

OLS		GLM		Panjang Selang Keyakinan	
Y	Selang Keyakinan 95%	Y	Selang Keyakinan 95%	OLS	GLM
46.7	(35.63 , 59.35)	47	(37.49 , 56.50)	32.70	19.01
57.9	(40.17 , 65.18)	52	(42.01 , 61.99)	25.01	19.98
46.7	(35.63 , 59.35)	47	(37.49 , 56.50)	23.70	19.01
57.9	(40.17 , 65.18)	52	(42.13 , 62.87)	25.01	19.98
42.9	(32.37 , 55.12)	43	(33.12 , 52.88)	22.75	18.18
87.9	(72.44 , 104.98)	88	(74.99 ,101.14)	32.54	26.70
42.9	(32.37 , 55.12)	43	(33.12 , 50.09)	22.75	18.18
87.9	(72.44 , 104.98)	88	(73.86 ,102.14)	32.54	26.01

Dari tabel di atas diperoleh bahwa semua panjang selang keyakinan 95% untuk metode GLM lebih pendek daripada metode OLS.

X. KESIMPULAN

Pada penelitian ini digunakan metode OLS dan GLM untuk mengatasi masalah respon tak normal dan variansi tak konstan untuk data yang berdistribusi Poisson. Kedua metode tersebut kemudian diaplikasikan pada data contoh dimana disain eksperimennya berbentuk faktorial 2^3. Kedua metode tersebut selanjutnya dibandingkan dengan melihat panjang selang keyakinan yang dihasilkan. Berdasarkan hasil pengolahan data yang menggunakan program S-Plus 2000, diperoleh bahwa metode GLM menghasilkan panjang selang keyakinan yang lebih pendek daripada metode OLS. Sehingga dapat disimpulkan bahwa metode GLM lebih baik daripada metode OLS dalam mengatasi masalah respon tak normal dan variansi tak konstan.

XI. UCAPAN TERIMA KASIH

Penelitian ini terlaksana atas bantuan dana dari Dana Rutin Universitas Andalas Tahun 2006. Ucapan terima kasih disampaikan kepada Lembaga Penelitian Universitas Andalas yang telah memberikan bantuan dana tersebut kepada peneliti. Ucapan terima kasih juga disampaikan kepada berbagai pihak yang tidak dapat disebutkan satu-persatu namanya dalam laporan ini yang telah membantu proses penelitian mulai dari pengumpulan data hingga penulisan laporan.

XII. DAFTAR PUSTAKA

Hamada, M. & Nelder, J.A. 1997. Generalized Linear Models For Quality Improvement Experiments. *Journal of Quality Technology*, **29** (3): 292-304.

John Peter W. M. 1990. *Statistical Methods In Engineering And Quality Assurance*. USA: John Wiley & Sons, Inc.

Lewis, S.L., Montgomery, D.C. & Myers, R.H. 2001. Examples of Designed Experiments With Nonnormal Responses. *Journal of Quality Technology*, **33** (3): 265-278.

Lewis, S.L., Montgomery, D.C. & Myers, R.H. 2001. Confidence Interval Coverage For Designed Experiments Analyzed With GLMs. *Journal of Quality Technology*, **33** (3): 279-292.

McCullagh, P. & Nelder, J.A. 1989. *Generalized Linear Models*. Ed ke-2. London: Chapman & Hall.

Myers, R.H. & Montgomery, D.C. 1997. A Tutorial On Generalized Linear Models. *Journal of Quality Technology*, 29 (3): 274-291.