REGRESI KUANTIL
(STUDI KASUS PADA DATA SUHU HARIAN)

Rita Rahmawati[1], Widiarti[2], Pepi Novianti[3]

[1)]Program Studi Statistika FMIPA Undip

[2)]Sekolah Pascasarjana Departemen Statistika IPB

[3)]Jurusan Matematika Universitas Negeri Bengkulu

Abstrak

Regresi merupakan teknik statistika untuk menentukan persamaan garis atau kurva dengan meminimumkan penyimpangan antara data pengamatan dan nilai-nilai dugaannya. Namun regresi dengan *Ordinary Least Square* (OLS) dianggap kurang tepat untuk menganalisis sejumlah data yang tidak simetris, karena nilai *mean* sebagai penduga bagi nilai tengah data menjadi sangat peka dengan adanya data *outlier*. Kemudian berkembanglah Median Regression dengan pendekatan LAD (*Least Absolute Deviation*) yang dikembangkan dengan mengganti pendekatan *mean* pada OLS menjadi median. Masalah selanjutnya adalah apabila terdapat kemungkinan bahwa kemiringan data bukan terletak pada mediannya melainkan pada potongan kuantil tertentu. Pendekatan dengan median dirasa kurang karena hanya melihat dua kelompok data yang dibagi pada nilai tengahnya saja. Sehingga berkembanglah metode Regresi Kuantil (*Quantile Regression*). Metode ini merupakan salah satu metode regresi dengan pendekatan memisahkan atau membagi data menjadi kuantil-kuantil tertentu dimana dicurigai terdapat perbedaan nilai dugaan. Untuk studi kasus pada makalah ini digunakan data suhu harian (hari ini dan kemarin) yang dalam diagram pencar terlihat tidak simetris dan dicurigai terjadi heteroskedastisitas.

Kata kunci : OLS, LAD, Regresi Kuantil, Heteroskedastisitas

1. Pendahuluan

Regresi merupakan suatu teknik dalam statistika untuk menentukan suatu persamaan garis atau kurva dengan cara meminimumkan penyimpangan atau deviasi antara data pengamatan dan nilai-nilai dugaannya. Regresi digunakan untuk menduga suatu variabel respon dari variabel (peubah) yang sudah diketahui atau diasumsikan ada hubungan dengan variabel respon. Pendekatan standar untuk mendapatkan nilai dugaan parameter dari model regresi linier adalah Metode OLS (*Ordinary Least Square*). Nilai dugaan bagi parameter dengan menggunakan metode OLS diperoleh dengan meminimumkan jumlah kuadrat sisaan. Namun OLS dianggap kurang tepat untuk menganalisis sejumlah data yang tidak simetris. Pertimbangannya adalah apabila data berbentuk lonceng tidak simetris, maka nilai *mean* menjadi sangat peka dengan adanya data *outlier*. Akibatnya, *mean* menjadi kurang tepat digunakan sebagai penduga bagi nilai tengah data.

Kemudian berkembanglah Median Regression dengan pendekatan LAD (*Least Absolute Deviation*) yang dikembangkan dengan mengganti pendekatan *mean* pada OLS menjadi median. Nilai dugaan bagi parameter dengan metode ini diperoleh dengan meminimumkan jumlah nilai mutlak dari sisaan. Sehingga penduga parameter mengarah pada nilai median data.

Permasalahan selanjutnya adalah apabila terdapat kemungkinan bahwa kemiringan data bukan terletak pada mediannya melainkan pada potongan quantile tertentu. Pendekatan dengan median dirasa kurang karena hanya melihat dua kelompok data yang dibagi pada nilai tengahnya saja. Sehingga berkembanglah metode Regresi Kuantil (*Quantile Regression*). Metode ini merupakan salah satu metode regresi dengan pendekatan memisahkan atau membagi data menjadi kuantil-kuantil tertentu dimana dicurigai terdapat perbedaan nilai dugaan.

2. Regresi Kuantil

Regresi Kuantil merupakan suatu pendekatan dalam analisis regresi yang dikenalkan oleh Koenker dan Bassett (1978). Pendekatan ini menduga berbagai fungsi kuantil dari suatu distribusi Y sebagai fungsi dari X. Regresi Kuantil sangat berguna jika distribusi data tidak homogen (*heterogenous*) dan tidak berbentuk standar seperti tidak simetris, terdapat ekor pada sebaran, atau *truncated distribution*.

Misalkan Y adalah peubah acak dengan fungsi distribusi F_Y dan θ adalah konstanta dimana $0 < \theta < 1$. Kuantil ke-θ dari F_Y, dinotasikan sebagai $q_Y(\theta)$ adalah solusi untuk $F_Y(q) = \theta$, yaitu :

$$q_y(\theta) := F_y^{-1}(\theta) = inf\{y : F_Y(y) \geq \theta\}$$

Sehingga $100\theta\%$ $(100(1-\theta)\%)$ dari masa peluang Y berada di bawah (di atas) $q_Y(\theta)$.

Seperti halnya dengan metode OLS yang meminimumkan jumlah kuadrat sisaan untuk mencari nilai dugaan bagi β, maka dalam regresi kuantil, kuantil ke-θ dari F_Y dapat diperoleh dengan meminimumkan fungsi berikut ini terhadap q:

$$\theta \int_{y>q} |y-q| dF_Y(y) + (1-\theta) \int_{y<q} |y-q| dF_Y(y)$$
$$= \theta \int_{y>q} |y-q| dF_Y(y) - (1-\theta) \int_{y<q} |y-q| dF_Y(y)$$

Dengan meminimumkan fungsi di atas, akan diperoleh persamaan berikut:

$$0 = -\theta \int_{y>q} dF_Y(y) + (1-\theta) \int_{y<q} dF_Y(y)$$
$$= -\theta[1 - F_Y(q)] + (1-\theta) F_Y(q)$$
$$= -\theta + F_Y(q)$$

Sehingga quantil ke θ merupakan solusi dari F_Y .

Jika Y sebagai fungsi dari X yang telah diketahui, memiliki peluang $F_{Y|X}(y)$, kuantil ke-θ dari fungsi tersebut dapat dituliskan sebagai $Q_{Y|X}(\theta) := F_{Y|X}^{-1}(\theta)$. $Q_{Y|X}(\theta)$ merupakan fungsi dari X dan diselesaikan dengan persamaan berikut:

$$\min_q \theta \int_{y>q} |y-q| dF_Y(y) + (1-\theta) \int_{y<q} |y-q| dF_Y(y) \qquad (1)$$

$Q_{Y|X}(0.5)$ adalah median Y (sebagai fungsi dari X) yang menunjukkan titik simetri dari $F_{Y|X}$; untuk θ mendekati 0 (atau 1), $Q_{Y|X}(\theta)$ menunjukkan ekor kiri (atau kanan) dari $F_{Y|X}$.

Dalam notasi matriks, jika $Q_{Y|X}(\theta)$ adalah fungsi linear **X'β**, maka persamaan 1 akan menjadi:

$$\min_q \theta \int_{y>X'\beta} |y-X'\beta| dF_Y(y) + (1-\theta) \int_{y<q} |y-X'\beta| dF_Y(y) \qquad (2)$$

Solusi dari persamaan 2 ini dinotasikan sebagai β_θ dan kuantil Y (sebagai fungsi dari X) ke-θ adalah $Q_{Y|X}(\theta) = X'\beta_\theta$

Misalnya diberikan data $(\mathbf{y_t}, \mathbf{x_t}')'$ untuk t=1, 2, …, T, dimana $\mathbf{x_t}$ berukuran kx1, maka model linier dari persamaan Regresi Kuantil dapat dituliskan sebagai:

$$y_t = \boldsymbol{x}_t'\boldsymbol{\beta} + e_t \tag{3}$$

dengan $Q_\theta(y_t|x_t) = \boldsymbol{x}_t'\boldsymbol{\beta}$ merupakan kuantil ke-θ (0 < θ < 1) dari y dengan suatu nilai x_t tertentu. Penduga bagi β dari Regresi Kuantil ke-θ diperoleh dengan meminimumkan jumlah nilai mutlak dari error dengan pembobot θ untuk error positif dan pembobot $(1-\theta)$ untuk error negatif yaitu:

$$\hat{\beta} = \min_\beta \left\{\theta \sum_{t:y_t \geq x_t} |y_t - x_t'\beta| + (1-\theta) \sum_{t:y_t < x_t} |y_t - x_t'\beta|\right\} \tag{4}$$

atau

$$\hat{\beta} = \min_\beta \frac{1}{T} \sum_{i=1}^{T} \rho_\theta u_t \tag{5}$$

dimana $\rho_\theta(u_t) = \begin{cases} \theta u_t & ,jika\ u_t \geq 0 \\ (\theta-1)u_t & ,jika\ u_t < 0 \end{cases}$

$\rho_\theta(u_t)$ disebut juga sebagai *Check Function* dan error dugaan dari y adalah $\hat{e}_t = y_t - \boldsymbol{x}_t'\boldsymbol{\beta}$. Solusi dari persamaan 4 tidak dapat diperoleh secara analitik, melainkan secara numerik seperti dengan metode *simplex*, metode *interior point* atau metode *smoothing*.

Selang Kepercayaan bagi β

Dalam analisis regresi quantil terdapat 3 metode untuk menghitung selang kepercayaan bagi β, yaitu:

- Metode *sparsity*
- Metode Rank
- Metode Resampling

Metode *sparsity* digunakan bila e_i diasumsikan i.i.d. dan berdistribusi F. Untuk data yang lebih dari 5000 pengamatan atau lebih dari 20 variabel , metode yang digunakan adalah metode MCMB (*markov chain marginal bootstrap*) Resampling.

Pengujian hipotesis

Misalkan model regresi kuantil sebagai berikut:

$$y_t = x_t' \beta + e_t$$

Pengujian hipotesisnya adalah:

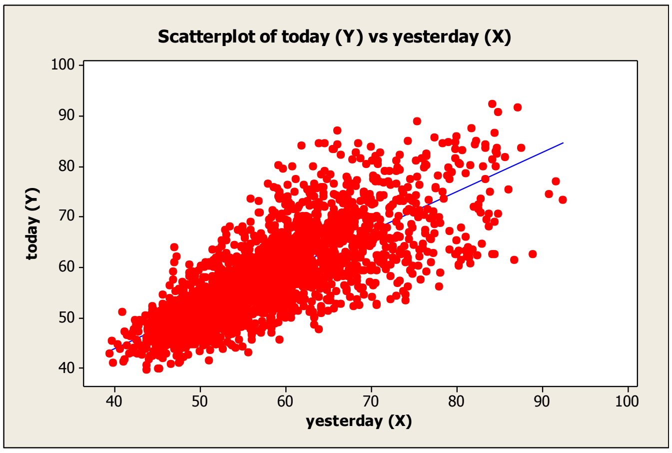
$H_0: \beta = 0 \qquad vs \qquad H_1: \beta \neq 0$

Statistic uji yang digunakan adalah statistic wald:

$$T_W(\tau) = \hat{\beta}_2'(\tau) \sum(\tau)^{-1} \hat{\beta}_2(\tau)$$

3. Studi Kasus

Studi kasus untuk aplikasi regresi kuantil ini diambil dari data temperatur suhu harian di Kota Sydney, dimana akan dilihat apakah suhu kemarin (X) berpengaruh terhadap suhu hari ini (Y). Scatter plot antara suhu hari ini dengan suhu kemarin dapat dilihat pada gambar berikut:



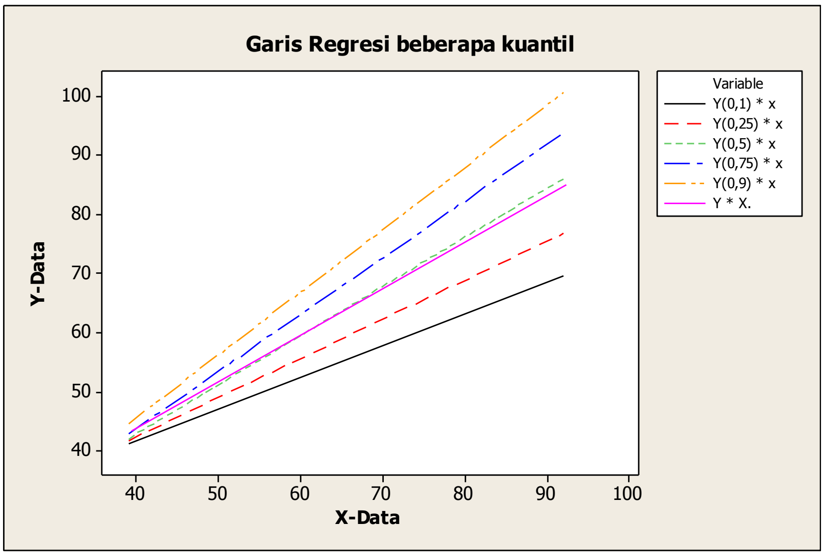
Gambar 1. Plot suhu hari ini terhadap suhu kemarin.

Dari scatter plot terlihat bahwa suhu harian bersifat tak simetris dan dicurigai tejadi heterokedastisitas. Dengan kondisi ini, apabila analisis regresi dilakukan dengan menggunakan metode OLS, maka akan mengakibatkan varians dari penduga parameter besar. Sehingga penduga parameternya menjadi tidak efisien. Untuk mengatasi hal tersebut dilakukan analisis dengan metode regresi kuantil. Analisis data dilakukan dengan menggunakan *software* SAS dengan listing program terlampir. Berikut ini beberapa nilai dugaan parameter pada beberapa nilai quantil, serta nilai dugaan parameter dengan menggunakan Metode OLS:

Tabel 1. Penduga Parameter

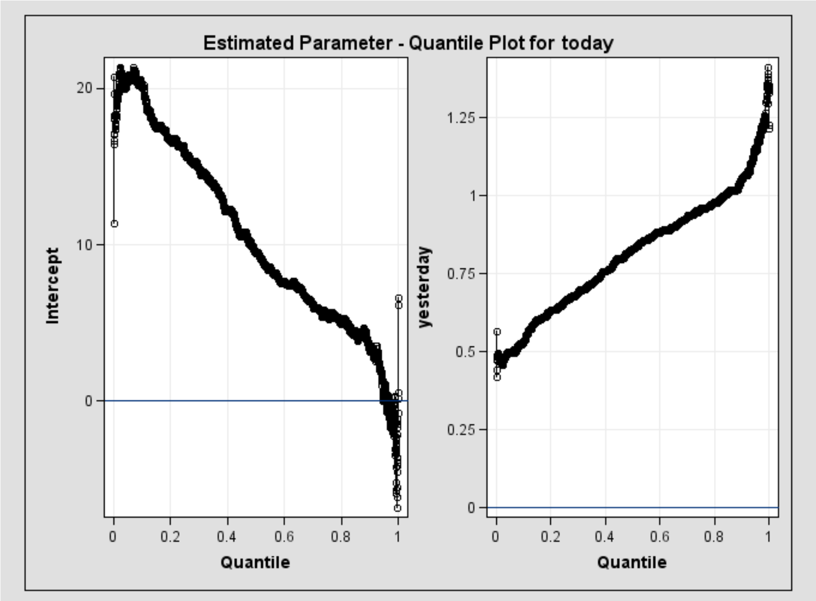
Kuantil	10%	25%	50%	75%	90%	OLS
Intersep	20.0759	15,8705	9,5127	5,3823	3,2631	12,4
Slope	0,5370	0,6615	0,8318	0,9597	1,0578	0,785

Pendugaan parameter dengan menggunakan metode OLS hanya menghasilkan satu pendugaan intersep dan slope, sehingga hanya menghasilkan satu persamaan linier untuk menduga seluruh nilai Y berdasarkan nilai X, dalam hal ini suhu hari ini dan suhu kemarin. Sementara itu dalam pendugaan parameter dengan menggunakan metode regresi kuantil menghasilkan pendugaan parameter yang berbeda-beda sesuai dengan nilai kuantil dari Y, sehingga nilai Y diduga berdasarkan nilai kuantilnya. Persamaan garis regresi Y terhadap X pada kuantil 10%, 25%, 50%, 75% dan 90% dapat dilihat pada gambar berikut:



Gambar 2. Beberapa garis regresi kuantil 10%, 25%, 50%, 75%, 90%

Secara grafis, pendugaan parameter intersep dan slope ditampilkan dalam output SAS berikut ini:



Gambar 3. Nilai dugaan parameter regresi kuantil

Nilai dugaan bagi parameter β_0 semakin mengecil dengan semakin besarnya kuantil yang dipilih, sebaliknya nilai dugaan bagi β_1 semakin besar dengan semakin besarnya kuantil yang dipilih. Sebagai contoh: persamaan regresi untuk kuantil 10%,25%, 50%, 75% dan 90%., yaitu:

$$Y_{10\%} = 20{,}0759 + 0{,}5370X$$

$$Y_{25\%} = 15{,}8705 + 0{,}6615X$$

$$Y_{50\%} = 9{,}5127 + 0{,}8318X$$

$$Y_{75\%} = 5{,}3823 + 0{,}9597X$$

$$Y_{90\%} = 3{,}2631 + 1{,}0578X$$

Selang kepercayaan bagi penduga parameter β_0 dan β_1 untuk 5 kuantil di atas adalah:

Tabel 2. Penduga Parameter

Kuantil	Intersep	Yesterday
10%	17,2947 – 21,1749	0,5160 – 0,5777
25%	13,6182 – 18,2770	0,6220 – 0,6990
50%	7,2736 – 11,7316	0.7861 – 0,7861
75%	3,7613 – 7,8850	0.9040 – 0,9960
90%	0,9181 – 7,3101	0.9853 – 1,0986

Dari table di atas dapat disimpulkan bahwa dengan tingkat kepercayaan 95% secara rata-rata kenaikan suhu hari ini pada Kuantil 10% terletak antara 0,5160 – 0,5777 dan

0,6220 – 0,6990 pada kuantil 25%. Sedangkan pada kuantil 50%, 75% dan 90% masing-masing adalah 0.7861 – 0,7861, 0.9040 – 0,9960 dan 0.9853 – 1,0986.

Untuk pengujian apakah β_1 bersifat signifikan digunakan uji wald. Berikut nilai statistic uji wald untuk beberapa kuantil:

Tabel 3. Nilai Statistic Uji Wald

Kuantil	Statistik Uji Wald	Chi Square	p-value
10%	613,2052	613,21	<0.0001
25%	1611,5592	1611,56	<0.0001
50%	2892,5423	2892,54	<0.0001
75%	3112,3260	3112,33	<0.0001
90%	1238,6863	1238,69	<0.0001

Berdasarkan uji wald, dapat disimpulkan bahwa pada kuantil 10%, 25%, 50%, 75%, dan 90% suhu kemarin berpengaruh terhadap suhu hari ini.

4. Kesimpulan

Metode Regresi Kuantil yang dikenalkan oleh Koenker & Bassett merupakan cara pendugaan dengan menggunakan pendekatan fungsi kuantil dari suatu distribusi Y sebagai fungsi dari peubah penjelas X. Analisis Regresi Kuantil sangat berguna untuk data yang tidak simetris, tidak homogen (heteroskedaktisitas) ataupun tidak beraturan. Penyelesaian pendugaan parameter tidak dapat dilakukan secara analitik, tetapi diselesaikan secara empirik. Nilai dugaan parameter dengan regresi kuantil memberikan hasil yang berbeda dengan analisis regresi yang menggunakan metode OLS.

Daftar Pustaka

Koenker, R and G. Basset (1978). *Regression Quantile Econometrica*, 46, 33-50.

Cung-Ming Kuan (2007), *An Introduction to Quantile Regression*, Institute of Economics, Academia Sinica