

ESTIMASI TITIK UBAH TUNGGAL PADA REGRESI LINIER DENGAN SATU PEUBAH BEBAS

Muhammad Bayu Nirwana¹⁾, Dewi Wulandari²⁾

¹⁾Ilmu Keperawatan, STIKES Muhammadiyah Kudus

email: mbnirwana@stikesmuhkudus.ac.id

²⁾FPMIPATI, Universitas PGRI Semarang

email: dewiwulandari1611@gmail.com

Abstract

Linear regression is a statistical method for estimating a linear relationship between one dependent variable and one or more independent variables. In simple linear regression, relationship analysis is carried out between one dependent variable and one independent variable to obtain a linear equation. The linear equation can then be used to predict the value of the dependent variable, if the value of the independent variable is known. The use of linear regression encounters a problem when the linear relationship between dependent variables and independent variables changes at a certain point. Segmentation regression is a regression model in which the relationship between variables is bound to independent variables in the form of piecewise linear relationships. The relationship is represented by two or more straight lines connected in a certain threshold value called a breakpoint. In this paper we will discuss the single breakpoint estimation in simple linear regression.

Keywords: *linear regression, segmented regression, single breakpoint*

1. PENDAHULUAN

Regresi linier merupakan metode statistika yang digunakan untuk memodelkan pengaruh dari satu atau lebih peubah bebas terhadap satu peubah terikat. Peubah terikat merupakan peubah dengan tipe data kontinu dengan skala pengukuran data interval atau rasio dan berdistribusi normal. Sesuai dengan namanya, regresi linier memodelkan hubungan linier antara peubah terikat dengan peubah bebas. Hal ini ditunjukkan pada persamaan model regresi linier dan pada kurva garis lurus pada grafik model regresi linier. Aplikasi dari regresi linier sangat luas dan mencakup berbagai bidang. Oleh karena itu, regresi linier merupakan salah satu metode statistika yang paling sering digunakan.

Penggunaan luas dari regresi linier salah satunya disebabkan oleh pemodelan dari regresi linier yang mudah dan sederhana. Selain itu, interpretasi dari model regresi linier dapat menjelaskan secara langsung seberapa besar pengaruh dari berubahnya nilai dari peubah bebas terhadap perubahan nilai peubah terikat. Namun, hubungan linier dari peubah bebas terhadap peubah terikat terkadang tidak terpenuhi. Dengan kata lain, terdapat kemungkinan hubungan antara peubah terikat dengan peubah bebas tidak linier. Salah satu bentuk ketidaklinieran tersebut yaitu efek terhadap peubah terikat yang ditunjukkan berupa *slope* pada garis regresi berubah pada satu atau lebih nilai tertentu. Nilai tersebut disebut dengan titik ubah, atau *break-point*. Titik ubah tersebut menjadikan terdapat dua atau lebih garis regresi yang masing-masing menunjukkan hubungan linier peubah terikat dan peubah bebas, di mana efek peubah bebas terhadap peubah terikat berbeda antara sebelum dan sesudah titik ubah. Regresi dengan titik ubah disebut dengan regresi segmentasi atau regresi sepotong-sepotong (*piecewise*).

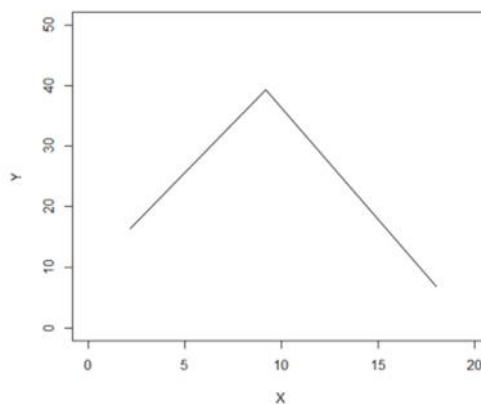
Regresi segmentasi dapat diaplikasikan dalam berbagai bidang. Toms dan Lesperance (2003) mengaplikasikan model regresi segmentasi untuk memodelkan ambang batas pada bidang ekologi. Tomoda dkk (2016) menggunakan regresi segmentasi untuk memprediksi

progresi dari penyakit glaukoma. Regresi segmentasi juga digunakan untuk memodelkan regresi Cox seperti yang dilakukan oleh Muggeo dan Attanasio (2011).

Pemodelan hubungan peubah terikat dan peubah bebas ini telah dibahas di berbagai tulisan. Muggeo (2003) menulis bahwa beberapa metode yang digunakan diantaranya dengan memilih secara manual titik ubah dari diagram pencar (scatter plot) antara peubah terikat dengan peubah bebas. Selain itu titik ubah diestimasi berdasarkan knot yang dihasilkan dari regresi spline, dan juga beberapa metode lainnya. Dalam tulisan ini akan dibahas estimasi titik ubah tunggal dan parameter dari regresi segmentasi sesuai dengan metode yang diajukan oleh Muggeo (2003).

2. KAJIAN LITERATUR

Regresi segmentasi merupakan model regresi di mana hubungan antara peubah terikat dengan peubah bebas sepotong-sepotong linier (*piecewise linear*), yang direpresentasikan dengan dua atau lebih garis lurus yang terhubung pada suatu nilai titik ubah yang tidak diketahui.



Gambar 1. Grafik regresi segmentasi dengan satu titik ubah

Hubungan dalam regresi segmentasi antara peubah terikat dengan satu peubah bebas ditunjukkan pada formula berikut

$$y_i = \gamma + S_1 x_i + S_2 (x_i - \mathbb{E})_+ \quad (1)$$

dengan y_i merupakan peubah terikat, x_i peubah bebas, γ intersep, S_1 slope pertama, S_2 slope kedua, dan \mathbb{E} adalah titik ubah. Selanjutnya

$$(x_i - \mathbb{E})_+ = (x_i - \mathbb{E}) \times I(x_i > \mathbb{E}) \quad (2)$$

Di mana $I(\cdot)$ merupakan fungsi indikator yang bernilai satu jika $(x_i > \mathbb{E})$ terpenuhi.

Estimasi Titik Ubah Tunggal dan Persamaan Regresi

Estimasi titik ubah tunggal untuk mengestimasi model (1) dilakukan dengan metode iterasi. Nilai awal untuk titik ubah $\tilde{\mathbb{E}}$ dipilih selanjutnya secara iteratif membentuk

$$y_i = \gamma + S_1 x_i + S_2 (x_i - \tilde{\mathbb{E}})_+ + \chi I(x_i > \tilde{\mathbb{E}})^- \quad (3)$$

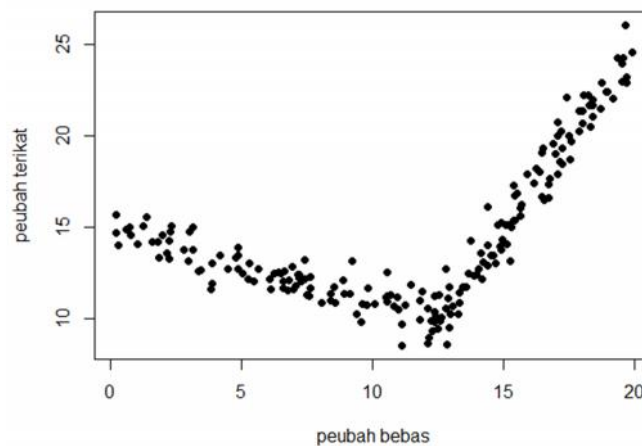
dengan $I(\cdot)^- = -I(\cdot)$ dan χ parameter yang digunakan sebagai reparameterisasi dari \mathbb{E} untuk mengakomodasi estimasi dari titik ubah. Pada setiap iterasi, suatu model linier dibentuk dan nilai titik ubah diganti dengan nilai yang baru dengan formula

$$\hat{\mathbb{E}} = \tilde{\mathbb{E}} + \hat{\chi} / \hat{S}_2 \quad (4)$$

Prosedur iterasi dilakukan sampai kekonvergenan dari $\hat{\epsilon}$ tercapai (Muggeo, 2008). Estimasi titik ubah $\hat{\epsilon}$ sekaligus juga memperoleh estimasi parameter regresi linier dari tiap-tiap segmen.

3. METODE PENELITIAN

Metode penelitian dalam menggunakan studi simulasi data. Simulasi bertujuan untuk membandingkan performa antara regresi linier sederhana dengan regresi segmentasi dengan satu titik ubah. Simulasi dilakukan dengan membangkitkan data berdasarkan model regresi linear dengan satu peubah bebas. Disimulasikan data dengan dua peubah berupa satu peubah bebas dan satu peubah terikat. Diagram pencar (*scatter plot*) dari data simulasi yaitu



Gambar 2. Diagram pencar hasil simulasi data

Tabel 1. Statistika deskriptif data simulasi

	Peubah Bebas (X)	Peubah Terikat (Y)	X segmen1	X segmen2	Y segmen1	Y segmen2
N	200	200	100	100	100	100
Rata-rata	11.32	12	6.76	15.87	12.27	16.59
Variansi	30.01	15.860	13.49	4.94	2.39	20.04
Maksimum	19.93	26.060	12.97	19.93	15.71	26.06
Minimum	0.18	8.531	0.18	12.14	8.53	8.59

Tabel 2. Parameter pada data simulasi

Psi	13
Alfa	15
Beta1	-0.4
Beta2	2

4. HASIL PENELITIAN

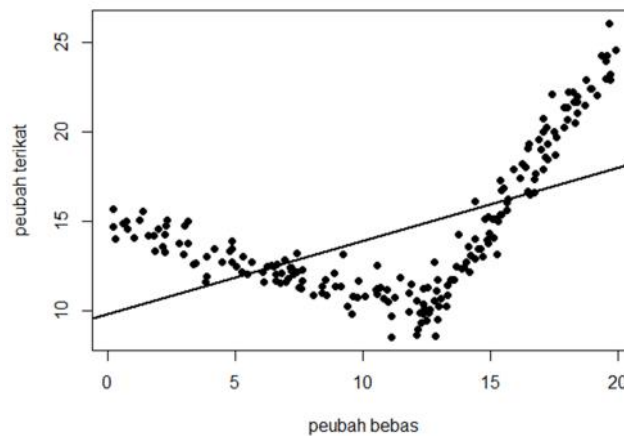
Model Regresi Linier pada Data Simulasi

Estimasi model regresi linier pada data simulasi diperoleh estimasi parameter model regresi sebagai berikut

Tabel 3. Estimasi parameter model regresi linier

	Estimasi	SE	t value	p-value
intersep	9.81901	0.53760	18.264	0.000
x	0.40751	0.04278	9.525	0.000
Uji-F	90.72	df	1: 198	0.000

Dari tabel 3 dapat disimpulkan bahwa model regresi linier signifikan pada tingkat signifikansi 5%. Selanjutnya diperoleh nilai R-square sebesar 0.3142, dan Adjusted R-squared sebesar 0.3108. Nilai R-Square yang diperoleh relatif kecil yang mengindikasikan bahwa model regresi linier tidak dapat mengakomodasi varian data dengan baik. Gambar 3 menunjukkan model fitting dari regresi linier terhadap data simulasi



Gambar 3. Diagram pencar hasil simulasi data dan model fitting regresi linier

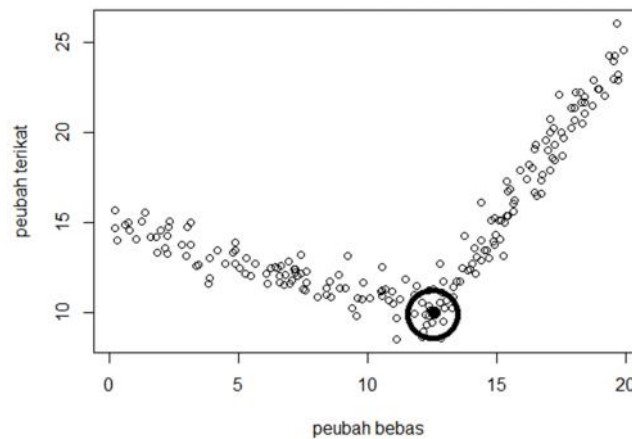
Model Regresi Segmentasi dengan satu titik ubah.

Estimasi titik ubah (*break-point*) tunggal regresi segmentasi pada data hasil simulasi diperoleh sebagai berikut

Tabel 4. Estimasi titik ubah tunggal regresi segmentasi

	psi
Estimasi	12.591
SE	0.095
Uji Davies (p-value)	0.000

Hasil estimasi diperoleh bahwa titik ubah (*break-point*) berada pada $x=12.591$, selanjutnya dari hasil uji Davies untuk menguji kewujudan titik ubah hasil estimasi (Muggeo, 2008), membuktikan bahwa titik ubah $x=12.591$ wujud. Lokasi titik ubah pada data hasil simulasi dapat dilihat pada grafik berikut



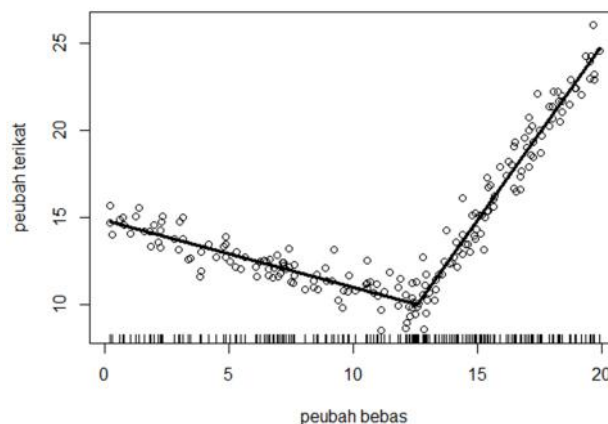
Gambar 4. Lokasi titik ubah (*break-point*)

Estimasi model regresi segmentasi dengan satu titik ubah pada data simulasi diperoleh estimasi parameter model regresi sebagai berikut

Tabel 5. Estimasi parameter model regresi segmentasi titik ubah tunggal

	Estimasi	SE	t value	p-value
intersep	14.84103	0.17702	83.84	0.000
x	-0.38470	0.02229	-17.26	0.000
x_u1	2.38333	0.04760	50.07	-

Dari tabel 5 dapat disimpulkan bahwa model regresi linier signifikan pada tingkat signifikansi 5%. Selanjutnya diperoleh nilai R-square sebesar 0.9548, dan Adjusted R-squared sebesar 0.9541. Nilai R-Square yang diperoleh relatif besar yang mengindikasikan bahwa model regresi segmentasi dengan titik ubah tunggal dapat mengakomodasi varian data dengan baik. Gambar 5 menunjukkan model fitting dari regresi segmentasi dengan titik ubah tunggal terhadap data simulasi



Gambar 5. Diagram pencar hasil simulasi data dan model fitting regresi segmentasi dengan titik ubah tunggal

Estimasi model regresi linier dan model regresi segmentasi dengan titik ubah tunggal menghasilkan pengujian estimasi parameter pada masing-masing model regresi. Hasil pengujian estimasi ditampilkan dalam tabel berikut

Tabel 6. Pengujian estimasi parameter model regresi segmentasi titik ubah tunggal dan model regresi linier

	Model regresi segmentasi dengan titik ubah tunggal	Model regresi linier
intersep	14.84103 (0.000)	9.81901 (0.000)
x	-0.38470 (0.000)	0.40751 (0.000)
x_u1	2.38333	-
R-Square	0.9548	0.3142
Adj. R-Square	0.9541	0.3108
AIC	509.8274	1049.891
BIC	526.3190	1059.786

Berdasarkan hasil pengujian estimasi pada Tabel 6, diperoleh bahwa model regresi segmentasi dengan titik ubah tunggal lebih baik daripada model regresi linier karena mempunyai nilai R-Square dan Adj.R-Square terbesar, serta nilai AIC dan BIC yang terkecil.

5. SIMPULAN

Regresi segmentasi memodelkan hubungan antara peubah terikat dengan peubah bebas di mana terdapat beberapa model linier yang berbeda untuk tiap interval pada peubah bebas yang berbeda. Antar model linier dipisahkan oleh nilai tertentu yang disebut dengan titik ubah (break-point). Pada studi kasus diperoleh hasil bahwa pemodelan menggunakan model regresi segmentasi dengan titik ubah tunggal dapat memodelkan data lebih baik daripada model regresi linier klasik. Hal ini dapat dilihat pada nilai R-Square, Adj. R-Square, AIC, dan BIC dari model regresi segmentasi yang lebih kecil daripada nilai R-Square, Adj. R-Square, AIC, dan BIC yang dihasilkan oleh model regresi linier klasik.

Model regresi segmentasi tidak hanya terbatas pada model titik ubah tunggal dengan satu peubah bebas. Model dapat dikembangkan dengan titik ubah ganda, maupun model regresi linier dengan lebih dari satu peubah bebas. Selain itu, dimungkinkan juga untuk memodelkan regresi segmentasi pada model regresi seperti regresi logistik, regresi poisson, maupun regresi Cox.

6. REFERENSI

- Muggeo, V. M. (2003). Estimating Regression Models with Unknown Break-points. *Statistics in Medicine*, 22(19), 3055-3071. doi:https://doi.org/10.1002/sim.1545
- Muggeo, V. M. (2008). segmented: An R package to Fit Regression Models with Broken-Line Relationships. *R NEWS*, 8(1), 20-25.
- Muggeo, V. M., & Attanasio, M. (2011). Segmented relationships to model erosion of regression effect in Cox regression. *Statistical Methods in Medical Research*, 20, 401-414.
- Tomoda, K., Morino, K., Murata, H., Asaoka, R., & Yamanishi, K. (2016). Predicting Glaucomatous Progression with Piecewise Regression Model. *BIOSTEC 2016 Proceedings of the International Joint Conference on Biomedical Engineering Systems and Technologies* (pp. 93-104). Rome: SCITEPRESS.
- Toms, J. D., & Lesperance, M. L. (2003). Piecewise Regression: A Tool for Identifying Ecological Thresholds. *Ecology*, 84(8), 2034-2041. doi:https://doi.org/10.1890/02-0472