KAJIAN ANALISIS REGRESI LINIER TERSEGMEN
(*Segmented Linear Regression*)

SKRIPSI

oleh :
ARINA SHOFIYATI
0410950005-95

PROGRAM STUDI STATISTIKA
JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS BRAWIJAYA
MALANG
2008

KAJIAN ANALISIS REGRESI LINIER TERSEGMEN
(*Segmented Linear Regression*)

SKRIPSI

Sebagai salah satu syarat untuk memperoleh gelar
Sarjana Sains dalam bidang Statistika

oleh :
ARINA SHOFIYATI
0410950005-95

PROGRAM STUDI STATISTIKA
JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS BRAWIJAYA
MALANG
2008

LEMBAR PENGESAHAN SKRIPSI

KAJIAN ANALISIS REGRESI LINIER TERSEGMEN
(*Segmented Linear Regression*)

oleh :
ARINA SHOFIYATI
0410950005-95

Setelah dipertahankan di depan Majelis Penguji
pada tanggal 3 Desember 2008
dan dinyatakan memenuhi syarat untuk memperoleh gelar
Sarjana Sains dalam bidang Statistika

Pembimbing I	Pembimbing II
Suci Astutik, SSi., MSi.	**Eni Sumarminingsih, SSi., MM**
NIP. 132 233 148	**NIP. 132 300 241**

Mengetahui,
Ketua Jurusan Matematika
Fakultas MIPA Universitas Brawijaya

Dr.Agus Suryanto, MSc.
NIP. 132 126 049

LEMBAR PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Arina Shofiyati
NIM : 0410950005-95
Jurusan : Matematika
Penulis skripsi berjudul : KAJIAN ANALISIS REGRESI LINIER TERSEGMEN (*Segmented Linear Regression*)

Dengan ini menyatakan bahwa:

1. **Skripsi ini adalah benar-benar karya sendiri dan bukan hasil plagiat dari karya orang lain. Karya-karya yang tercantum dalam Daftar Pustaka skripsi ini, semata-mata digunakan sebagai acuan/referensi.**
2. **Apabila kemudian hari diketahui bahwa isi skripsi saya merupakan hasil plagiat, maka saya bersedia menanggung akibat hukum dari keadaan tersebut.**

Demikian pernyataan ini dibuat dengan segala kesadaran.

Malang, 3 Desember 2008
Yang menyatakan,

Arina Shofiyati
NIM. 0410950005

KAJIAN ANALISIS REGRESI LINIER TERSEGMEN

ABSTRAK

Analisis regresi digunakan untuk menggambarkan hubungan antara peubah respon Y dengan peubah penjelas X dalam suatu bentuk model regresi. Akan tetapi, belum tentu model yang telah dispesifikasikan tersebut cocok dengan data. Hal ini bisa saja dikarenakan setelah nilai x tertentu, titik-titik data pengamatan berada jauh dari garis regresi, atau dengan kata lain peubah respon menunjukkan pola yang berbeda setelah nilai x tertentu. Dengan demikian, satu model kurang sesuai jika digunakan untuk mewakili hubungan antara peubah respon dan peubah penjelas. Oleh karena itu, dibutuhkan suatu model yang terdiri dari beberapa submodel, di mana masing-masing submodel digunakan pada rentang nilai x yang berbeda, yang dikenal dengan model regresi linier tersegmen. Bila digambarkan, model ini berupa rangkaian garis linier yang patah-patah. Tujuan penelitian ini adalah menerapkan penggunaan analisis regresi linier tersegmen sebagai pendekatan linier terhadap bentuk kurva model regresi yang tidak linier. Dalam penelitian ini, digunakan tiga data sekunder, masing-masing data dianalisis dengan analisis regresi linier tersegmen, di mana satu set data dibagi menjadi dua bagian berdasarkan titik tertentu yang disebut *breakpoint.* Selain itu, masing-masing data juga dianalisis dengan polinom derajat 2 (regresi kuadratik) yang digunakan sebagai model pembanding. Berdasarkan hasil analisis, model regresi linier tersegmen memberikan nilai R^2 *adjusted* yang lebih besar dibanding model regresi kuadratik dengan selisih sebesar 2.96%, 8.45% dan 1.01% untuk data 1,2 dan 3 berturut-turut.

Kata kunci : regresi linier tersegmen, *breakpoint*, regresi kuadratik

SEGMENTED LINEAR REGRESSION

ABSTRACT

Regression analysis is used to describe the relationship between response variable Y and explanatory variable X in a regression model. In fact, may be the regression model that ha ve been specified have no good fit to the data. This situation happen because the data point lies far from the regression line after a certain point. In other word, the response variable shows the different pattern after a certain value of x. Because of that, segmented linear regression analysis is presented. This model consists of some submodels which is used for different ranges of X. The segment ed linear regression model is described as a sequence of broken line. The purpose of this study is to apply the segmented linear regression analysis to the data that tend to have nonlinear regression curve. In this study, three secondary data is used. The segmented linear regression analysis is applied to each data, where a set of data divided into two segments based on a certain value, called breakpoint. Beside that, each data is analyzed with second degree polynomial (cuadratic regression) as comparison model. The result shows that segmented linear regression model gives greater R^2 adjusted than cuadratic regression do, with difference 2.96%, 8.45% and 1.01% for data 1,2 and 3 respectively.

Keywords : segmented linear regression, breakpoint, cuadratic regression

KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Allah SWT. atas limpahan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul "*Kajian Analisis Regresi Linier Tersegmen (Segmented Linear Regression)*" sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains dalam bidang Statistika. Oleh karena itu penulis mengucapkan rasa hormat dan terima kasih kepada:

1. Ibu Suci Astutik, S.Si., M.Si., selaku Dosen Pembimbing I dan ibu Eni Sumarminingsih, S.Si., MM selaku Dosen Pembimbing II atas arahan serta nasehat yang telah diberikan kepada penulis selama penyusunan skripsi ini.
2. Ibu Dr. Ir. Maria Bernadetha Mitakda selaku dosen pembimbing akademik yang telah memberikan pengarahan dan masukan selama penyusunan proposal skripsi.
3. Ibu Ir. Soepraptini, bapak Adji Achmad R.F. dan bapak Prof.Dr.Ir. Loekito A.S. selaku Dosen Penguji.
4. Segenap Bapak dan Ibu Dosen Statistika atas kesabaran dalam mendidik selama kuliah hingga selesai.
5. Bapak, Ibu, Mbak, Adik dan segenap keluarga yang senantiasa memberikan doa dan semangat.
6. Teman-teman Program Studi Statistika 2004 yang telah memberikan dukungan, semangat dan bantuan.
7. Semua pihak yang tidak dapat penulis sebutkan satu per satu yang telah banyak membantu dan memberikan bantuan selama penulisan skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna, mengingat keterbatasan kemampuan penulis. Untuk itu dengan segala kerendahan hati penulis mengharap kritik dan saran. Akhir kata, penulis berharap semoga skripsi ini dapat bermanfaat bagi pembaca dan khususnya bagi penulis.

Malang, Desember 2008

Penulis

DAFTAR ISI

Halaman

BAB III BAHAN DAN METODE

BAB IV HASIL DAN PEMBAHASAN

BAB V KESIMPULAN DAN SARAN

DAFTAR GAMBAR

Halaman

DAFTAR TABEL

Halaman

DAFTAR LAMPIRAN

Halaman

BAB I
PENDAHULUAN

1.1 Latar Belakang

Analisis regresi digunakan untuk menggambarkan hubungan antara peubah respon Y dengan peubah penjelas X dalam suatu bentuk model regresi, dengan harapan model tersebut dapat menjelaskan sebanyak mungkin informasi yang ada pada data. Akan tetapi, belum tentu model yang telah dispesifikasikan tersebut cocok dengan data. Hal ini bisa saja dikarenakan setelah nilai X tertentu, titik-titik data pengamatan berada jauh dari garis regresi, atau dengan kata lain peubah respon menunjukkan pola yang berbeda setelah nilai X tertentu. Dengan demikian, satu model kurang sesuai jika digunakan untuk mewakili hubungan antara peubah respon dan peubah penjelas. Sehingga, dibutuhkan suatu model yang terdiri dari beberapa submodel, di mana masing-masing submodel digunakan pada rentang nilai X yang berbeda.

Pada analisis regresi linier sederhana, model regresi digambarkan sebagai sebuah garis linier. Analisis ini bisa dilakukan terhadap seluruh data atau pun membagi nilai-nilai peubah penjelas menjadi beberapa bagian (segmen) kemudian menerapkan analisis regresi pada setiap segmen, yang dikenal dengan analisis regresi linier tersegmen (*Segmented Linear Regression*)(Anonymous, 2006). Model regresi linier tersegmen terdiri dari beberapa submodel linier, bila digambarkan model ini berupa rangkaian garis linier yang patah-patah. Oleh karena itu, model ini bisa digunakan sebagai pendekatan terhadap bentuk kurva model regresi yang tidak linier.

Dalam analisis regresi linier tersegmen, terdapat suatu titik yang disebut *breakpoint*, yaitu titik batas antar tiap segmen (titik patah). Pada titik ini diduga mulai terjadi perubahan bentuk hubungan matematis antara peubah respon dengan peubah penjelas. Titik ini juga digunakan sebagai indikator banyaknya segmen (s), di mana s = banyaknya *breakpoint* + 1.

1.2 Rumusan Masalah

Bagaimana menerapkan model regresi linier tersegmen pada pasangan data yang mempunyai kecenderungan bentuk kurva model regresi yang tidak linier?

1.3 Batasan Masalah

Studi ini dibatasi pada analisis regresi dengan satu peubah bebas dan analisis regresi linier tersegmen dengan satu *breakpoint* yang bersifat kontinu, dengan anggapan bahwa semua asumsi analisis regresi terpenuhi.

1.4 Tujuan

Tujuan penelitian ini adalah menerapkan penggunaan analisis regresi linier tersegmen sebagai pendekatan linier untuk pasangan data yang mempunyai kecenderungan bentuk kurva model regresi yang tidak linier.

1.5 Manfaat

Manfaat penelitian ini yaitu mendapatkan model yang lebih representatif terhadap data. Di mana, satu set data diwakili oleh dua model regresi linier, masing-masing untuk nilai x yang berbeda.

BAB II
TINJAUAN PUSTAKA

2.1 Diagram Pencar

Sebelum memastikan model regresi yang sesuai untuk pasangan data, sebaiknya dibuat plot data terlebih dahulu pada susunan sumbu (X,Y), di mana X sebagai absis dan Y sebagai ordinat. Titik-titik yang diperoleh dari pengeplotan tersebut disebut sebagai diagram pencaran. Hal ini dimaksudkan untuk memperkecil kesalahan dalam menentukan bentuk hubungan antara peubah respon dan peubah penjelas (Yitnosumarto, 1985).

Riduwan (2006) menyatakan bahwa diagram pencar disebut juga dengan diagram titik, yaitu diagram yang menunjukkan gugusan titik-titik pada garis koordinat tanpa garis penghubung antar tiap titik. Diagram ini biasanya digunakan untuk menggambarkan titik data korelasi atau regresi yang terdiri dari peubah respon dan peubah penjelas.

2.2 Model Regresi

Menurut Sembiring (1995), analisis regresi adalah sebuah teknik statistika untuk membentuk suatu model dalam menentukan hubungan kausal antara dua peubah atau lebih. Model ini merupakan fungsi dari peubah-peubah tersebut dan digunakan untuk memahami, menerangkan dan memprediksikan perilaku sistem yang diamati.

Model regresi linier sederhana untuk populasi adalah:

$$y_i = \alpha + \beta x_i + \varepsilon_i \tag{2.1}$$

di mana:
- i = 1,2,...,N
- y_i = nilai pengamatan peubah respon individu ke-i
- x_i = nilai pengamatan peubah penjelas individu ke-i
- α = titik potong garis regresi dengan sumbu Y (intersep)
- β = koefisien regresi (slope)
- ε_i = galat individu ke-i
- N = ukuran populasi

Untuk menduga parameter α dan β, digunakan Metode Kuadrat Terkecil Biasa (MKTB) yang meminimumkan jumlah kuadrat sisa:

$$JK_S = \sum_{i=1}^{N} \varepsilon_i^2 = \sum_{i=1}^{N} (y_i - \alpha - \beta x_i)^2$$

Dengan menurunkan $\sum_{i=1}^{N} \varepsilon_i^2$ terhadap α dan β kemudian disamakan dengan nol, diperoleh persamaan normal:

$$\frac{\partial JK_S}{\partial \alpha} = -2\sum (y_i - \hat{\alpha} - \hat{\beta} x_i) = 0$$

$$\frac{\partial JK_s}{\partial \beta} = -2\sum (y_i - \hat{\alpha} - \hat{\beta} x_i) x_i = 0 \qquad (2.2)$$

Penggantian α dan β dengan penduganya pada gugus persamaan (2.2) menghasilkan:

$$na + b\sum x_i = \sum y_i$$

$$a\sum x_i + b\sum x_i^2 = \sum x_i y_i \qquad (2.3)$$

Persamaan 2.3 dapat ditulis dalam bentuk matriks:

$$\begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix} \text{ atau } (\mathbf{X'X})\mathbf{b} = \mathbf{X'Y}$$

di mana: $\mathbf{Y}_{n \times 1} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ $\mathbf{X}_{n \times 2} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$ dan $\mathbf{b}_{2 \times 1} = \begin{bmatrix} a \\ b \end{bmatrix}$

Dengan demikian, penduga bagi $\boldsymbol{\beta}$ adalah:

$$\mathbf{b} = (\mathbf{X'X})^{-1}\mathbf{X'Y} \qquad (2.4)$$

Persamaan (2.4) adalah rumus umum untuk menduga parameter model regresi dengan k peubah penjelas (Draper and Smith, 1992).

2.2.1 Pengujian koefisien regresi

Pengujian terhadap koefisien regresi populasi dilakukan dengan metode analisis ragam yang memecah keragaman total menjadi komponen regresi dan sisa. Hipotesis yang melandasi pengujian adalah:

$H_0 : \beta = 0$

$H_1 : \beta \neq 0$

Pengujian koefisien regresi populasi disajikan dalam tabel 2.1.

Tabel 2.1. Tabel analisis ragam untuk pengujian koefisien regresi

SK	db	JK	KT	F
Regresi	1	$\sum(\hat{y}_i - \bar{y})^2$	JK_R	KT_R/KT_S
Sisa	n-2	$\sum(y_i - \hat{y}_i)^2$	JK_S/ n-2	
Total	n-1	$\sum(y_i - \bar{y})^2$		

H_0 ditolak apabila statistik uji F > $F^{\alpha}_{(1,n-2)}$ atau *p-value* < α, yaitu taraf nyata yang digunakan.

Selain dengan analisis ragam, pengujian koefisien regresi juga dapat dilakukan dengan uji T, berlandaskan hipotesis:

$H_0 : \beta = 0$

$H_1 : \beta \neq 0$

dengan statistik uji:

$$t = \frac{|b|}{\sqrt{s^2 / \sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

Hipotesis nol ditolak apabila statistik uji t > $t^{\alpha/2}_{(n-2)}$ atau *p-value* < α (Yitnosumarto, 1985).

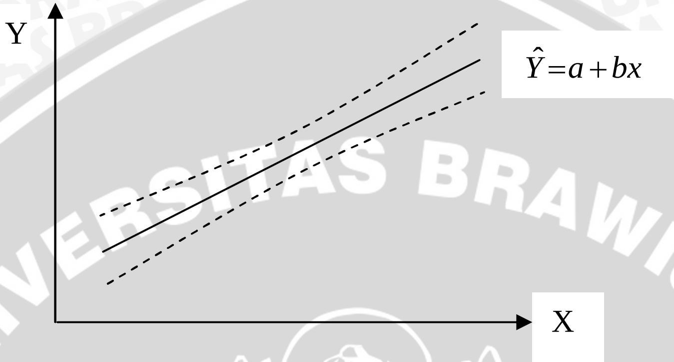
2.2.2 Selang kepercayaan (1-α)% untuk $E(Y|x_0)$

Berdasarkan model contoh $\hat{Y}_0 = a + bx_0$ untuk nilai x = x_0, maka taksiran nilai Y untuk nilai x tertentu (x_0) yang tidak diamati dalam contoh bisa ditentukan, yaitu $E(Y|x_0)$. Sedangkan selang kepercayaan (1-α)% untuk $E(Y|x_0)$ adalah:

$$\hat{y}_0 - t_{(n-2;\alpha/2)}.s(\hat{y}_0) \leq E(Y|x_0) \leq \hat{y}_0 + t_{(n-2;\alpha/2)}.s(\hat{y}_0)$$

dengan $s(\hat{y}_0) = \left(\sigma^2\left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right]\right)^{\frac{1}{2}}$

Apabila selang kepercayaan ini dikerjakan pada banyak nilai x_0, kemudian menghubungkan ujung-ujung selang tersebut, maka diperoleh gambar seperti tampak pada Gambar 2.1.



Gambar 2.1. Pita selang kepercayaan untuk model $\hat{Y} = a + bx$ (Sembiring, 1995).

2.2.3 Koefisien determinasi (R^2)

Setelah menduga model regresi dari data, masalah selanjutnya adalah menilai baik buruknya kecocokan model regresi yang digunakan dengan data. Sehingga, diperlukan ukuran tentang kecocokan model, yaitu koefisien determinasi (R^2) dengan definisi matematis:

$$R^2 = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} = \frac{JK_R}{JK_T} = 1 - \frac{JK_S}{JK_T} \qquad (2.5)$$

Semakin dekat R^2 dengan 1 semakin baik kecocokan data dengan model, dan sebaliknya, semakin dekat R^2 dengan 0 semakin jelek kecocokan data dengan model.

Besarnya R^2 dipengaruhi oleh banyaknya peubah penjelas dalam model, semakin banyak peubah penjelas yang masuk dalam model semakin besar R^2 yang diperoleh. Untuk mengatasi hal tersebut, digunakan R^2 *adjusted* (R^2 disesuaikan) yang dilakukan dengan membagi JK_S dan JK_T dengan derajat bebasnya masing-masing sehingga diperoleh:

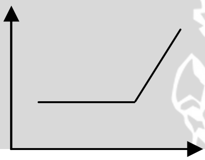
$$R^2 adj. = 1 - \frac{JK_S/(n-k)}{JK_T/n-1} \qquad (2.6)$$

(Sembiring, 1995).

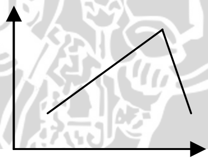
2.3 Bentuk Model Regresi Linier Tersegmen

Model regresi linier tersegmen adalah model regresi yang terdiri dari dua submodel yang bersifat linier. Model ini banyak digunakan di berbagai bidang, seperti ekonomi, kedokteran dan lain sebagainya dan digunakan apabila terdapat indikasi perubahan parameter setelah nilai tertentu pada peubah penjelas. Dengan demikian, nilai-nilai pada peubah penjelas terbagi menjadi dua bagian (segmen).

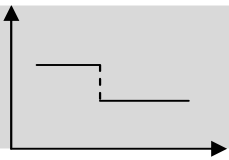
Secara umum, model regresi linier tersegmen dibagi menjadi dua tipe, yaitu model regresi yang kontinu, di mana model regresi linier di segmen pertama bertemu dengan model regresi linier di segmen kedua pada *breakpoint*, seperti tampak pada Gambar 2.2a dan 2.2b dan model regresi yang tidak kontinu, di mana model regresi linier di segmen pertama tidak bertemu dengan model regresi linier di segmen kedua pada *breakpoint*, seperti terlihat pada Gambar 2.2c dan 2.1d (Diniz and Brochi, 2005).



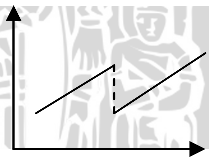
a. Regresi linier tersegmen yang kontinu pada *breakpoint* dengan slope1 = 0 dan slope2 > 0



b. Regresi linier tersegmen yang kontinu pada *breakpoint* dengan slope1 > 0 dan slope2 < 0



c. Regresi linier tersegmen yang tidak kontinu pada *breakpoint* dengan slope = 0



d. Regresi linier tersegmen yang tidak kontinu pada *breakpoint* dengan slope1 dan slope2 >0

Gambar 2.2. Bentuk fungsi regresi linier tersegmen

Secara khusus, ada 7 tipe model pada analisis regresi linier tersegmen, yaitu:

1. Tipe 0 – sebuah garis horizontal tanpa *breakpoint*
2. Tipe 1 – sebuah garis miring tanpa *breakpoint* (seperti garis regresi linier sederhana)
3. Tipe 2 – garis miring pada segmen pertama dan kedua (seperti Gambar 2.2b)

4. Tipe 3 – garis horizontal pada segmen pertama dan garis miring pada segmen kedua (seperti Gambar 2.2a)
5. Tipe 4 – garis miring pada segmen pertama dan garis horizontal pada segmen kedua
6. Tipe 5 – garis horizontal pada segmen pertama dan kedua dengan level yang berbeda (seperti Gambar 2.2c)
7. Tipe 6 – dua garis miring yang terpisah (seperti Gambar 2.2d)

Model yang telah diduga digunakan untuk memprediksi perilaku sistem yang diamati, bukan meramalkan. Istilah prediksi mempunyai arti khusus yaitu interpolasi, yakni mencari suatu nilai fungsi yang tidak diketahui di antara beberapa nilai fungsi yang diketahui.

2.4 Analisis Regresi Linier Tersegmen (*Segmented Linear Regression*)

Analisis regresi linier tersegmen adalah suatu metode dalam analisis regresi yang membagi peubah penjelas menjadi beberapa segmen berdasarkan nilai tertentu yang disebut dengan *breakpoint* di mana, pada setiap segmen data terdapat model regresi linier (Anonymous, 2006).

Analisis regresi linier tersegmen dengan satu *breakpoint* disebut juga dengan analisis regresi linier dua fase. Pada metode ini, garis regresi tidak lagi disajikan dalam satu garis linier, melainkan disajikan dalam dua garis linier yang bertemu pada suatu titik, yaitu titik x = c. Dengan demikian, terdapat dua model regresi berikut:

$$Y = a_1 + b_1X \qquad x \leq c$$
$$Y = a_2 + b_2X \qquad x \geq c$$

Pada saat titik x = c,

$$a_1 + b_1c = a_2 + b_2c \tag{2.7}$$

di mana, titik x = c disebut sebagai *breakpoint*.

Persamaan 2.5 juga dapat dituliskan dalam bentuk:

$$a_2 = a_1 + c\,(b_1 - b_2)$$

Jika a_2 disubstitusikan ke model regresi $Y = a_2 + b_2X$, maka akan diperoleh bentuk lain dari model analisis regresi linier tersegmen yaitu:

$$Y = a_1 + b_1X \qquad x \le c$$
$$Y = \{a_1 + c\,(b_1-b_2)\} + b_2X \qquad x \ge c$$

(Ryan and Porth, 2007)

Dalam analisis regresi linier tersegmen, *breakpoint* bisa saja sudah diketahui sebelum analisis. Tapi, pada umumnya titik ini tidak diketahui dan harus diduga terlebih dahulu.

2.4.1 Pendugaan kuadrat terkecil jika *breakpoint* diketahui

Misalkan terdapat suatu model regresi dua fase:

$$y_{si} = \alpha_s + \beta_s x_{si} + \varepsilon_{si} \qquad i = 1,2,...,n_s \quad s = 1,2$$

dengan $x_{[11]} < x_{[12]} < ... < x_{[1n_1]} < \gamma < x_{[21]} < x_{[22]} < ... < x_{[2n_2]}$ dan γ diketahui

di mana : $x_{[si]}$ = nilai peubah penjelas yang telah diurutkan dari nilai terkecil ke nilai terbesar

γ = *breakpoint*

n_s = banyaknya pasangan data pada segmen ke - s

s = banyaknya segmen

Jika *breakpoint* diketahui, maka pendugaan parameter model regresi linier tersegmen dilakukan dengan metode kuadrat terkecil terkendala (MKTK) dengan kendala pada persamaan (2.7), yang diselesaikan dengan metode *Lagrange*. Sehingga, jumlah kuadrat sisa pada model regresi linier tersegmen, adalah:

$$JK_{SK} = \sum_{s=1}^{2}\sum_{i=1}^{n_s}\left(y_{si} - \alpha_s - \beta_s x_{si}\right)^2 + 2\lambda\left(\alpha_2 - \alpha_1 + \gamma\left(\beta_2 - \beta_1\right)\right)$$

-2λ adalah pengganda *Lagrange*. Setelah menurunkan JK_{SK} terhadap α_s, β_s kemudian disamakan dengan nol, diperoleh persamaan:

$$-2\left(\sum_i y_{1i} - n_1\tilde{\alpha}_1 - \sum_i x_{1i}\tilde{\beta}_1\right) - 2\lambda = 0 \qquad (2.8)$$

$$-2\left(\sum_i y_{2i} - n_2\tilde{\alpha}_2 - \sum_i x_{2i}\tilde{\beta}_2\right) + 2\lambda = 0 \qquad (2.9)$$

$$-2\left(\sum_i x_{1i}\left(y_{1i} - \tilde{\alpha}_1 - \tilde{\beta}_1 x_{1i}\right)\right) - 2\lambda\gamma = 0 \qquad (2.10)$$

$$-2\left(\sum_i x_{2i}\left(y_{2i}-\tilde{\alpha}_2-\tilde{\beta}_2 x_{2i}\right)\right)+2\lambda\gamma=0 \qquad (2.11)$$

Dari persamaan (2.8) dan (2.9) didapatkan :

$$\tilde{\alpha}_s=\bar{y}_s-\tilde{\beta}_s\bar{x}_s+(-1)^{s-1}\lambda n_s^{-1} \qquad (2.12)$$

$\tilde{\alpha}_s, \tilde{\beta}_s$ adalah penduga parameter α_s, β_s dari metode kuadrat terkecil terkendala.

Dengan mensubstitusikan persamaan (2.12) ke dalam persamaan (2.7) diperoleh:

$$\lambda=w\left\{\bar{y}_2-\bar{y}_1+\tilde{\beta}_1\left(\bar{x}_1-\gamma\right)-\tilde{\beta}_2\left(\bar{x}_2-\gamma\right)\right\} \qquad (2.13)$$

di mana: $w = n_1 n_2/(n_1+n_2)$.

Selanjutnya, substitusi persamaan (2.12) dan (2.13) ke persamaan (2.10) dan (2.11), menghasilkan sistem persamaan untuk menghitung nilai $\tilde{\beta}_s$:

$$c_{11}\tilde{\beta}_1+c_{12}\tilde{\beta}_2=c_{13}$$

$$c_{21}\tilde{\beta}_1+c_{22}\tilde{\beta}_2=c_{23}$$

di mana: $c_{ss}=\sum_i\left(x_{si}-\bar{x}_s\right)^2+w\left(\bar{x}_s-\gamma\right)^2$

$$c_{12}=c_{21}=-w\left(\bar{x}_1-\gamma\right)\left(\bar{x}_2-\gamma\right)$$

dan, $c_{s3}=\sum_i\left(y_{si}-\bar{y}_s\right)\left(x_{si}-\bar{x}_s\right)+\left(-1\right)^s w\left(\bar{y}_2-\bar{y}_1\right)\left(\bar{x}_s-\gamma\right)$

Setelah $\tilde{\beta}_1$ dan $\tilde{\beta}_2$ diketahui, diperoleh nilai λ pada persamaan (2.13) dan $\tilde{\alpha}_s$ pada persamaan (2.12). Dengan demikian, nilai minimum $\varepsilon'\varepsilon$ adalah:

$$\sum_{s=1}^{2}\sum_{i=1}^{n_s}\left(y_{si}-\tilde{\alpha}_s-\tilde{\beta}_s x_{si}\right)^2$$

$$=\sum\sum\left\{y_{si}-\bar{y}_s-\tilde{\beta}_s\left(x_{si}-\bar{x}_s\right)+\left(-1\right)^s\lambda n_s^{-1}\right\}^2$$

$$=\sum\sum\left(y_{si}-\bar{y}_s\right)^2-2\sum\sum\tilde{\beta}_s\left(y_{si}-\bar{y}_s\right)\left(x_{si}-\bar{x}_s\right)$$

$$+\sum\sum\tilde{\beta}_s^2(x_{si}-\bar{x}_s)^2+\frac{\lambda^2}{w}$$

yang selanjutnya disebut JK_{SK} (Seber, 1977).

2.4.2 Pendugaan kuadrat terkecil jika *breakpoint* tidak diketahui

Apabila *breakpoint* tidak diketahui, maka harus diduga terlebih dahulu. Langkah pertama adalah membagi titik-titik data menjadi dua segmen, hanya untuk menentukan n_1 dan n_2 tanpa mengetahui nilai γ sesungguhnya. Dari setiap bagian dengan n_s pasangan data, dibentuk model regresi linier sederhana untuk memperoleh nilai $\hat{\alpha}_s$ dan $\hat{\beta}_s$. Sehingga, berdasarkan persamaan 2.7 dan asumsi bahwa $x_{1n_1} < \gamma < x_{21}$, nilai duga untuk γ adalah:

$$\hat{\gamma} = -\left(\frac{\hat{\alpha}_1 - \hat{\alpha}_2}{\hat{\beta}_1 - \hat{\beta}_2}\right)$$

Pendugaan terhadap *breakpoint* dilakukan tidak hanya pada satu titik, melainkan pada beberapa titik dengan tujuan mendapatkan *breakpoint* optimum. *Breakpoint* optimum ditentukan dengan memilih *breakpoint* yang memberikan jumlah kuadrat sisa paling kecil dan R^2 paling besar di antara penduga *breakpoint* yang lain. Di mana, R^2 untuk model regresi linier tersegmen didefinisikan:

$$R^2 = 1 - \frac{JK_{SK}}{JK_T}$$

Koefisien determinasi pada model regresi linier tersegmen pada dasarnya sama dengan koefisien determinasi pada model regresi linier sederhana, perbedaannya terletak pada jumlah kuadrat sisa yang digunakan. Jumlah kuadrat sisa pada model regresi tersegmen diperoleh dengan menjumlahkan nilai kuadrat dari sisaan (y_i- $\hat{y}$),di mana $\hat{y}$ dihitung berdasarkan submodel pertama untuk nilai x ≤ *breakpoint* dan $\hat{y}$ dihitung berdasarkan submodel kedua untuk nilai x ≥ *breakpoint* (Oosterbaan, et al., 1990).

Selang kepercayaan untuk model regresi linier tersegmen juga mempunyai prinsip yang sama seperti selang kepercayaan pada analisis regresi linier sederhana, n (ukuran contoh) yang digunakan bukan n pada setiap segmen, melainkan n secara keseluruhan.

2.5 Regresi Polinom

Dalam analisis regresi linier dengan model regresi umum:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k + \varepsilon$$

di mana k = banyaknya peubah penjelas, dimisalkan bahwa hubungan antara peubah respon (Y) dengan peubah penjelas X_1, X_2, ... , X_k berbentuk linier. Dalam banyak hal, X_1, X_2, ... , X_k terikat dalam suatu hubungan yang berbentuk polinom $X_1 = X$, $X_2 = X^2$, ... , $X_k = X^k$. Sehingga, terbentuk suatu model regresi polinom:

$$Y = \beta_0 + \beta_1 X^1 + \beta_2 X^2 + ... + \beta_k X^k + \varepsilon$$

Polinom banyak digunakan dalam menghampiri suatu kurva, artinya suatu kurva selalu dapat dihampiri oleh suatu deret polinom. Melalui 2 titik selalu dapat dibuat satu garis lurus. Melalui 3 titik dapat dibuat suatu parabol, melalui n titik dapat dibuat suatu polinom berderajat n-1. Jadi, kalau tersedia n titik data kemudian dicocokkan polinom derajat n-1, maka akan diperoleh kecocokan sempurna, R^2 = 100%. Akan tetapi, tidak ada pengetahuan tambahan apapun yang diperoleh.

Dalam menentukan derajat suatu polinom, perlu diperhatikan unsur kesederhanaan. Semakin sederhana suatu model semakin baik model tersebut. Penentuan derajat polinom dengan derajat tinggi sebaiknya dihindari. Karena penambahan parameter pada model regresi polinom akan menyebabkan derajat bebas galat berkurang. Akibatnya, presisi dari penduga parameter akan semakin menurun. Model yang diharapkan yaitu model polinom dengan derajat serendah mungkin tapi dengan kecocokan yang cukup tinggi (Sembiring, 1995).

Pendugaan parameter pada model regresi polinom dilakukan dengan membuat peubah baru X_2, X_3,..., X_k yang merupakan transformasi dari X^2, X^3, ... , X^k. Kemudian meregresikan peubah respon (Y) terhadap peubah penjelas (X) dan peubah baru X_2, X_3,..., X_k dengan metode kuadrat terkecil biasa (MKTB).

Menurut Gujarati (1988), selama model regresi polinom berderajat k masih bersifat linier dalam parameter, maka parameter model regresi ini masih bisa diduga dengan MKTB, dengan menggunakan persamaan 2.4.

2.6 Aplikasi Analisis Regresi Linier Tersegmen

Pada penelitian ini, digunakan tiga data sekunder hasil penelitian, yaitu :

1. Ni Luh Putu Novi Anggraeni Ayu (2003), yang melakukan penelitian untuk mengkaji sifat zeolit terhadap adsorpsi kation Fe^{3+} dan anion Cl^-. Salah satu tahapan penelitian yang dilakukan

adalah uji adsorpsi anion Cl^- yang diukur pada berbagai panjang gelombang. Uji ini dilakukan untuk menentukan panjang gelombang yang dapat memberikan absorbansi anion Cl^- maksimum.

2. Hermawan Guntoro (2001), yang melakukan penelitian untuk menentukan pH optimal penggunaan membran misel balik untuk memisahkan kation Fe(III) dari larutan akuatik. Salah satu tahapan penelitian tersebut adalah menentukan konsentrasi kritis misel (KKM) pada rentang konsentrasi antara 5 $x10^{-3}$M hingga 9.5 $x10^{-3}$M, di mana pada titik tersebut berlangsung transport terbaik bagi kation.
3. Aftiajeng Yulia Safitri (2006), yang meneliti hubungan antara jumlah sel somatis dengan kadar lemak susu. Berdasarkan literatur, peningkatan jumlah sel somatis berhubungan dengan penurunan kandungan lemak susu, karena kemampuan kelenjar susu untuk memproduksi lemak susu menurun. Di mana kadar sel somatis yang tinggi dalam susu mengindikasikan bahwa susu abnormal, salah satunya disebabkan kelenjar susu yang terinfeksi oleh bakteri.

BAB III
BAHAN DAN METODE

3.1 Bahan

Pada penelitian ini, digunakan 3 data sekunder, yaitu:

Data	Sumber	Peubah
1	Ayu, N.L.P.N.A. 2003. Pengaruh Fosfatasi pada Zeolit Alam Turen Terhadap Absorpsi Kation Fe^{3+} dan Anion Cl^-	Y = Absorbansi larutan X = Panjang gelombang (nm)
2	Guntoro, H. 2001. Kajian Awal Pengaruh pH Fasa Eksternal Terhadap Pemisahan Kation Fe(III) Dengan Menggunakan Membran Misel Balik	Y = daya hantar spesifik (ms/cm) X = konsentrasi $NaC_{12}H_{25}SO_4$ (M)10^3
3	Safitri, A.Y. 2006. Hubungan Antara Jumlah Sel Somatis Dengan Kadar Lemak Dan Profil Asam Susu	Y = kadar lemak (%) X = jumlah sel somatis (10^6sel/ml)

3.2 Metode

Pada penelitian ini, digunakan dua metode analisis yang diterapkan pada masing-masing data. Yang pertama adalah analisis regresi linier tersegmen dengan program *Segreg for Windows* dan yang kedua adalah analisis regresi kuadratik dengan *Software Minitab 13*, sebagai pembanding terhadap model yang diperoleh dari analisis regresi linier tersegmen.

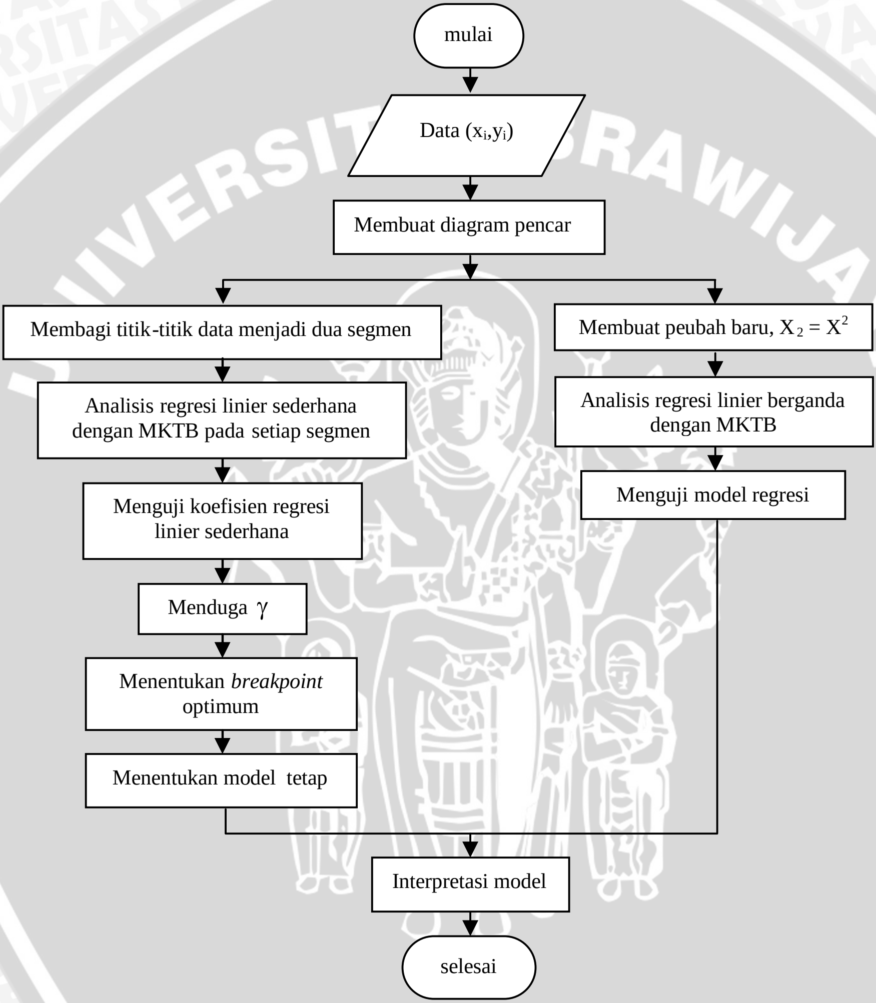
Adapun langkah-langkah metode analisis adalah sebagai berikut:

a. Analisis regresi linier tersegmen
   1. Membuat diagram pencar
   2. Membagi titik-titik data menjadi dua segmen
   3. Melakukan analisis regresi linier sederhana untuk setiap segmen dengan MKTB

4. Menguji koefisien model regresi linier pada setiap segmen
5. Menduga nilai γ dengan rumus $\hat{\gamma} = -\left(\frac{\hat{\alpha}_1 - \hat{\alpha}_2}{\hat{\beta}_1 - \hat{\beta}_2}\right)$
6. Mengulangi langkah 2 – langkah 5 dengan segmen yang berbeda (dengan n_1 dan n_2 yang berbeda pula)
7. Menentukan *breakpoint* optimum berdasarkan SSE terkecil dan R^2 terbesar
8. Menentukan model tetap untuk setiap segmen
9. Interpretasi model

b. Analisis regresi kuadratik
1. Membuat peubah baru X_2, di mana $X_2 = X^2$
2. Melakukan analisis regresi linier berganda dengan MKTB
3. Menguji model regresi
4. Interpretasi model

Tahapan-tahapan untuk metode analisis ditampilkan pada Gambar 3.1.

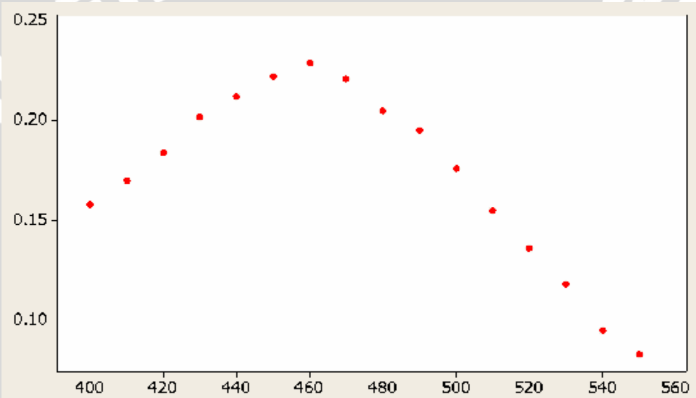


Gambar 3.1. Tahapan metode analisis

BAB IV
HASIL DAN PEMBAHASAN

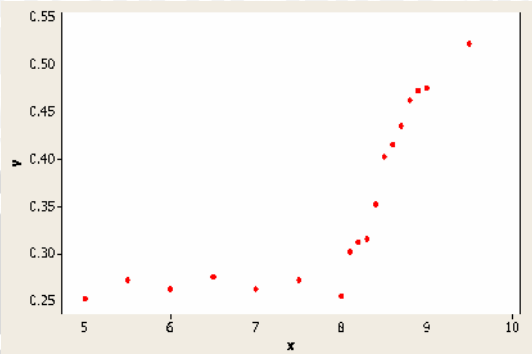
4.1 Diagram Pencar

Pembuatan diagram pencar pada analisis regresi sangat membantu guna mengetahui bagaimana bentuk hubungan antara peubah respon dengan peubah penjelas, apakah bersifat linier atau tidak sehingga dapat memperkirakan model yang sesuai untuk data. Untuk diagram pencar data 1, dapat dilihat pada Gambar 4.1.

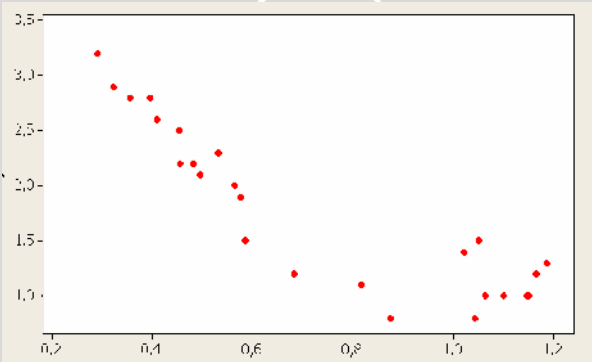


Gambar 4.1 Diagram pencar pada data 1

Berdasarkan Gambar 4.1 terlihat bahwa nilai-nilai peubah respon (Y) semakin meningkat seiring dengan peningkatan nilai X, namun setelah titik tertentu, peningkatan nilai x justru diikuti oleh penurunan nilai Y. Sedangkan dari diagram pencaran data 2 pada Gambar 4.2, tampak bahwa peubah respon memberikan nilai yang relatif sama (konstan) seiring dengan peningkatan nilai x sampai titik tertentu, kemudian semakin meningkat seiring dengan peningkatan nilai x. Pola yang serupa juga ditunjukkan oleh gambar 4.3 yang merupakan diagram pencaran data 3. Akan tetapi, pola diagram pencaran pada data 3 ini diawali dengan penurunan nilai y seiring dengan peningkatan nilai x sampai batas tertentu, dilanjutkan dengan nilai y yang tidak menunjukkan peningkatan ataupun penurunan nilai yang signifikan.



Gambar 4.2   Diagram pencar data 2



Gambar 4.3   Diagram pencar data 3

Dari ketiga diagram pencar tersebut, dapat disimpulkan bahwa baik data 1, 2 maupun data 3 menunjukkan perubahan pola setelah titik tertentu. Di mana, masing-masing data mempunyai dua pola yang berbeda dan masing-masing pola tersebut dapat diwakili dengan sebuah garis linier. Oleh karena itu, analisis regresi linier tersegmen sesuai untuk diterapkan pada data 1, 2 dan 3.

Namun begitu, ketiga data tersebut juga bisa dianalisis dengan analisis regresi polinom, dengan dasar bahwa suatu kurva selalu dapat dihampiri dengan suatu deret polinom. Dalam penelitian ini, model regresi polinom derajat 2 (regresi kuadratik) digunakan sebagai pembanding terhadap model regresi linier tersegmen pada data 1, 2 dan 3.

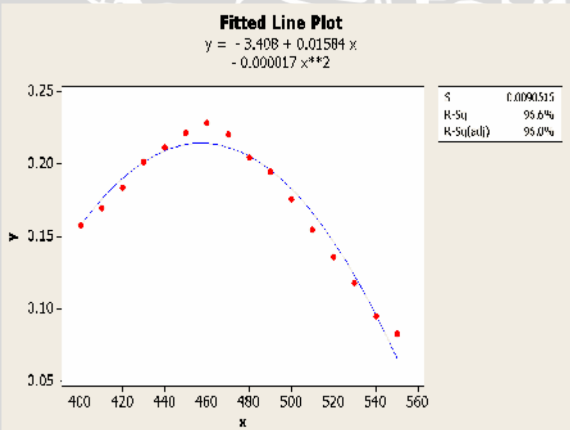
4.2 Analisis regresi kuadratik

Seperti yang telah dijelaskan pada subbab 4.1, pada penelitian ini, model regresi polinom derajat 2 (regresi kuadratik) digunakan sebagai pembanding terhadap model regresi linier tersegmen pada data 1, 2 dan 3. Adapun model regresi kuadratik untuk ketiga data tersebut tertera pada tabel 4.1.

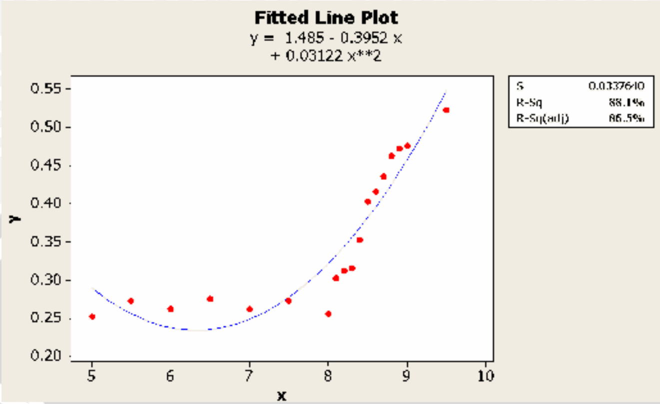
Tabel 4.1 Model regresi kuadratik untuk data 1, 2 dan 3

Data	Model regresi	R^2 *adj*.
1	$y = -3.408 + 0.01584x - 0.000017x^2$	96.0%
2	$y = 1.485 - 0.3952x + 0.03122x^2$	86.5%
3	$y = 5.524 - 9.145x + 4.643x^2$	92.5%

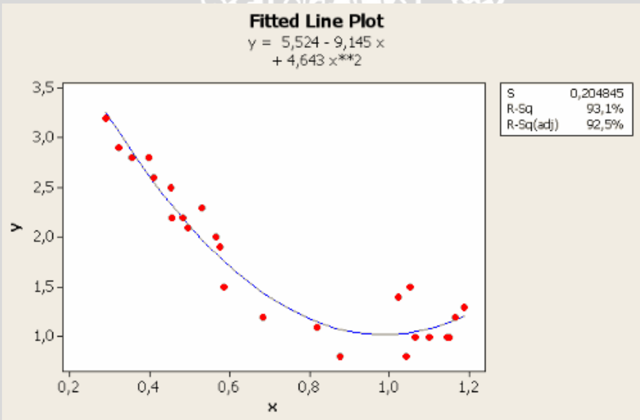
Dari Tabel 4.1 tampak bahwa model regresi kuadratik untuk data 1,2 dan 3 mempunyai nilai R^2 *adjusted* yang cukup besar, yaitu 96.0%, 86.5% dan 92.5%, artinya model regresi kuadratik cukup representatif untuk mewakili data 1, 2 dan 3. Namun tidak tertutup kemungkinan pula bahwa model regresi linier tersegmen akan memberikan R^2 *adjusted* yang lebih besar. Sedangkan bentuk kurva model regresi kuadratik untuk data 1, 2 dan 3 dapat dilihat pada Gambar 4.4, Gambar 4.5 dan Gambar 4.6 dan uji signifikansi untuk model ini dapat dilihat pada Lampiran 7, Lampiran 8 dan Lampiran 9.



Gambar 4.4 Model regresi kuadratik pada data 1



Gambar 4.5 Model regresi kuadratik pada data 2



Gambar 4. 6 Model regresi kuadratik pada dat a 3

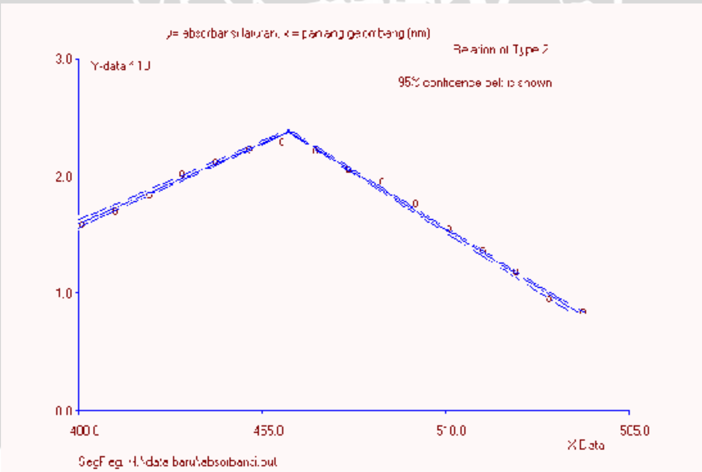
4.3 Analisis regresi linier tersegmen

Hal pokok yang dilakukan pada analisis regresi linier tersegmen adalah menentukan *breakpoint*, yaitu titik yang membagi data menjadi dua bagian, sehingga diperoleh dua submodel masing - masing untuk rentang nilai x y ang berbeda. Hasil analisis regresi linier tersegmen untuk data 1, 2 dan 3 dapat dilihat pada Lampiran 4, Lampiran 5 dan Lampiran 6, sedangkan model yang diperoleh diringkas pada Tabel 4.2.

Tabel 4.2 Model regresi linier tersegmen pada data 1,2 dan 3

Data	Model	R^2 *adj.*
1	y = 0.00125x – 0.342 $x \le 463$ y = -0.00179x + 1.07 $x \ge 463$	98.96%
2	y = 0.267 $x \le 7.92$ y = 0.194x – 1.27 $x \ge 7.92$	94.95%
3	y = -4.83 + 4.58 $x \le 0.72$ y = 1.05 $x \ge 0.72$	93.51%

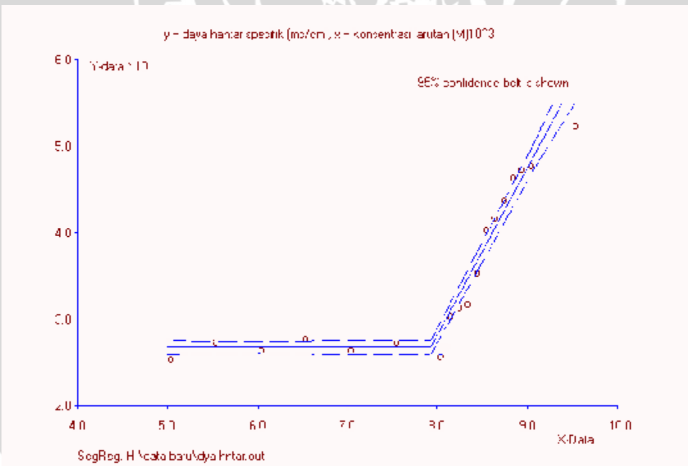
Pada data 1, diperoleh *breakpoint* x = 463, titik ini membagi data menjadi dua bagian. Bagian pertama terdiri dari 7 data pengamatan mulai x = 400 sampai x = 460, sedangkan bagian kedua terdiri dari 9 data pengamatan mulai x = 470 sampai x = 5 50. Submodel pertama mempunyai slope bertanda negatif, yang mengindikasikan bahwa terdapat hubungan negatif antara peubah X dan Y, di mana peningkatan nilai x akan diikuti oleh penurunan nilai y. Sebaliknya, submodel kedua mempunyai slope yang bertanda positif, yang mengindikasikan bahwa terdapat hubungan positif antara peubah X dengan peubah Y, artinya peningkatan nilai x akan diikuti oleh peningkatan nilai y. Model regresi linier tersegmen untuk data 1 dapat dilihat pada Gambar 4.7.



Gambar 4.7. Model regresi linier tersegmen untuk data 1

Berdasarkan hasil analisis pada Lampiran 4, dapat diinterpretasikan bahwa pada panjang gelombang kurang dari 463 nm absorbansi Cl^- meningkat seiring dengan peningkatan panjang gelombang, selanjutnya dengan panjang gelombang yang lebih besar dari 463 nm, absorbansi Cl^- pada larutan NaCl akan semakin menurun. Artinya, titik x = 463 nm merupakan panjang gelombang optimal di mana anion Cl^- menunjukkan absorbansi maksimum.

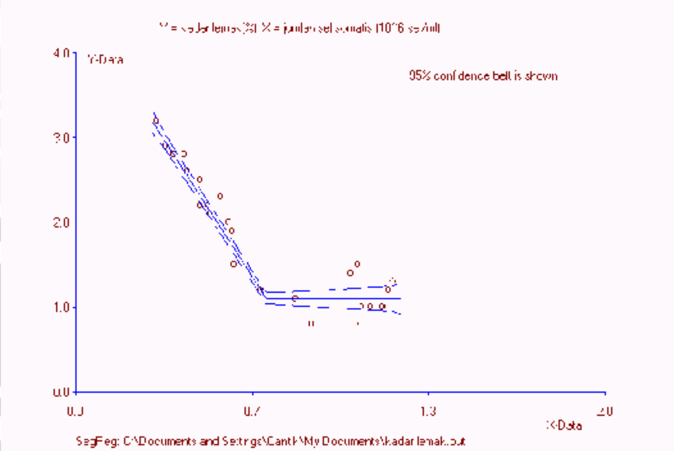
Untuk data 2 diperoleh *breakpoint* x = 7.92. Titik-titik data dengan nilai $x \leq 7.92$ termasuk dalam submodel pertama yang mencakup 6 data pengamatan mulai dari x = 5 sampai x = 7.5. Sedangkan titik data dengan nilai $x \geq 7.92$ termasuk dalam submodel kedua yang mencakup 12 data pengamatan mulai dari x = 8 sampai x = 9.5. Model pada bagian pertama mempunyai slope yang tidak signifikan secara statistik, yang menunjukkan bahwa tidak ada pengaruh linier peubah X terhadap peubah Y, sehingga nilai y relatif konstan berapapun nilai x yang diamati. Oleh karena itu digunakan model konstan $y = \bar{y}$. Sedangkan model pada bagian kedua mempunyai slope yang bertanda positif yang berarti setiap kenaikan x akan diikuti oleh kenaikan y, seperti tampak pada Gambar 4.8. Uji signifikansi pada masing-masing segmen untuk data 1, 2 dan 3 dapat dilihat pada Lampiran 10, Lampiran 11 dan Lampiran 12



Gambar 4.8. Model regresi linier tersegmen untuk data 2

Berdasarkan Gambar 4.8 terlihat bahwa penambahan konsentrasi larutan tidak berpengaruh secara signifi kan terhadap daya hantar spesifik yang dihasilkan. Meskipun konsentrasi larutan ditingkatkan, daya hantar spesifik larutan tidak menunjukkan peningkatan ataupun penurunan yang berarti. Akan tetapi, keadaan ini hanya terjadi pada peningkatan konsentrasi lar utan sampai 7.92 M(10^{-3}). Selanjutnya, peningkatan konsentrasi larutan lebih dari 7.92 M(10^{-3}) akan semakin meningkatkan daya hantar spesifik larutan $NaC_{12}H_{25}SO_4$. Dengan demikian, titik x = 7.92 (M)10^{-3} merupakan konsentrasi kritis yang menunjukkan mulai t erjadinya peningkatan daya hantar spesifik larutan $NaC_{12}H_{25}SO_4$. Sesuai dengan literatur, pada konsentrasi rendah, surfaktan (dalam hal ini adalah larutan $NaC_{12}H_{25}SO_4$) mempunyai sifat seperti elektrolit. Sedang pada konsentrasi tinggi beberapa sifat fisik s urfaktan (di antaranya adalah daya hantar spesifik) mengalami perubahan. Hal ini terjadi karena surfaktan membentuk misel. Konsentrasi pada saat surfaktan membentuk misel inilah yang dikenal sebagai konsentrasi kritis misel (KKM).

Sama seperti data 2, data 3 mempunyai dua submodel, di mana salah satu modelnya merupakan model konstan, y = $\bar{y}$, dan kedua submodel ini dibatasi oleh *breakpoint* x = 0.729. Bagian pertama terdiri dari 14 data pengamatan mencakup nilai x = 0.290×10^6sel/ml sampai x = 0.683×10^6sel/ml. Sedangkan bagian kedua terdiri dari 11 data pengamatan yang mencakup nilai x = 0.817 $\times 10^6$sel/ml sampai x = 1.186 $\times 10^6$sel/ml. Model pada bagian pertama mempunyai slope yang bertanda negatif, artinya apabila nilai x semakin meningkat, maka nilai y akan semakin menurun. Untuk bagian kedua digunakan model konstan dikarenakan slope model regresi pada bagian kedua tidak signifikan secara statistik. Adapun model regresi linier tersegmen pada data 3 dapat dilihat pada Gambar 4.9.



Gambar 4.9. Model regresi linier tersegmen untuk data 3

Berdasarkan Gambar 4.9, terlihat bahwa kadar lemak susu semakin menurun seiring dengan jumlah sel somatis yang semakin meningkat. Namun begitu, pada saat jumlah sel somatis melebihi 0.72×10^6 sel /ml kadar lemak yang terkandung dalam susu dapat dikatakan relatif konstan, berkisar pada angka 1.05%. Penurunan kandungan lemak dalam susu ini dikarenakan kemampuan kelenjar susu untuk memproduksi lemak susu menurun. Berdasarkan literatur, peningkatan jumlah sel somatis dari 0.083×10^6sel/ml sampai 0.87×10^6sel/ml akan berakibat konsentrasi lemak dalam susu berkurang karena penurunan sintesis lemak oleh sel epitel dari kelenjar ambing.

Ditinjau dari kesesuaian model, secara keseluruhan dapat diambil kesimpulan bahwa model regresi linier tersegmen memberikan nilai koefisien determinasi yang lebih besar dibandingkan model regresi kuadratik. Pada data 1, model regresi linier tersegmen memberikan R^2 *adjusted* lebih besar 2.96% dari R^2 *adjusted* yang diberikan oleh model regresi kuadratik, yaitu 98.96%. Untuk data 2, penggunaan model regresi linier tersegmen meningkatkan kesesuaian model sebesar 8.45% dibanding model regresi kuadratik. Sedangkan pada data 3, kesesuaian model menjadi lebih besar, yaitu 93.51% setelah digunakan model regresi linier tersegmen. Hal ini berarti bahwa model regresi linier tersegmen lebih representatif terhadap data dan mampu menjelaskan keragaman data lebih besar dari keragaman yang bisa dijelaskan oleh model regresi kuadratik, di mana kurva model regresi yang berbentuk kuadratik didekati dengan rangkaian dua garis linier.

Akan tetapi, perlu diperhatikan bahwa dalam menentukan model terbaik untuk data perlu mempertimbangkan beberapa hal, salah satunya adalah dasar teori pada data yang bersangkutan. Kalau memang secara teori menunjukkan bahwa data membentuk suatu pola kuadratik, misalnya, maka model regresi kuadratiklah yang memang lebih sesuai untuk data, meskipun model yang lain memberikan R^2 yang lebih besar.

BAB V
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil dan pembahasan pada bab IV, disimpulkan bahwa model yang diperoleh dari analisis regresi linier tersegmen memberikan nilai koefisien determinasi lebih besar dibandingkan model regresi polinom berderajat 2 (regresi kuadratik), yang berarti bahwa model regresi linier tersegmen mampu menjelaskan keragaman data lebih besar dari keragaman yang bisa dijelaskan oleh model regresi kuadratik.

5.2 Saran

Dari penelitian ini, disarankan untuk mempelajari penerapan analisis regresi linier tersegmen pada sekelompok data dengan banyak *breakpoint* lebih dari satu, mempelajari analisis regresi linier tersegmen pada data yang tidak memenuhi asumsi analisis regresi dan mengkaji analisis regresi tersegmen yang mempunyai segmen nonlinier.

DAFTAR PUSTAKA

Anonymous. 2006. *Segmented Regression*. http://en.wikipedia.org/wiki/Segmented_regression tanggal akses: 15 Agustus 2006

Ayu, N.L.P.N.A. 2003, *Pengaruh Fosfatasi pada Zeolit Alam Turen Terhadap Absorpsi Kation* Fe^{3+} *dan Anion* Cl^-. Jurusan Kimia. Fakultas Matematika dan Ilmu Pengetahuan Alam. Universitas Brawijaya. Tidak dipublikasikan.

Draper, N.R. dan H. Smith. 1992. *Analisis Regresi Terapan.* Edisi Kedua. Terjemah B. Sumantri. Gramedia Pustaka Utama. Jakarta

Diniz, C.A.R and L.C. Brochi. 2005. *Robustness of Two-Phase Regression Tests*. REVSTAT-Statistical Journal 3:3 http://www.ine.pt/revstat/pdf/rs050101.pdf tanggal akses: 01 Januari 2008

Gujarati, D.N. 1988. *Basic Econometrics*. Mc Graw Hill Book Company. USA

Guntoro, H. 2001. *Kajian Awal Pengaruh pH Fasa Eksternal Terhadap Pemisahan Kation Fe(III) Dengan Menggunakan Membran Misel Balik*. Jurusan Kimia. Fakultas Matematika dan Ilmu Pengetahuan Alam. Universitas Brawijaya. Tidak dipublikasikan.

Oosterbaan, R.J., D.P. Sharma and K.N. Singh. 1990. *Crop production and soil salinity:Evaluation of field data from India by segmented linear regression*. Symposium on Land Drainage for Salinity Control in Arid and Semi-Arid Regions.Vol.3.Cairo http://waterlog.info/pdf/segmregr.pdf tanggal akses: 25 November 2007

Riduwan. 2006. *Dasar-dasar Statistika*. Alfabeta. Bandung

Ryan, S.E. and L.S.Porth. 2007. *A Tutorial on the Piecewise Regression Approach Applied to Bedload Transport Data*. General Technical Report RMRS-GTR-189 http://www.fs.fed.us/rm/pubs/rmrs_gtr189.pdf tanggal akses: 09 November 2007

Safitri, A.Y. 2006. *Hubungan Antara Jumlah Sel Somatis Dengan Kadar Lemak Dan Profil Asam Susu*. Fakultas Peternakan. Universitas Brawijaya. Tidak dipublikasikan.

Seber, G.A.F.1977. *Linear Regression Analysis*. John Wiley and Sons.New York

Sembiring, R.K..1995. *Analisis Regresi*. ITB. Bandung

Yitnosumarto, S. 1985. *Regresi dan Korelasi: Teori dan Penggunaannya*. Universitas Brawijaya. Malang. (tidak dipublikasikan)

Lampiran 1. Nilai absorbansi cl^- 8 ppm pada berbagai panjang gelombang

y = absorbansi larutan	x = panjang gelombang (nm)
0.158	400
0.170	410
0.184	420
0.202	430
0.212	440
0.222	450
0.229	460
0.221	470
0.205	480
0.195	490
0.176	500
0.155	510
0.136	520
0.118	530
0.095	540
0.083	550

Lampiran 2. Daya hantar spesifik larutan $NaC_{12}H_{25}SO_4$ dalam berbagai konsentrasi dengan konduktor

y= daya hantar spesifik (ms/cm)	x= konsentrasi $NaC_{12}H_{25}SO_4$ (M)10^{-3}
0.253	5.0
0.273	5.5
0.263	6.0
0.276	6.5
0.263	7.0
0.273	7.5
0.256	8.0
0.303	8.1
0.313	8.2
0.316	8.3
0.353	8.4
0.403	8.5
0.416	8.6
0.436	8.7
0.463	8.8
0.473	8.9
0.476	9.0
0.523	9.5

Lampiran 3. Jumlah sel somatis dan kadar lemak susu

Y = kadar lemak(%)	X = jumlah sel somatis (10^6 sel/ml)
3.2	0.29039
2.9	0.32311
2.8	0.35583
2.8	0.39673
2.6	0.40899
2.5	0.45399
2.2	0.45499
2.2	0.48262
2.1	0.49489
2.3	0.53169
2	0.56442
1.9	0.57669
1.5	0.58487
1.2	0.68304
1.1	0.81799
0.8	0.87526
1.5	1.05113
1.4	1.02249
1.3	1.18609
1.2	1.16564
1	1.06339
1	1.10020
1	1.14928
1	1.14519
0.8	1.04295

Lampiran 4. Analisis regresi linier tersegmen pada data 1

Results of program SEGREG for the segmented linear regression of Y upon X.

X is the independent variable.
There can be different types of functions ranging from type 0 to type 6 and (for types 2, 3 and 4) two methods of calculation.
The types and methods are determined with the procedure of best fit.
For explanations, use the symbols function in the output scroll menu.

Name of this output file: H:\data baru\absorbansi.var
Name of input file used : H:\data baru\absorbansi.inp

y= absorbansi larutan, x = panjang gelombang (nm)
Minimum confidence % : 95

Regression of Y upon X without breakpoint (BPx=Xmin).
Only one independent variable (X) is used.
The table below gives the following series of values respectively:

Breakpoint(BPx)	number of data	Av.Y	Av.X
Regr.Coeff.(RC)	Corr.Coeff.Sq.	St.Dev.RC	Y(X=0)
St.Dev.Y	St.Dev.Yr	St.Dev.X	
BPx= 400.00	1.60E+001	1.73E-001	4.75E+002
-6.13E-004	4.12E-001	1.96E-004	4.64E-001
4.54E-002	3.48E-002	4.76E+001	

Results of regression of Y upon X with optimal breakpoint (BPx).
The table below gives the following series of values respectively:

Breakpoint(BPx)	number of data	Av.Y	Av.Z
Regr.Coeff.(RC)	Corr.Coeff.Sq.	St.Dev.RC	Y(X=0)
St.Dev.Y	St.Dev.Yr	St.Dev.X	

for the data with X-values smaller and greater than BPx followed by the function parameters.

Data with X < BPx :

BPx= 463.00	7.00E+000	1.97E-001	4.30E+002
1.23E-003	9.85E-001	6.89E-005	-3.33E-001
2.68E-002	3.33E-003	2.16E+001	

Data with X > BPx :

BPx= 463.00	9.00E+000	1.54E-001	5.10E+002
-1.79E-003	9.95E-001	4.91E-005	1.07E+000
4.92E-002	3.56E-003	2.74E+001	

Lampiran 4. (Lanjutan)

Parameters for function type 2 and method

Slope < BPx	St.Err.Slope	method
1.25E-003	4.16E-005	2
Y at BPx	Expl.Coeff.	
2.38E-001	9.95E-001	

SUMMARY OF THE Y-X REGRESSION.
Function type : 2 - two segments with sloping lines.
See help functions on Intro tabsheet

Optimal breakpoint of X (BPx) : 4.63E+002

There are two regression equations:

when X is smaller than BPx: Y = AsX + Cs
when X is greater than BPx: Y = AgX + Cg

As = 1.25E-003	Ag = -1.79E-003
Cs = -3.42E-001	Cg = 1.07E+000

Lampiran 5. Analisis regresi linier tersegmen pada data 2

Results of program SEGREG for the segmented linear regression of Y upon X.

X is the independent variable.
There can be different types of functions ranging from type 0 to type 6 and (for types 2, 3 and 4) two methods of calculation.
The types and methods are determined with the procedure of best fit.
For explanations, use the symbols function in the output scroll men u.

Name of this output file: H:\data baru\dya hntar.var
Name of input file used : H:\data baru\dya hntar.inp

y = daya hantar spesifik (ms/cm), x = konsentrasi larutan (M)10^ -3
Minimum confidence % : 95

Regression of Y upon X without breakpoint (BP x=Xmin).
Only one independent variable (X) is used.
The table below gives the following series of values respectively:

Breakpoint(BPx)	number of data	Av.Y	Av.X
Regr.Coeff.(RC)	Corr.Coeff.Sq.	St.Dev.RC	Y(X=0)
St.Dev.Y	St.Dev.Yr	St.Dev.X	
BPx= 5.00	1.80E+001	3.52E-001	7.81E+000
5.65E-002	6.26E-001	1.09E-002	-8.94E-002
9.19E-002	5.62E-002	1.29E+000	

Results of regression of Y upon X with optimal breakpoint (BPx).
The table below gives the following series of values respectively:

Breakpoint(BPx)	number of data	Av.Y	Av.Z
Regr.Coeff.(RC)	Corr.Coeff.Sq.	St.Dev.RC	Y(X=0)
St.Dev.Y	St.Dev.Yr	St.Dev.X	

for the data with X-values smaller and greater than BPx followed by the function parameters.

Data with X < BPx :

BPx= 7.92	6.00E+000	2.67E-001	6.25E+000
4.74E-003	2.58E-001	4.02E-003	2.37E-001
8.73E-003	7.52E-003	9.35E-001	

Data with X > BPx :

BPx= 7.92	1.20E+001	3.94E-001	8.58E+000
1.89E-001	9.24E-001	1.72E-002	-1.23E+000
8.44E-002	2.33E-002	4.28E-001	

Lampiran 5. (Lanjutan)

Parameters for function type 3 and method

Slope < BPx	St.Err.Slope	method
0.00E+000	2.33E-003	2
Y at BPx	Expl.coeff.	
2.67E-001	9.56E-001	

SUMMARY OF THE Y-X REGRESSION.

Function type : 3 - first a horizontal segment, then sloping.

See help functions on Intro tabsheet

Optimal breakpoint of X (BPx) : 7.92E+000

There are two regression equations:

when X is smaller than BPx: Y = AsX + Cs

when X is greater than BPx: Y = AgX + Cg

As = 0.00E+000 Ag = 1.94E-001

Cs = 2.67E-001 Cg = -1.27E+000

Lampiran 6. Analisis regresi linier tersegmen pada data 3

Results of program SEGREG for the segmented linear regression of Y upon X.

X is the independent variable.

There can be different types of functions ranging from type 0 to type 6 and (for types 2, 3 and 4) two methods of calculation.

The types and methods are determined with th e procedure of best fit.

For explanations, use the symbols function in the output scroll menu.

Name of this output file: C:\ My Documents\kadar lemak.var

Name of input file used : C:\ My Documents\kadar lemak.inp

Y = kadar lemak(%), X = jumlah sel s omatis (10^6 sel/ml)

Minimum confidence % : 95

Regression of Y upon X without breakpoint (BPx=Xmin).

Only one independent variable (X) is used.

The table below gives the following series of values respectively:

Breakpoint(BPx)	number of data	Av.Y	Av.X
Regr.Coeff.(RC)	Corr.Coeff.Sq.	St.Dev.RC	Y(X=0)
St.Dev.Y	St.Dev.Yr	St.Dev.X	
BPx= 0.29	2.50E+001	1.77E+000	7.29E-001
-2.09E+000	7.88E-001	2.26E-001	3.30E+000
7.46E-001	3.43E-001	3.17E-001	

Results of regression of Y upon X with optimal breakpoint (BPx).

The table below gives the following series of values respectively:

Breakpoint(BPx)	number of data	Av.Y	Av.Z
Regr.Coeff.(RC)	Corr.Coeff.Sq.	St.Dev.RC	Y(X=0)
St.Dev.Y	St.Dev.Yr	St.Dev.X	

for the data with X-values smaller and greater than BPx followed by the function parameters.

Data with X < BPx :

BPx= 0.72	1.40E+001	2.30E+000	4.72E-001
-4.80E+000	9.29E-001	3.82E-001	4.56E+000
5.52E-001	1.47E-001	1.11E-001	

Data with X > BPx :

BPx= 0.72	1.10E+001	1.10E+000	1.06E+000
4.32E-001	4.96E-002	6.30E-001	6.44E-001
2.28E-001	2.22E-001	1.18E-001	

Lampiran 6. (Lanjutan)

Parameters for function type 4 and method

Slope > BPx	St.Err.Slope	method
0.00E+000	4.04E-001	2
Y at BPx	Expl.Coeff.	
1.10E+000	9.40E-001	

SUMMARY OF THE Y-X REGRESSION.
Function type : 4 - first a sloping segment, then horizontal.
See help functions on Intro tabsheet

Optimal breakpoint of X (BPx) : 7.20E-001

There are two regression equations:

when X is smaller than BPx: Y = AsX + Cs
when X is greater than BPx: Y = AgX + Cg

As = -4.82E+000 Ag = 0.00E+000
Cs = 4.58E+000 Cg = 1.10E+000

Lampiran 7. Analisis regresi kuadratik pada data 1

Polynomial Regression Analysis: y versus x

The regression equation is
y = - 3.408 + 0.01584 x - 0.000017 x**2

S = 0.00905150   R-Sq = 96.6%   R-Sq(adj) = 96.0%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	0.0299089	0.0149544	182.53	0.000
Error	13	0.0010651	0.0000819		
Total	15	0.0309739			

Sequential Analysis of Variance

Source	DF	SS	F	P
Linear	1	0.0127676	9.82	0.007
Quadratic	1	0.0171413	209.22	0.000

Lampiran 8. Analisis regresi kuadratik pada data 2

Polynomial Regression Analysis: y versus x

The regression equation is
y = 1.485 - 0.3952 x + 0.03122 x**2

S = 0.0337640   R-Sq = 88.1%   R-Sq(adj) = 86.5%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	0.126567	0.0632835	55.51	0.000
Error	15	0.017100	0.0011400		
Total	17	0.143667			

Sequential Analysis of Variance

Source	DF	SS	F	P
Linear	1	0.0899746	26.81	0.000
Quadratic	1	0.0365924	32.10	0.000

Lampiran 9. Analisis regresi kuadratik pada data 3

Polynomial Regression Analysis: y versus x

```
The regression equation is
y = 5.524 - 9.145 x + 4.643 x**2


S = 0.204845   R-Sq = 93.1%   R-Sq(adj) = 92.5%


Analysis of Variance

Source       DF        SS        MS       F      P
Regression    2   12.4273   6.21363  148.08  0.000
Error        22    0.9231   0.04196
Total        24   13.3504


Sequential Analysis of Variance

Source     DF       SS      F      P
Linear      1  10.5216  85.55  0.000
Quadratic   1   1.9057  45.42  0.000
```

Lampiran 10. Uji signifikansi model untuk setiap segmen pada data 1

- Uji signifikansi pada segmen pertama $(x \leq 463)$

Regression Analysis: y1 versus x1

```
The regression equation is
y1 = - 0.333 + 0.00123 x1

Predictor         Coef     SE Coef       T      P
Constant      -0.33311     0.02968  -11.22  0.000
x1          0.00123214  0.00006894   17.87  0.000

S = 0.00364790   R-Sq = 98.5%   R-Sq(adj) = 98.2%

Analysis of Variance

Source          DF         SS         MS       F      P
Regression       1  0.0042509  0.0042509  319.44  0.000
Residual Error   5  0.0000665  0.0000133
Total            6  0.0043174
```

- Uji signifikansi pada segmen kedua $(x \geq 463)$

Regression Analysis: y2 versus x2

```
The regression equation is
y2 = 1.07 - 0.00179 x2

Predictor          Coef     SE Coef       T      P
Constant        1.06838     0.02508   42.60  0.000
x2          -0.00179333  0.00004911  -36.52  0.000

S = 0.00380392   R-Sq = 99.5%   R-Sq(adj) = 99.4%

Analysis of Variance

Source          DF        SS        MS        F      P
Regression       1  0.019296  0.019296  1333.55  0.000
Residual Error   7  0.000101  0.000014
Total            8  0.019398
```

Lampiran 11. Uji signifikansi model untuk setiap segmen pada data 2

- Uji signifikansi pada segmen pertama ($x \leq 7.92$)

Regression Analysis: y1 versus x1

```
The regression equation is
y1 = 0.237 + 0.00474 x1

Predictor       Coef   SE Coef     T      P
Constant     0.23719   0.02534  9.36  0.001
x1          0.004743  0.004017  1.18  0.303

S = 0.00840266   R-Sq = 25.8%   R-Sq(adj) = 7.3%

Analysis of Variance

Source          DF          SS          MS     F      P
Regression       1  0.00009841  0.00009841  1.39  0.303
Residual Error   4  0.00028242  0.00007060
Total            5  0.00038083
```

- Uji signifikansi pada segmen kedua ($x \geq 7.92$)

Regression Analysis: y2 versus x2

```
The regression equation is
y2 = - 1.23 + 0.189 x2


Predictor      Coef  SE Coef      T      P
Constant    -1.2318   0.1478  -8.33  0.000
x2          0.18945  0.01720  11.01  0.000


S = 0.0244301   R-Sq = 92.4%   R-Sq(adj) = 91.6%


Analysis of Variance

Source          DF        SS        MS       F      P
Regression       1  0.072378  0.072378  121.27  0.000
Residual Error  10  0.005968  0.000597
Total           11  0.078346
```

Lampiran 12. Uji signifikansi model untuk setiap segmen pada data 3

- Uji signifikansi pada segmen pertama $(x \leq 0.72)$

Regression Analysis: y1 versus x1

```
The regression equation is
y1 = 4.56 - 4.80 x1

Predictor     Coef  SE Coef       T      P
Constant    4.5625   0.1849   24.68  0.000
x1         -4.7976   0.3823  -12.55  0.000

S = 0.152857   R-Sq = 92.9%   R-Sq(adj) = 92.3%

Analysis of Variance

Source          DF      SS      MS       F      P
Regression       1  3.6796  3.6796  157.48  0.000
Residual Error  12  0.2804  0.0234
Total           13  3.9600
```

- Uji signifikansi pada segmen kedua $(x \geq 0.72)$

Regression Analysis: y2 versus x2

```
The regression equation is
y2 = 0.644 + 0.432 x2

Predictor    Coef  SE Coef     T      P
Constant   0.6437   0.6695  0.96  0.361
x2         0.4319   0.6302  0.69  0.510

S = 0.234333   R-Sq = 5.0%   R-Sq(adj) = 0.0%

Analysis of Variance

Source          DF       SS       MS     F      P
Regression       1  0.02579  0.02579  0.47  0.510
Residual Error   9  0.49421  0.05491
Total           10  0.52000
```