Jurnal Peluang, Volume 1, No. 1, Oktober 2012, ISSN: 2302-5158

Mendapatkan Algoritma Menghitung Koefisien-koefisien Persamaan Regresi Linear Dua Variabel untuk Dikoding ke dalam Bahasa Pemrograman

Anwar[1]

[1]Dosen Program Studi Pendidikan Matematika FKIP Unsyiah

Abstract

Data in research usually are provided in discrete value. Many researchers want to predict the value of the discrete values. The strategy is to find the simple function that appropriates of the value. One of the ways that is used is least square method that is linear regression. This paper aims to describe to readers about how to find the algorithm of counting the coefficient a and b of linear regression y = a + bx for coding to programming language.

Keynotes*: linear regression, algorithm, programming language.*

Pendahuluan

Ketika melakukan pengolahan data penelitian, parapeneliti pemula dan mahasiswa kerap kali ingin menentukan hubungan linear sederhana antara dua variabel data, misalnya X dan Y. Hubungan ini pada umumnya dinyatakan oleh suatu persamaan matematika yang menyatakan hubungan fungsional kedua variabel tersebut, yang dikenal sebagai persamaan regresi dengan bentuk y = a +bx. Koefisien a dan b sangat menentukan makna hubungan antara dua variabel dari persamaan regresi tersebut. Koefisien b menyatakan perubahan rata-rata variavel y untuk setiap perubahan variabel x sebesar satu unit. Perubahan ini merupakan pertambahan bila b bernilai positif dan perurunan atau pengurangan bila bernilai negatif (Sujana, 1992: 318).

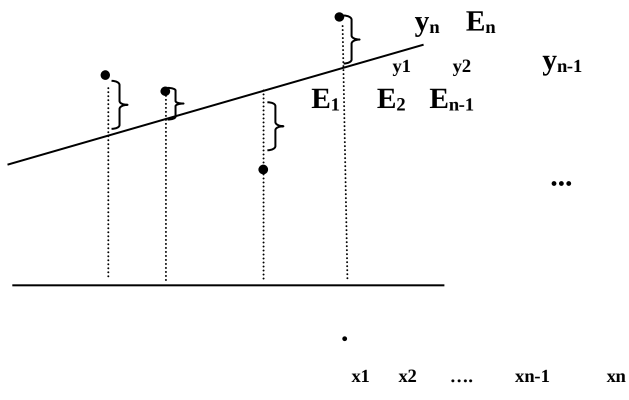
Untuk menghitung koefisien a dan b dari persamaan regresi di atas, parapeneliti dan mahasiswa pada umumnya melakukan perhitungan menggunakan kalkulator atau perhitungan tangan biasa berdasarkan formula yang terdapat dalam buku –buku penelitian dan statistika, atau memanfaatkan software statistika yang ada di pasaran sekarang ini. Menurut pengamatan penulis, sedikit sekali mereka yang mengetahui dengan persis dari mana formula a dan b itu diperoleh bahkan sangat sedikit pula yang mau dan dapat membuat program sendiri. Padahal, bila kita mau memanfaatkan komputer pribadi yang kita miliki, kita dapat mengolah data penelitian meskipun kita tidak mempunyai software-sofware pengolah data yang canggih. Untuk itu, secara sangat sederhana bahasan ini akan menjelaskan kepada pembaca tentang bagaimana mendapatkan formula perhitungan koefisien a dan b dari persamaan regresi y = a + bx, mendapatkan algoritmenya dan menerjemahkannya (koding) ke dalam suatu bahasa pemrograman, dalam hal ini PASCAL, QBASIC atau JUSTBASIC. Melalui bahasan ini pula diharapkan agar pembaca memiliki pengetahuan tambahan mendapatkan formula koefsien-koefisien regresi a dan b serta memiliki program sendiri yang dapat dimodifikasi untuk melakukan perhitungan berdasarkan algoritme yang tersedia.

Regresi Linear Sederhana

Data suatu penelitian sering kali diberikan dalam bentuk nilai diskret dan akan ditaksir pada titik-titik di antara nilai-nilai diskret itu. Untuk keperluan ini, terdapat beberapa pendekatan umum yang berbeda antara satu dengan yang lain. Salah satu di antaranya adalah regresi linear. Stateginya adalah menurunkan fungsi yang sederhana yang mencocoki nilai-nilai itu atau menurunkan kurva tunggal yang menyatakan kecenderungan umum data. Kurva didesain untuk mengikuti pola titik - titik tersebut yang diambil dari sebagian atau sekelompok data. Salah satu pendekatannya adalah dengan metode kuadrat terkecil.

Secara sederhana, masalah kita adalah mencocokkan garis lurus pada himpunan pasangan pengamatan (x_1. y_1), (x_2, y_2), …(x_n, y_n) seperti pada gambar 1. Ekepresi matematis untuk garis lurus yang menghampiri seluruh titik-titik data tersebut adalah Y = a + bx + E, dimana E adalah galat atau sisa antara model dan pengamatan. Kriteria untuk pencocokan terbaik yang kita lakukan terhadap data di atas adalah meminimumkan galat, yakni dengan metode kuadrat terkecil. Dalam hal ini misalkan D adalah jumlah kuadrat galat. Maka D dapat dinyatakan sebagai

$$D = \sum_{i=1}^{n} E_i^2 = \sum_{i=1}^{n} (y_i - a_o - a_1 x_i)^2$$



Gambar 1.

Untuk selanjutnya, untuk penyingkatan maka sumasi pada notasi sigma tidak dituliskan lagi. Perhatikan bahwa fungsi D minimum bila turunan parsial tingkat pertama terhadap a dan b sama dengan nol, yakni $\frac{\partial D}{\partial a} = 0$ dan $\frac{\partial D}{\partial b} = 0$. Untuk kasus pertama, bila $\frac{\partial D}{\partial a} = 0$ maka

$$2\sum (y_i - a - bx_i)(-1) = 0 \text{ atau}$$

$$\sum y_i - \sum a - \sum bx_i = 0 \quad \text{atau}$$
$$\sum y_i - na - (\sum x_i)b = 0 \quad \text{atau}$$
$$na + (\sum x_i)b = \sum Y_i \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(1)$$

Untuk kasus ke dua, bila $\frac{\partial D}{\partial b} = 0$ maka

$$2\sum (y_i - a - bx_i)(-x_i) = 0$$

$$\sum x_i y_i - \sum a x_i - \sum b x_i^2 = 0 \quad \text{atau}$$
$$(\sum x_i)a + (\sum x_i^2)\, b = \sum x_i y_i \quad \ldots\ldots(2)$$

Dengan menggunakan determinan untuk menyelesaikan system

$$\left.\begin{aligned} na - (\sum x_i)b &= \sum Y_i \\ (\sum x_i)a - (\sum x_i^2)\, b &= \sum x_i y_i \end{aligned}\right\}$$

diperoleh

$$b = \frac{\begin{vmatrix} n & \sum y_i \\ \sum x_i & \sum x_i y_i \end{vmatrix}}{\begin{vmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{vmatrix}} \text{ atau}$$

$$b = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2} \ldots\ldots\ldots\ldots(3)$$

dan

$$a = \frac{1}{n}(\sum y_i - b\sum x_i)\ldots\ldots\ldots\ldots\ldots\ldots(4)$$

Formula (3) dan (4) inilah yang sering digunakan parapeneliti dan mahasiswa untuk mendapatkan koefisien a dan b dari persamaan regresi y = a + bx, seperti yang terdapat dalam kebanyakan buku statistika dan penelitian. Akhirnya, kita dapat menghitung besarnya nilai taksiran y, dengan simbol (Yt) berdasarkan nilai taksiran X, dengan simbol (Xt) berdasarkan persamaan Yt = a + b Xt.

Untuk membantu pemahaman pembaca terhadap penggunakan formula di atas, berikut diberikan sebuah contoh.

Contoh:
Misalkan seorang peneliti mencatat hasil pengamatan mengenai banyaknya orang yang datang (X) dan banyaknya orang yang berbelanja (Y) pada sebuah toko selama 30 hari. Ingin diketahui hubungan fungsional secara linear ke dua variabel tersebut dan selanjutnya akan diramal banyaknya pembeli berdasarkan banyaknya orang yang datang ke toko tersebut. Adapun datanya sebagai berikut:

X_i	34	38	34	40	30	40	40	34	35	39	33	32	42	42	41	32	34	36	37	36	37	36	37	39	40	33	34	36	37	38
Y_i	32	36	31	38	29	35	33	30	32	36	31	31	36	37	35	38	37	30	30	30	33	32	34	35	36	32	32	34	32	34

Secara manual biasa atau menggunakan kalkulator, berdasarkani data di atas diperoleh $\sum x_i$ = 1105, $\sum Y_i$ = 1001, $\sum x_i^2$ = 41029, $\sum x_i y_i$ = 37094. Berdasarkan nilai-nilai tersebut, hasil perhitungan koefisien a dan b dengan menggunakan formula (3) dan (4) adalah

a = (1101-0,68(1105))/30 = 8,24
b = {30(37094) – 1105(1001)}/{30(41029) – (1105x1105)} = 0.68,
sehingga persamaan regresinya adalah Y = 8,24 + 0.68X. Menurut Sujana (1992:319), hal ini berarti untuk setiap X (pengunjung) bertambah dengan seorang maka rata-rata Y (pembeli) bertambah 0,68 orang.

Bila kita ingin meramal berapa banyak pembeli, misalnya dari Xt=30 pengunjung ke toko, maka ramalan untuk Y adalah
Yt = 8,24 + 0.68 (30) = 28,6.

Hal ini berarti diperkirakan rata-rata ada 28,6 orang pembeli untuk setiap 30 orang pengunjung yang masuk ke toko itu.

Perhitungan dengan cara yang telah dijelaskan di atas seperti menggunakan kalkulator atau perhitungan biasa akan sangat membosankan bila datanya cukup besar. Oleh karena itu, kita dapat melakukan perhitungan tersebut dengan membuat suatu program sederhana. Sebelum masalah ini dibahas lebih lanjut, perlu dijelaskan terlebih dahulu bagaimana mendapatkan algoritme perhitungan koefisien a dan b.

Algoritme

Menurut Liu (1993:3), suatu algoritme adalah suatu prosedur langkah demi langkah tentang bagaimana mengerjakan suatu tugas tertentu. Secara khusus, suatu algoritme adalah uraian lengkap dari operasi-operasi hitungan dan logika secara sistematis yang mentransformasikan himpunan data masukan (input) hingga diperoleh suatu keluaran (output) dalam sejumlah berhingga langkah. Ditambahkan oleh Johnsonbaugh (1997:135), ciri-cirinya adalah berbentuk seumum mungkin, jelas dan lengkap, minimal menghasilkan satu keluaran, hasil akhir (output) tidak bergantung pada siapa yang melaksanakan algoritme. Ia dibuat dengan mengikuti suatu pengkodean tertentu yang dikenal dengan istilah *Kode Pseudo* yang melibatkan *masukan, keluaran dan langkah-langkah (proses).* “kode Pseudo” adalah kalimat dengan kata-kata yang sudah mempunyai makna tertentu. Menurut (Susila, 1993:3), kerangka algoritme adalah

Masukan : ---------------

Keluaran : ---------------

Langkah-langkah : ------------

Karena dalam perhitungan a dan b banyak melibatkan perhitungan deret seperti $\sum x_i$, $\sum Y_i$, $\sum x_i^2$ dan $\sum x_i y_i$, maka sebagai contoh dijelaskan sedikit tentang bagaimana mendapatkan algoritme perhitungan $D = \sum_{i=1}^{\infty} a_i = a_1 + a_2 + \ldots$.

Karena kita bekerja dengan numerik, perhitungan D di atas ditulis sebagai

$$D = \sum_{i=1}^{n} a_i = a_1 + a_2 + \ldots + a_n \ .$$

Prosesnya:

D = 0

Jumlah hingga 1 suku $D = a_1 = D + a_1$, yakni jumlah nilai D yang sebelumnya dengan a_1

Jumlah hingga 2 suku $D = a_1 + a_2 = D + a_2$, yakni jumlah nilai D yang sebelumnya dengan a_2

Jumlah hingga 3 suku $D = a_1 + a_2 + a_3 = D + a_3$, jumlah nilai D yang sebelumnya dengan a_3

....

Jumlah hingga n suku $D = a_1 + a_2 + a_3 + \ldots + a_n = D + a_n$

Berdasarkan proses beberapa langkah di atas dapat diperkatikan bahwa untuk langkah ke-i, i dari 1 sampai n, formula umumnya adalah D = D + a_i, sementara nilai D yang pertama bernilai 0, yakni D= 0. Berdasarkan formula umum tersebut disusun algoritme perhitungan D seperti berikut.

Masukan : N (banyak data)

a_i ; i = 1, 2, …, n (Data dari $a_1, a_2, a_3, \ldots, a_n$)

Keluaran: D (Jumlah yang dicari dari $a_1 + a_2 + a_3 + \ldots + a_n$)

Langkah-langkah:

D = 0

Untuk i = 1, 2 , …, n

D = D + a_i

Dengan mengikuti algoritme di atas, untuk membuat algoritme menghitung $\sum x_i$, yang disingkat denga Jx, misalnya, maka dapat dituliskan sebagai

Jx = 0

Untuk i = 1, 2 , …, n

Jx = Jx + x_i

Adapun algoritme lengkap regresi linear dapat dilakukan dengan mengikuti kode pseudo berikut.

Masukan: N , X_i , Y_i, XP
Keluaran: a, b, YT
Langkah-langka:

JX = 0, JXX = 0, JXY = 0, JXX = 0
Untuk i = 1, 2, …, n lakukan
JX = JX + X_i
JY = JY + Y_i
JXY = JXY + X_i Y_i
JXX = JXX + X_i X_i

b = $\frac{nJxx - JxJy}{n(Jxx) - JxJy}$

a = $\frac{Jy - A1Jx}{n}$

Yp = a + b x_p

Keterangan:

N : banyak data

X_i : Data untuk variabel X

Y_i, : Data untuk variabel Y

XP : Nilai taksiran untuk X

a : Koefisien a

b : Koefisien variabel X

JX : $\sum x_i$

JXX : $\sum Y_i$

JXY : $\sum x_i^2$

JXX : $\sum x_i y_i$

Yt : Nilai ramalan untuk Y

Akhirnya, koding algoritme di atas dapat dibuat dengan mengikuti pengkodingan dalam bahasa pemrograman PASCAL atau QBASIC seperti berikut.

Koding Dalam PASCAL

```
Program Regresi_Linear;
Uses Dos, crt;
 Var
        X, Y :  array[1..99] of real;
        n, i   :  integer;
        A, B,JX, JY,JXX,JXY, Xp, Yp : REAL;
        ch        : char;
 Begin
 Clrscr;
   WriteLN ('PROGRAM REGRESI LINEAR dibuat oleh …….');
   Write ('BERAPA BANYAK DATA = '); READLN(N);
   CLRSCR;
   Writeln ('MASUKKAN DATANYA');
Masukan
   FOR I:= 1 TO N DO
   BEGIN
      Write ('X[',I,']='); READLN(X[I]);
      Write ('Y[',I,']='); READLN(Y[I]);
```

```
        writeln;
    END;
    clrscr;
    Write ('BERAPA NILAI X YANG DITAKSIR?=');READLN(XP);
    JX:= 0; JY:= 0; JXX:= 0 ; JXY:= 0;
    FOR I:= 1 TO N DO
    BEGIN
        JX := JX + X[I];
```

Langkah-Langkah

```
        JY := JY + Y[I];
        JXY := JXY + X[I]*Y[I];
        JXX := JXX + X[I]*X[I];
    END;
    B := (N *JXX -JX*JY)/(N*JXX - JX*JY);
    A := (JY - B * JX)/N;
    YP := A + B*XP;

    WRITE  ('HASILNYA ADALAH');
    WRITE ('a=',a:2:2, '   ', 'b=', b:2:2,  '   ',  'YP=',YP:2:2);
    WRITELN;
    WRITELN; ('JADI, PERSAMAAN  REGRESINYA DARI');
    WRITELN;
    WRITELN ('X','          ','Y');
    FOR I:= 1 TO N DO
    BEGIN
        WRITE (X[I]:2:2,'         ', Y[I]:2:2);
        WRITELN;
    END;
    WRITE ('ADALAH  Y = ',A:2:2,'+',B:2:2, 'X');
    CH := READKEY;
END.
```

Keluaran (Output)

Koding Dalam QBASIC

```
    Rem Program Regresi-Linear
    Cls
    Dim x(N), y(N)
    PRINT "PROGRAM REGRESI LINEAR dibuat oleh ……."
    PRINT "BERAPA BANYAK DATA = ": INPUT N
```

```
CLS
PRINT “MASUKKAN DATANYA”
FOR I = 1 TO N
PRINT “X[“;I;”]=”: INPUT X(I)                                   Masukan
PRINT “Y[“;I;”]=”: INPUT Y(I)
PRINT
NEXT I
CLS
PRINT “BERAPA NILAI X YANG DITAKSIR?=”: INPUT XT
JX= 0 : JY= 0 : JXX = 0 : JXY= 0
FOR  I = 1 TO N
      JX = JX + X(I)
     JY = JY + Y(I)
     JXY = JXY + X(I)*Y(I)                        Langkah-Langkah
     JXX = JXX + X(I)*X(I)
 NEXT
 B = (N *JXX -JX*JY)/(N*JXX - JX*JY)
 A = (JY - B * JX)/N
 YT = A + B*XT

 PRINT “HASILNYA ADALAH”
 PRINT  “Koefisien A=';A
 PRINT  “Koefisien B=”;B
 PRINT   “Nilai ramalan YT=”;YT                                  KELUARAN
 PRINT
 PRINT “'JADI, PERSAMAAN  REGRESINYA ADALAH”
 PRINT  “Y = “;A;”+”;B; “X”
END
```

Dengan mengetikkan salah satu program di atas sesuai dengan softwarenya, maka kita dapat melakukan perhitungan dengan cepat koefisien-koefisien a dan b dari sejumlah besar data untuk mendapatkan persamaan regresi $y = a + bx$ dan meramal nilai y berdasarkan nilai x.

Penutup

Perhitungan koefisien a dan b dari persamaan regresi $y = a + bx$ secara sangat sederhana sekali dapat diklakukan dengan membuat program sendiri oleh para pembaca. Dengan cara ini kita akan memiliki program yang dapat dimodifikasikan sendiri tanpa harus terkendala oleh program yang telah dibuat oleh orang lain.

Pustaka

Johnsonbaugh, R. (1997). *Matematika Diskrit* (Edisi keempat). Prentice-Hall, Inc.

Liu, C.L (1995). *Dasar-Dasar Matematika Diskret*. Jakarta: Gramedia Pustaka Utama.

Mathews, J.H. (1992). *Numerical Methods for Mathematics, Science and Engineering* (Second Edition). America: Prentice-Hall Inc.

Sudjana (1992). *Metoda Statistika*. Bandung: Tarsito.

Susila, N.I. (1993). *Dasar-Dasar Metode Numerik*. Jakarta: Depdikbud Dikti Proyek Pembinaan Tenaga Kependidikan Pendidikan Tinggi