



Bahan Kuliah Ekonometrika I

Analisis Regresi Dua Variabel: Beberapa Ide Dasar

IMAM AWALUDDIN

Analisis Regresi Sederhana

Menduga rata-rata peubah tak bebas berdasarkan nilai peubah (satu) bebas yang diketahui

Diilustrasikan dengan data dari Gujarati (2009), dengan populasi beranggotakan 60 keluarga

- X_i : pendapatan/minggu per keluarga
- Y_i : konsumsi/minggu per keluarga
- $i = 1, \dots, 60$ (60 keluarga yang diamati)

Dari 60 keluarga tersebut dikelompokkan ke dalam 10 kelas pendapatan

Pendapatan Keluarga Mingguan $X, \$$

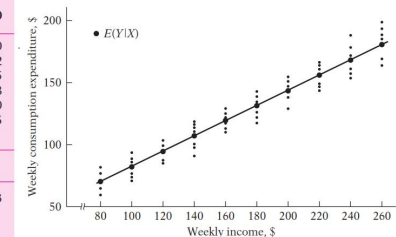
$Y \downarrow X \rightarrow$	80	100	120	140	160	180	200	220	240	260
Weekly family consumption expenditure $Y, \$$	55 60 65 70 75 — —	65 70 74 80 85 88 —	79 84 90 94 98 — —	80 93 95 103 108 113 —	102 107 110 116 118 125 —	110 115 120 130 135 140 —	120 136 140 144 145 160 —	135 137 140 152 157 162 —	137 145 155 165 175 189 —	150 152 175 178 180 185 191
Total	325	462	445	707	678	750	685	1043	966	1211
Conditional means of $Y, E(Y X)$	65	77	89	101	113	125	137	149	161	173

Ke-60 keluarga dibagi menjadi 10 kelompok pendapatan (dari \$80 sampai \$260)

Untuk setiap kelas pendapatan/minggu terdapat variasi jumlah konsumsi/minggu

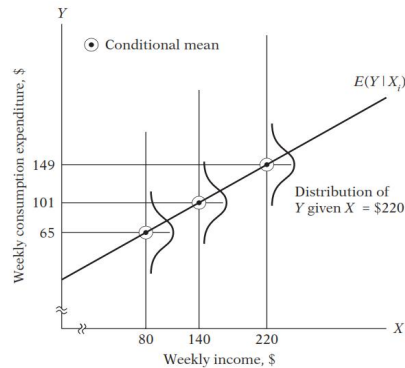
Pada tingkat pendapatan mingguan \$80, rerata pengeluaran konsumsinya adalah \$65, dan seterusnya.

Distribusi bersyarat dari pengeluaran untuk berbagai tingkat pendapatan



Nilai rerata pengeluaran konsumsi ini adalah nilai ekspektasi bersyarat (*conditional expected values*) karena nilainya bergantung pada nilai tertentu dari setiap variabel X , dilambangkan dengan $E(Y|X)$ dibaca sebagai nilai ekspektasi dari Y untuk setiap nilai X . $E(Y|X = 80) = 65$

Konsep Fungsi Regresi Populasi (Population Regression Function – PRF)



} PRF: garis yang menghubungkan nilai harapan bersyarat untuk seluruh kemungkinan nilai X

Konsep Fungsi Regresi Populasi (PRF)

Setiap rerata kondisional $E(Y|X_i)$ adalah fungsi dari X_i , di mana X_i merupakan nilai yang telah ditentukan. Secara simbolis dituliskan:

$$E(Y|X_i) = f(X_i)$$

Jika diasumsikan bahwa hubungan kedua peubah tersebut linier, maka digunakan fungsi linier dari X :

$$E(Y|X_i) = b_1 + b_2 X_i$$

Model/Persamaan Regresi

Dibutuhkan metode tertentu untuk menduga parameter model (intersep dan slope)

Arti dari Linier

- Ø Linier dalam peubah/variabel
- Ø Linier dalam parameter

$$E(Y|X_i) = b_1 + b_2 X_i \longrightarrow \text{Linier dalam peubah}$$

$$E(Y|X_i) = b_1 + b_2 X_i^2 \longrightarrow \text{Non Linier dalam peubah}$$

$$E(Y|X_i) = b_1 + b_2 X_i \longrightarrow \text{Linier dalam parameter}$$

$$E(Y|X_i) = b_1 + b_2^2 X_i \longrightarrow \text{Non Linier dalam parameter}$$

Arti dari Linier

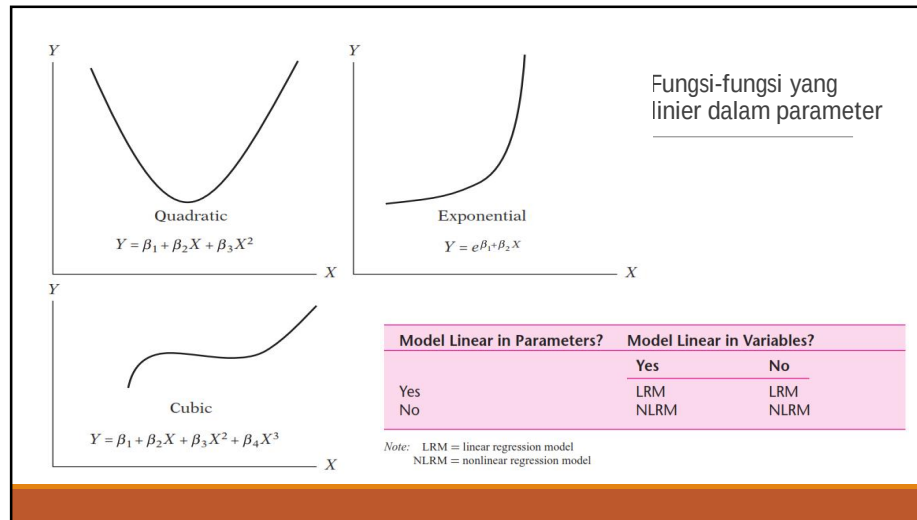
$$E(Y|X_i) = b_1 + b_2 X_i \quad \text{Linier dalam peubah maupun parameter}$$

- Ø Di dalam analisis regresi sederhana, LINIER berarti linier dalam PARAMETER
- Ø Parameter berpangkat paling tinggi 1
- Ø Diperbolehkan pangkat lebih dari satu untuk peubah/ non linier

$$E(Y|X_i) = b_1 + b_2 X_i$$

$$E(Y|X_i) = b_1 + b_2 X_i^2$$

Keduanya Linier dalam parameter:
Model Regresi Linier Sederhana



Fungsi Regresi Populasi Secara Stokastik

- Ø Untuk model konsumsi sebagai fungsi dari pendapatan,
- Ø Dimungkinkan bahwa faktor selain pendapatan juga mempengaruhi konsumsi
- Ø Tidak semua titik tepat pada garis regresi
- Ø Faktor-faktor lain tsb dirangkum dalam komponen error/galat disimbolkan u_i
- Ø u_i dikenal sebagai gangguan stokastik (*stochastic disturbance*) atau faktor kesalahan stokastik (*stochastic error term*) adalah sebuah variable acak yang tidak dapat diamati dan dapat bernilai positif maupun negatif

$$Y_i = E(Y|X_i) + u_i = b_1 + b_2 X_i + u_i$$

Garis Regresi Error/galat
Komponen Komponen
Sistematik/Deterministik Nonsistematik/Stokastik

Fungsi Regresi Populasi Secara Stokastik

$$Y_i = E(Y|X_i) + u_i = b_1 + b_2 X_i + u_i$$

$Y \downarrow$	$X \rightarrow$	
	80	$Y_1 = 55 = b_1 + b_2(80) + u_1$
		$Y_2 = 60 = b_1 + b_2(80) + u_2$
		$Y_3 = 65 = b_1 + b_2(80) + u_3$
		$Y_4 = 70 = b_1 + b_2(80) + u_4$
		$Y_5 = 75 = b_1 + b_2(80) + u_5$
Weekly family consumption expenditure Y , \$	55 60 65 70 75 — —	
Total	325	
Conditional means of Y , $E(Y X)$	65	

Nilai harapan di ruas kiri dan kanan dengan syarat X pada komponen stokastik

$$Y_i = E(Y|X_i) + u_i$$

$$E(Y_i|X_i) = E(E(Y|X_i)) + E(u_i|X_i)$$

Fungsi Regresi Populasi Secara Stokastik

$$E(Y_i|X_i) = E(Y|X_i) + E(u_i|X_i)$$

$$E(Y|X_i) = E(Y_i|X_i)$$

Tujuan dari analisis regresi

$$E(Y_i|X_i) = E(Y_i|X_i) + E(u_i|X_i)$$

$$E(u_i|X_i) = 0$$

Asumsi utama untuk galat/error

Signifikansi Faktor Gangguan Stokastik (Galat/Error Term)

Faktor gangguan u_i merupakan pengganti dari semua variable yang dihilangkan dari model, namun secara kolektif mempengaruhi Y .

Ada beberapa alasan kita tidak memasukkan sebanyak mungkin variable secara eksplisit ke dalam model:

- § Teori yang belum pasti
- § Ketidakterediaan data
- § Peubah utama vs peubah tambahan
- § Sifat alami perilaku manusia (acak)
- § Peubah *proxy* yang kurang berkualitas
- § Model sesederhana mungkin (*Principle of Parsimony*)
- § Kemungkinan hubungan fungsional yang kurang tepat

Fungsi Regresi Sampel (*Sample Regression Function – SRF*)

- ✗ Data pendapatan dan konsumsi: diasumsikan berasal populasi 60 keluarga
- ✗ Fungsi Regresi Populasi (PRF)
- ✗ Secara praktek: tidak mungkin memperoleh informasi secara keseluruhan dari populasi
- ✗ Pengambilan sampel pasangan nilai pendapatan (X) dan konsumsi (Y) dari populasi tersebut
- ✗ Menduga PRF berdasarkan informasi dari sampel
- ✗ Akibat fluktuasi sampel: kemungkinan pendugaan tidak akurat

Ilustrasi Fungsi Regresi Sampel (SRF)

TABLE 2.4
A Random Sample from the
Population of Table 2.1

Y	X
70	80
65	100
90	120
95	140
110	160
115	180
120	200
140	220
155	240
150	260

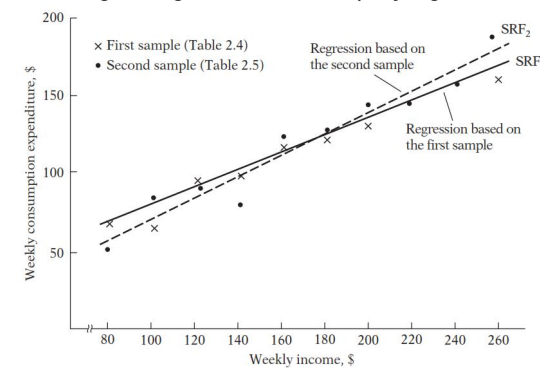
TABLE 2.5
Another Random Sample from the
Population of Table 2.1

Y	X
55	80
88	100
90	120
80	140
118	160
120	180
145	200
135	220
145	240
175	260

Pasangan konsumsi dan pendapatan dari 2 sampel berukuran 10 keluarga yang diambil dari populasi 60 keluarga

Ilustrasi Fungsi Regresi Sampel (SRF)

Garis-garis regresi dari dua sampel yang berbeda



Yang mana yang lebih tepat menggambarkan populasi?

Fungsi Regresi Populasi Dalam prakteknya tidak akan pernah diketahui

Fungsi Regresi Sampel (SRF)

- Ø Regresi yang dibentuk dari sampel
- Ø Dipakai untuk menduga regresi populasi
- Ø Tidak akan pernah sama untuk sampel yang berbeda

$$\hat{Y}_i = \hat{b}_1 + \hat{b}_2 X_i$$

$\hat{Y}_i$: penduga untuk $E(Y|X_i)$ $\hat{b}_1$: penduga untuk b_1

$\hat{b}_2$: penduga untuk b_2

FRS dalam bentuk stokastik: $Y_i = \hat{b}_1 + \hat{b}_2 X_i + \hat{u}_i$
 Faktor residual/sisa, merupakan estimator dari u_i

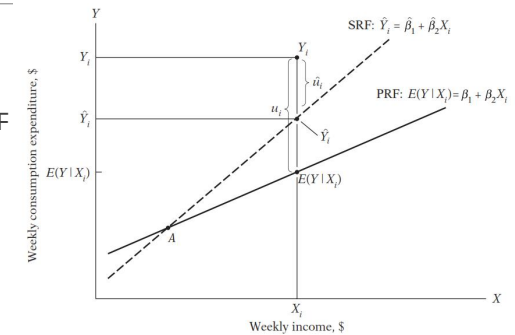
Tujuan Analisis Regresi

- Ø Menduga PRF dengan SRF
- Ø Dengan adanya sampel yang berfluktuasi, SRF hanya pendekatan dari PRF

SRF *underestimate* PRF untuk X di kiri titik A

SRF *overestimate* PRF untuk X di kanan titik A

Bagaimana membentuk SRF sedekat mungkin dengan PRF?



Sumber Bacaan

Gujarati, Damodar N. and Dawn C. Porter (2009). *Basic Econometrics*, Fifth Edition. McGraw-Hill/Irwin. New York. Chapter 2.