

Analisa Regresi Dua Variabel: Estimasi

Tjipto Juwono, Ph.D.

May 2018



SURYA
UNIVERSITY

PRL vs SRL

Apa Perbedaan PRL dan SRL

Population Regression Line Diperoleh dari populasi dengan dengan cara menghubungkan semua *conditional mean*.

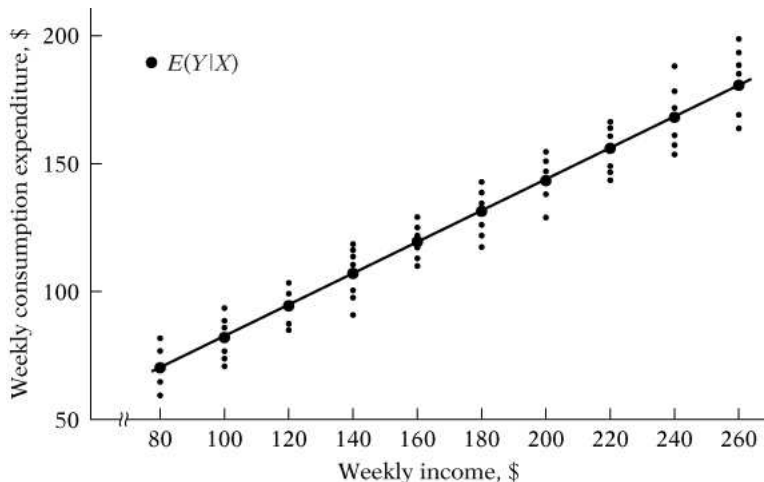
Sample Regression Line Diperoleh dari sample dengan cara metode *least squares*

PRL vs SRL

$Y \downarrow \quad X \rightarrow$	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 95 103 108 113 115	102 107 110 116 118 125 –	110 115 120 130 135 140 –	120 136 140 144 145 – –	135 137 140 152 157 160 162	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

Gambar 1: Populasi terdiri dari 60 keluarga terbagi atas 10 income groups

Mean Values Vs Fixed Values



Gambar 2: Grafik $E(Y|X)$ vs X . Populasi terdiri dari 60 keluarga terbagi atas 10 income groups

Sampel

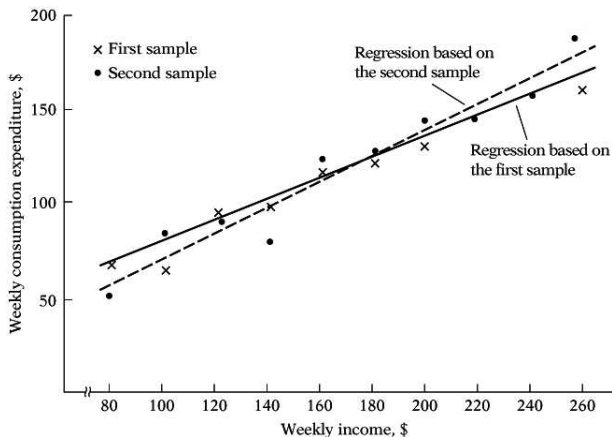
Y	X
70	80
65	100
90	120
95	140
110	160
115	180
120	200
140	220
155	240
150	260

Gambar 3: Sampel-1, dari data pada tabel Gbr (1)

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

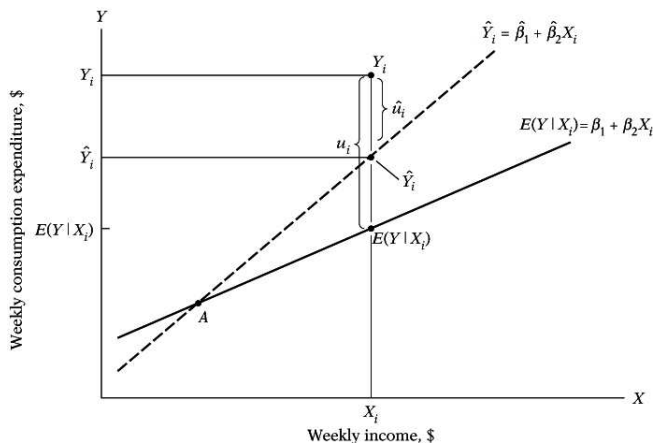
Gambar 4: Sampel-2, dari data pada tabel Gbr (1)

Sample Regression Line



Gambar 5: SRL, diperoleh dari dua sample pada Gbr. (3,4)

PRL vs SRL



Gambar 6: Perbandingan antara PRL dan SRL

Least Squares Principles

- Ingat kembali PRF:

$$Y_i = \beta_1 + \beta_2 X_i + u_i \quad (1)$$

Apa yang terjadi jika kita tidak mempunyai data populasi, dan hanya mempunyai data sampel? Artinya: PRF (Pers. (1)) **tidak diketahui**.

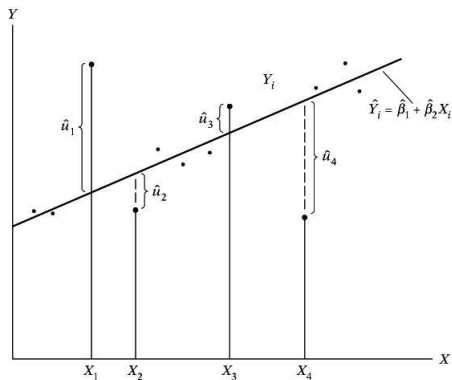
- Kita meng-estimasi PRF, dengan SRF:

$$\begin{aligned} Y_i &= \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i \\ &= \hat{Y}_i + \hat{u}_i \end{aligned} \quad (2)$$

Pertanyaannya adalah: **Bagaimana cara memperoleh SRL?**

Least Squares Principles

Meminimalisasi penjumlahan semua u_i ? \rightarrow Meminimalkan $\sum_i u_i$?



Gambar 7: Bagaimana memperoleh SRF

Least Squares Principles

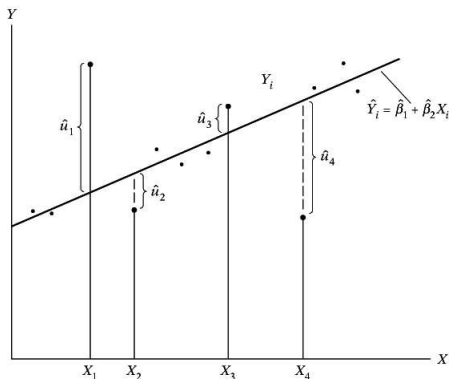
Bagaimana memperoleh SRF?

Meminimalisasi penjumlahan semua u_i ? \rightarrow Meminimalkan $\sum_i u_i$?

Bukan langkah yang tepat. Walaupun sebenarnya u_i mempunyai nilai absolut yang besar sehingga SRF jelas bukan yang terbaik, namun u_i yang positif dan negatif dapat saling meniadakan sehingga diperoleh $\sum_i u_i$ minimum. Walaupun SRF yang diperoleh jauh dari SRF yang terbaik.

Least Squares Principles

Meminimalisasi penjumlahan semua u_i^2 ? \rightarrow Meminimalkan $\sum_i u_i^2$?



Least Squares Principles

Bagaimana memperoleh SRF?

Meminimalisasi penjumlahan semua u_i^2 ? → Meminimalkan $\sum_i u_i^2$?

Langkah yang **tepat!** Dengan mengkuadratkan u_i , maka kita memperoleh nilai yang positif semua. Meminimalisasi $\sum_i u_i^2$ akan menghasilkan SRF yang terbaik. Proses meminimalisasi $\sum_i u_i^2$ merupakan asal usul istilah "*least squares*". Metode ini disebut metode *least squares*

Hasil minimalisasi $\sum u_i^2$

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i \quad (3)$$

$$\hat{\beta}_2 = r \frac{s_y}{s_x} \quad (4)$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} \quad (5)$$

Dengan s_x dan s_y adalah Standard Deviasi Sample:

$$s_x = \sqrt{\frac{\sum (X - \bar{X})^2}{n - 1}} \quad (6)$$

$$s_y = \sqrt{\frac{\sum (Y - \bar{Y})^2}{n - 1}} \quad (7)$$

Dan r adalah koefisien Korelasi:

$$r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{(n - 1)s_x s_y} \quad (8)$$

OLS: Ordinary Least Squares

Untuk selanjutnya, metode yang kita gunakan akan kita sebut **OLS: Ordinary Least Squares**. (Selain *OLS*, ada juga *GLS*, Generalized Least Squares, yang akan kita bahas di Ekonometrika II). Beberapa catatan tentang *OLS*:

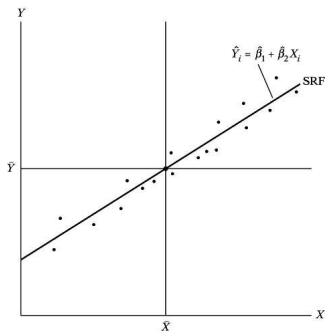
- 1 Estimator-estimator OLS dinyatakan hanya dalam observables X dan Y . Dengan demikian estimator-estimator itu (β_1 dan β_2) dapat dihitung dengan mudah.
- 2 Estimator-estimator itu adalah **point estimators**. Artinya, untuk setiap sampel yang diberikan, estimator itu hanya memberikan satu angka saja (bukan interval) untuk nilai parameter populasi yang relevan.
- 3 Setelah estimator-estimator itu diperoleh (melalui OLS), maka **SRL, Sample Regression Line** dapat diperoleh.

Sample Regression Line, SRL

Sifat-sifat SRL

- 1 SRL melalui (\bar{X}, \bar{Y})

$$\begin{aligned}\hat{\beta}_1 &= \bar{Y} - \hat{\beta}_2 \bar{X} \\ \bar{Y} &= \hat{\beta}_1 + \hat{\beta}_2 \bar{X}\end{aligned}\tag{9}$$



Gambar 8: SRL melalui (\bar{X}, \bar{Y})

Sifat-sifat SRL

- ② Harga rata-rata hasil estimasi $\bar{\hat{Y}}$ sama dengan harga rata-rata dari Y yang sesungguhnya.

$$\bar{\hat{Y}} = \bar{Y} \quad (10)$$

- ③ Harga rata-rata residual adalah nol.

$$\bar{\hat{u}} = 0 \quad (11)$$

PRF,SRF,SRL

Mengulang lagi: Apa bedanya PRF,SRF dan SRL? Perhatikan persamaan-persamaan berikut.

PRF:

$$Y_i = \beta_1 + \beta_2 X_i + u_i \quad (12)$$

SRF:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i \quad (13)$$

SRL:

$$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X_i \quad (14)$$

Bentuk Lain SRF

Persamaan SRF:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i \quad (15)$$

Dapat ditulis:

$$y_i = \hat{\beta}_2 x_i + \hat{u}_i \quad (16)$$

Dengan

$$\begin{aligned} y_i &= Y_i - \bar{Y} \\ x_i &= X_i - \bar{X} \end{aligned}$$

Pers. (16) disebut **bentuk deviasi** (deviation form) dari SRF. Dengan cara yang sama, SRL dapat ditulis dalam bentuk deviasi.

$$\hat{y}_i = \hat{\beta}_2 x_i \quad (17)$$

Sifat \hat{u}_i

- Residu \hat{u}_i tidak berkorelasi dengan hasil prediksi y_i

$$\sum_i \hat{y}_i \hat{u}_i = 0 \quad (18)$$

- Residu \hat{u}_i tidak berkorelasi dengan X_i

$$\sum_i X_i \hat{u}_i = 0 \quad (19)$$

Presi atau Standard Error dari Estimator OLS

Nilai estimator ($\hat{\beta}_1$, $\hat{\beta}_2$) berbeda-beda untuk sample yang berbeda yang diambil dari populasi yang sama. Karena itu kita perlu alat ukur untuk menentukan apakah estimator dari sampel yang satu lebih bagus daripada estimator dari sampel yang lain. Alat ukur itu adalah presisi atau standard error, yang rumusnya adalah:

$$\begin{aligned} se(\hat{\beta}_2) &= \frac{\hat{\sigma}}{\sqrt{\sum(X - \bar{X})^2}} \\ \hat{\sigma} &= \sqrt{\frac{\sum(Y - \hat{Y})^2}{n - 2}} \end{aligned} \quad (20)$$

Presisi atau Standard Error dari Estimator OLS

$$se(\hat{\beta}_1) = \left[\sqrt{\frac{\sum X_i^2}{n \sum (X_i - \bar{X})^2}} \right] \hat{\sigma} \quad (21)$$

BLUE

Best Linear Unbiased Estimator

- 1 Linear
- 2 Unbiased: Nilai ekspektasi dari estimator sama dengan nilai sesungguhnya
- 3 Minimum variance dari estimator

Tugas Kelas

Tugas Kelas

1. Jelaskan apa perbedaan antara PRL dan SRL!
2. Jelaskan bagaimana cara memperoleh PRL!
3. Jelaskan bagaimana cara memperoleh SRL!
4. Mengapa dalam kita perlu meminimalisasi $\sum u_i^2$ (sehingga disebut "least squares"?) Mengapa kita tidak meminimalisasi $\sum u_i$ saja?

Tugas Kelas

- 5 Perhatikan tabel berikut. Untuk setiap kelompok income, ambillah satu nilai yang paling kecil selisihnya dengan condition mean value, lalu gunakan untuk memperoleh SRF!

$Y \downarrow$ \ $X \rightarrow$	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 95 103 108 113 115	102 107 110 116 118 125 –	110 115 120 130 135 140 –	120 136 140 144 145 – –	120 136 140 144 145 – –	135 137 140 152 157 160 162	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	