Chapter 3

REGRESI LINIER SEDERHANA

(MASALAH ESTIMASI)

PowerPoint® Slides

by **Yana Rohmana**

Education University of Indonesian

Jl. Dr. Setiabudi 229 Bandung, Telp. 022 2013163 - 2523

Hal-hal yang akan dipelajari:

- Metode kuadrat terkecil yang biasa (OLS)
- Ukuran tingkat ketepatan suatu perkiraan
- Sifat-sifat yang dimiliki pemerkira OLS dan koefisien determinasi
- Koefisien determinasi, suatu ukuran ketepatan/kecocokan
- Bentuk-bentuk fungsi model regresi
- Asumsi kenormalan

ε_{ij}^2

Metode Kuadrat Terkecil Biasa (OLS)

- $Y_i = A + BX_i + \varepsilon_i$ (sebenarnya)
- $Y_i = a + bX_i + e_i$ (perkiraan)
- Untuk menghitung a dan b berdasarkan data sampel, ada berberapa cara atau metode salah satu diantaranya ialah metode kuadarat terkecil yang biasa (*Ordinary Least Square* = OLS).
- Metode ini ditemukan oleh ahli matematika Jerman bernama Carl Friedrich Gauss, sering disingkat **Gauss**.
- **Gauss** membuat beberapa asumsi (asumsi Klasik) sebagai berikut.

Asumsi	Dinyatakan dalam ε	Dinyatakan dalam Y
1.	$E(\varepsilon_i / X_i) = 0$	$E(Y_i / X_i) = A + BX_i$
2.	$\text{kov}(\varepsilon_i, \varepsilon_j) = 0, \quad i \neq j$	$\text{kov}(Y_i, Y_j) = 0, \quad I \neq j$
3.	$\text{var}(\varepsilon_i, X_i) = \sigma^2$	$\text{var}(Y_i / X_i) = \sigma^2$

Prinsip Metode Kuadrat Terkecil

- Model regresi linear sebenarnya dari populasi yang tidak diketahui dan harus diperkirakan berdasarkan data empiris dari sampel.
- Perhatikan model regresi linear dari sampel:
- $Y_i = a + bX_i + e_i$

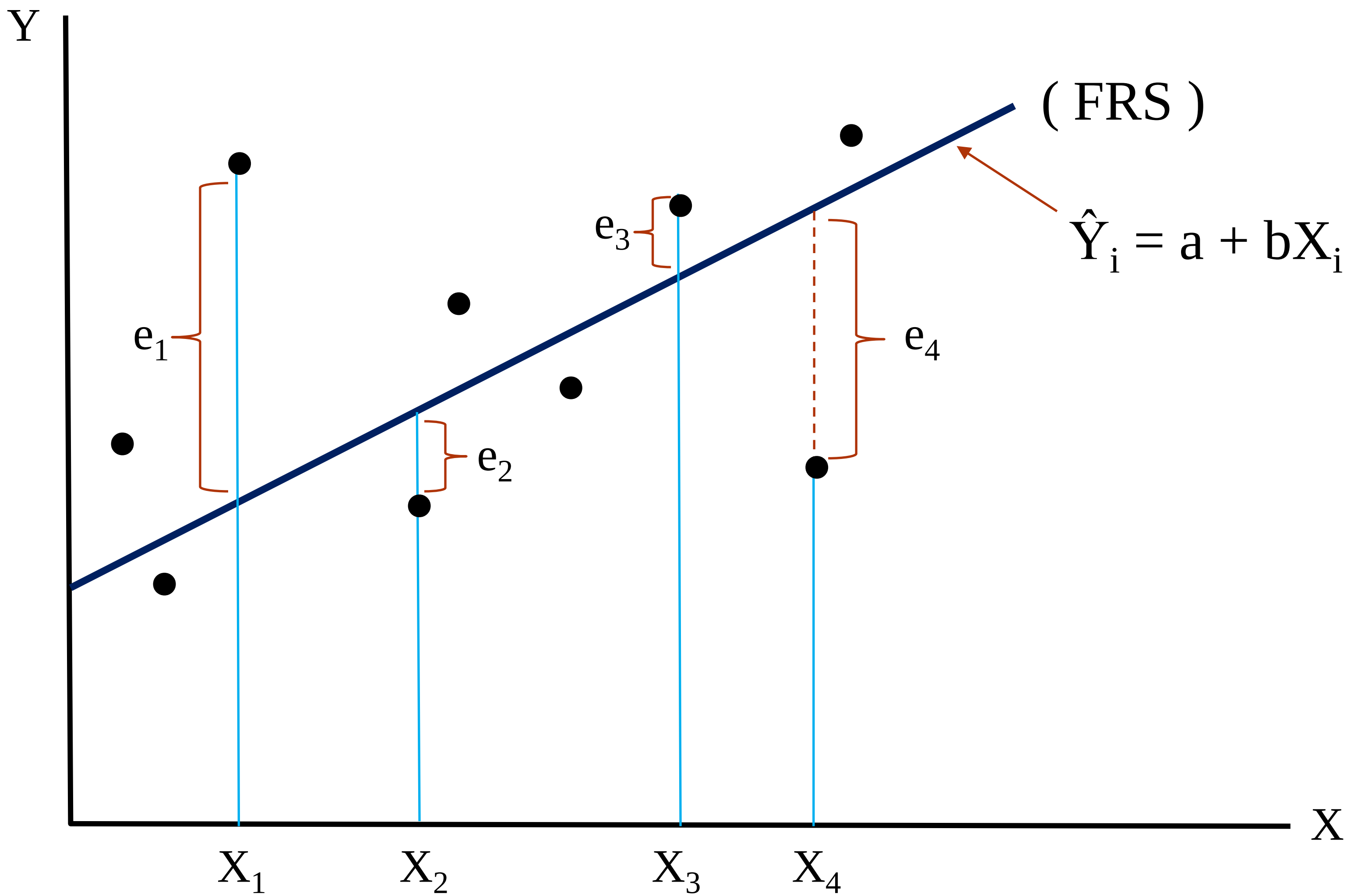
  $= \hat{Y}_i + e_i$

  $\hat{Y}_i = a + bX_i$
- $\hat{Y}_i$ dibaca Y topi merupakan perkiraan/ ramalan dari Y, karena

  $Y_i = \hat{Y}_i + e_i$, maka

  $e_i = Y_i - \hat{Y}_i$
- $\mathbf{e_i = Y_i - a - bX_i}$
- Metode OLS menyatakan bahwa berdasarkan nilai observasi X dan Y sebanyak n pasang akan menentukan nilai a dan b sebagai perkiraan A dan B, sehingga $\Sigma e_i^2 = \Sigma(Y_i - Y_i)^2 = \Sigma(Y_i - a - bXi)^2 = \text{minimum}$

Prinsip Metode Kuadrat Terkecil



Kesalahan Pengganggu Sampel

Prinsip Metode Kuadrat Terkecil

- Apabila kita perhatikan $\Sigma e_i^2 = f(a, b)$, yaitu jumlah kesalahan pengganggu kuadarat merupakan fungsi a dan b, artinya nilainya tergantung kepada nilai a dan b.
- Untuk nilai a dan b yang berlainan , nilai Σe_i^2 juga akan berlainan.
- Dengan metode kuadrat terkecil kita peroleh a dan b yang membuat Σe_i^2 = minimum. Itulah sebabnya mengapa cara ini disebut *least square error*.

$$a = \overline{Y} - b\overline{X}$$

$$b = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum X_i^2 - \left(\sum X_i\right)^2} \quad \text{atau} \quad b = \frac{\sum x_i y_i}{\sum x_i^2}$$

Ukuran Tingkat Ketelitian/Ketepatan Suatu Perkiraan

- Perkiraan a dan b akan bervariasi dari sampel ke sampel, jadi mempunyai standar deviasi, yang disebut *standard error*, sebagai ukuran tingkat ketelitian (*reliability* atau *precision*).
- Standar deviasi dari suatu perkiraan yang disebut *standar error* merupakan akar varian (var) dari perkiraan tersebut.
- Makin kecil *standar error* suatu perkiraan, makin tinggi tingkat ketelitian perkiraan tersebut.
- Kesalahan baku (*standard error*) ialah penyimpangan baku (*standard deviation*) ditribusi sampling untuk pemerkira (*estimato*r) dan distribusi sampling dari suatu pemerkira, merupakan distribusi probabilitas (frekuensi) dari pemerkira, yaitu suatu distribusi dari himpunan nilai-nilai pemerkira (*set of valuesof the estimators*) yang diperoleh dari semua kemungkinan sampel dengan jumlah elemen (n) yang sama dari suatu populasi tertentu.

Ukuran Tingkat Ketelitian/Ketepatan Suatu Perkiraan

- Dalam praktiknya, σ^2 tidak diketahui dan harus diperkirakan dengan S_e^2, di mana :

$$S_e^2 = \frac{1}{n-2}\sum e_i^2 = \frac{1}{n-2}\sum\left(Y_i - \hat{Y}_i\right)^2$$

Oleh karena S_e^2 sering dipergunakan dalam praktik, perhitungannya sebagai berikut.

$$S_e^2 = \frac{1}{n-2}\left(\sum y_i^2 - b^2\sum x_i^2\right)$$

atau ;

$$S_e^2 = \frac{1}{n-2}\left(\sum y_i^2 - b\sum x_i y_i\right)$$

Sifat-sifat yang dimiliki pemerkira OLS

- Pemerkira a dan b yang diperoleh dengan menggunakan metode kuadrat terkecil (*least square method*) disebut *best linear unbiased estimator*, disingkat dengan BLUE.
- Suatu pemerkira, katakan $\hat{\theta}$ (dibaca teta topi atau *cap*) dikatakan pemerkira linear tanpa bias dan terkait (BLUE) dan parameter θ (teta), kalau:

1. Linear ;
2. Tak bias (*unbiased*)
3. Mempunyai *variance* terkecil di dalam kelas seluruh pemerkira tanpa bias dari θ .

Goodness of Fit

- Koefisien Determinasi, Suatu Ukuran Ketepatan/ Kecocokan
- Kita hanya mengharapkan bahwa kesalahan pengganggu berada di sekitar garis regresi dengan jarak sedekat-dekatnya.
- Sebelum menjelaskan arti koefisien determinasi/ penentuan (*coefficient of determination*), terlebih dahulu akan diterangkan arti koefisien korelasi (*coefficient correlation*).

$$r = \frac{\sum x_i y_i}{\sqrt{\sum x_i^{\,2} \sum y_i^{\,2}}}$$

$$r = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{\sqrt{n\sum X_i^{\,2} - \left(\sum X_i\right)^2}\sqrt{n\sum Y_i^{\,2} - \left(\sum Y_i\right)^2}}$$

Goodness of Fit

$$r^2 = \frac{\sum \hat{y}_i^2}{\sum y_i^2}$$

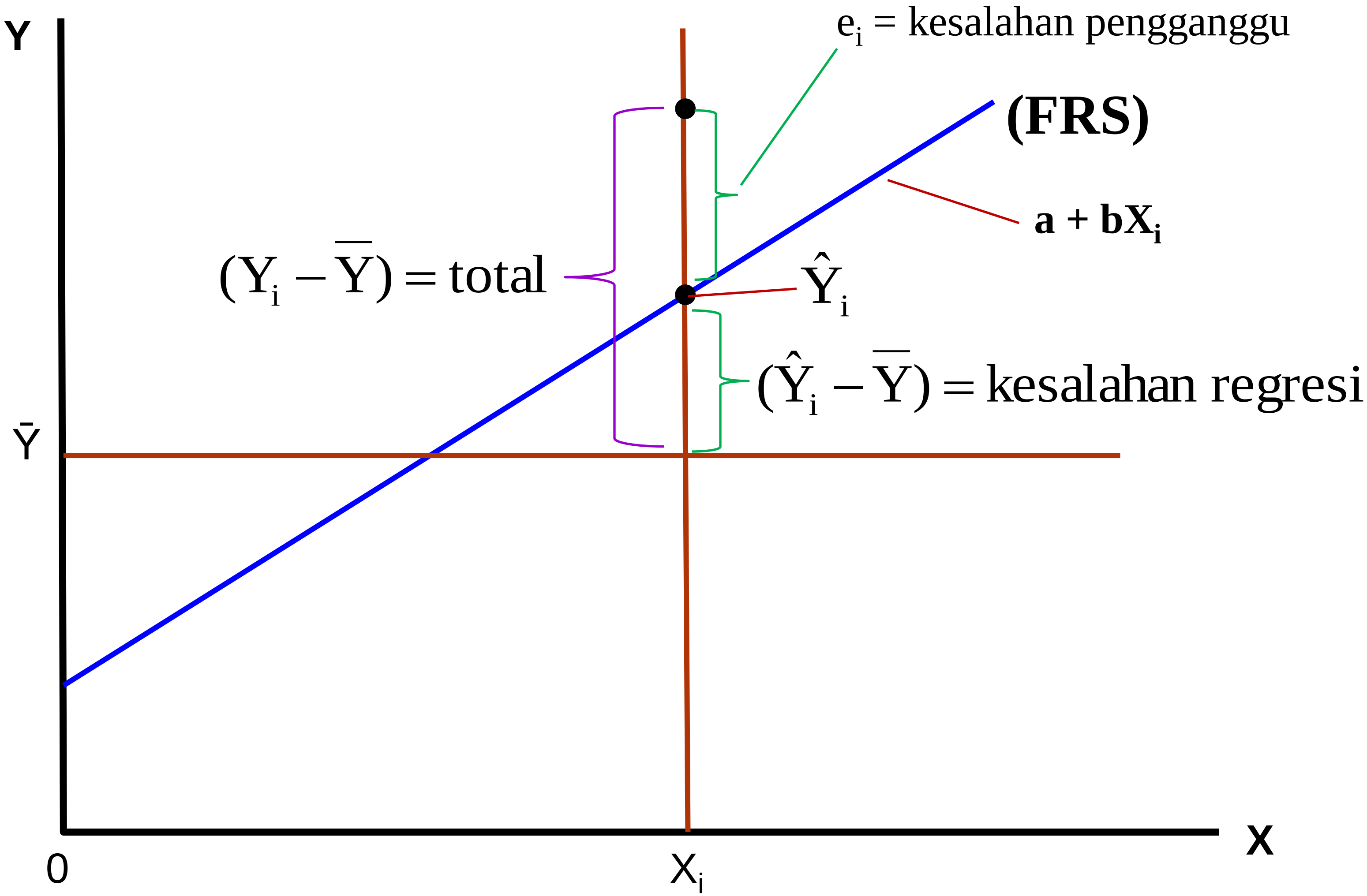
$$r^2 = 1 - \frac{\sum e_i^2}{\sum y_i^2}$$

- Koefisien determinasi merupakan kuadrat koefisien korelasi (r^2).
- Nilai maksimum/ terbesar koefisien determinasi 1 terjadi kalau $e_i^2 = 0$, yaitu kalu semua nilai $e_i = 0$.
- Koefisien determinasi merupakan nilai yang dipergunakan untuk mengukur besarnya sumbangan / andil (*share*) variabel X terhadap variasi atau naik turunnya Y, kalau persamaan regresi $\hat{Y} = a + b X$.
- Jika r = 0,9 ; r^2 = 0,81, berarti sumbangan X terhadap naik turunnya Y sebesar 81%, sedangkan sisanya sebesar 19% merupakan faktor lainnya.

Goodness of Fit

- Pengertian variasi (variation) dengan varian itu berbeda.
- Variasi berarti jumlah deviasi kuadrat (*sum of square of deviation*) suatu variabel terhadap rata-ratanya, yaitu $y_i^2 = (Y_i - \bar{Y})^2$ = TSS, singkatan *Total Sum of Squares*.
- Sedangkan, varian adalah variasi dibagi dengan derajat kebebasan yang tepat (*the appropriate degrees of freedom* = df).
- Jadi varian = variasi / df.
- Untuk keperluan analisis varian:
- $\sum y_i^2$ = TSS, $\sum \hat{y}_i^2$ = ESS (= *explained sum of square*), dan $\sum e^2$= RSS (*residual or unexplained sum of square*)
- $\sum y_i^2 = \sum \hat{y}_i^2 + \sum e_i^2 \rightarrow$ TSS = ESS + RSS
- untuk mengetahui hubungan antara e_i, y_i, dan $\hat{y}_i$ lihat gambar sbb:

Goodness of Fit



Pembagian Variasi Y dalam Dua Komponen

Goodness of Fit

- r^2 yang disebut koefisien determinasi/ penentuan mempunyai dua kegunaan yaitu sebagai berikut:

1. Sebagai ukuran ketepatan/ kecocokan suatu garis regresi yang diterapkan terhadap suatu kelompok data hasil observsasi (*a measure of goodness of fit*). Makin besar nilai r^2, makun bagus atau makin tepat/ cocok suatu garis regresi, sebaliknya, makin kecil makin tidak tepat garis regresi tersebut untuk mewakili data hasil observasi. Nilai r^2 terletak antara 0 dan 1 ($0 \leq r^2 \leq 1$)
2. Untuk mengukur besarnya proporsi (presentase) jumlah variasi Y yang diterangkan oleh model regresi. Atau secara mudah untuk mengukur besarnya sumbangan (*share*) variabel bebas X (= *explanatory/ independent variable*) terhadapa variasi (naik turunnya) Y.

Contoh

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

- X = pendapatan bulanan karyawan perusahaan swasta (ribuan Rp)
- Y = konsumsi bulanan (ribuan Rp)

Ditanyakan:

1. Hitung a, b, dan tulis persamaan regresi linear $\hat{Y} = a + bX$ (dengan metode kuadrat terkecil)! Apa arti b?
2. Hitung var (a), S_a !

   Hitung var (b), S_b !
3. Hitung r^2 ! Apa arti r^2 ?

Σ

Contoh

JAWABAN

X	Y	X^2	Y^2	XY
80	70	6400	4900	5600
100	65	10000	4225	6500
120	90	14400	8100	10800
140	95	19600	9025	13300
160	110	25600	12100	17600
180	115	32400	13225	20600
200	120	40000	14400	24000
220	140	48400	19600	30800
240	155	57600	24025	37200
260	150	67600	22500	39000
$\sum X_i$ 1700	$\sum Y_i$ 1110	$\sum X_i^2$ 322000	$\sum Y_i^2$ 132100	$\sum XY$ 205500

$$\overline{X} = \frac{1}{n}\sum X_i = \frac{1700}{10} = 170$$

$$\overline{Y} = \frac{1}{n}\sum Y_i = \frac{1110}{10} = 111$$

Contoh

- Rumus Praktis:

$$\sum x_i^{\,2} = \sum \left(X_i - \overline{X}\right)^2 = \sum X_i^{\,2} - \left(\sum X_i\right)^2 / n$$

$$\sum y_i^{\,2} = \sum Y_i^{\,2} - \left(\sum Y_i\right)^2 / n$$

$$\sum x_i y_i = \sum \left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right) = \sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n}$$

- Maka:

$$\sum x_i^{\,2} = 322000 - 1700^2 / 10 = 322000 - 289000 = 33000$$

$$\sum x_i y_i = 205500 - (1700)(1110) / 10 = 205500 - 188700 = 16800$$

- Sehingga :

$$b = \frac{\sum x_i y_i}{\sum x_i^{\,2}} = \frac{16800}{33000} = 0{,}5090909$$

$$a = \overline{Y} - b\overline{X} = 111 - 0{,}5090909(170) = 24{,}4545$$

Contoh

a) $\hat{Y} = a + bX = 24{,}4545 + 0{,}5091\ X$

b = 0,5091, artinya jika pendapatan bulanan naik Rp 1000, maka konsumsi bulanan akan naik Rp 509,10

b) $$\mathrm{var}(a) = S_e^{\,2}\frac{\sum X_i^{\,2}}{n\sum x_i^{\,2}}, S_e^{\,2} = \frac{\sum e_i^{\,2}}{n-2} = \frac{\sum y_i^{\,2} - b^2\sum x_i^{\,2}}{8}$$

$\sum y_i^2 = 132100 - (1110)^2/10 = 132100 - 123210 = 8890$

$b^2 \sum x_i^2 = (0{,}5090909)^2 \,.\, 33000 = 8552{,}726952$

$$S_e^2 = \frac{8890 - 8552{,}726952}{8} = 42{,}1591$$

$$\mathrm{var}(a) = 42{,}1591\frac{322000}{10(33000)} = 41{,}137$$

Contoh

$$S_a = \sqrt{\mathrm{var}(a)} = 6{,}4138$$

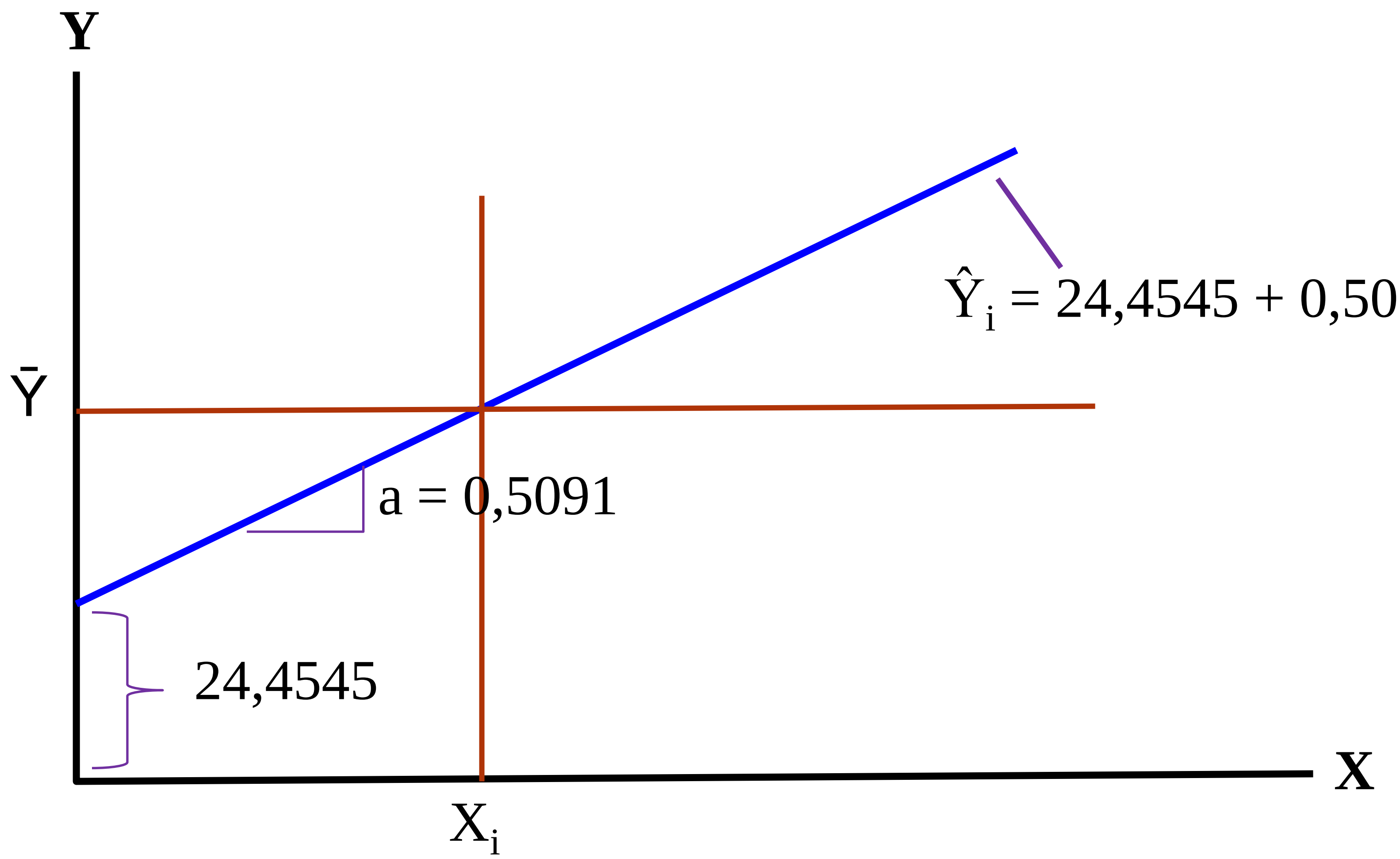
$$\mathrm{var}(b) = S_e^2 / \sum x_i^2 = 42{,}1591 / 33000 = 0{,}0013$$

$$S_b = \sqrt{\mathrm{var}(b)} = 0{,}0357$$

c) $$r^2 = \frac{\left(\sum x_i y_i\right)^2}{\sum x_i^2 \sum y_i^2} = \frac{(16800)^2}{(33000)(8890)} = \frac{282240000}{293370000} = 0{,}96$$

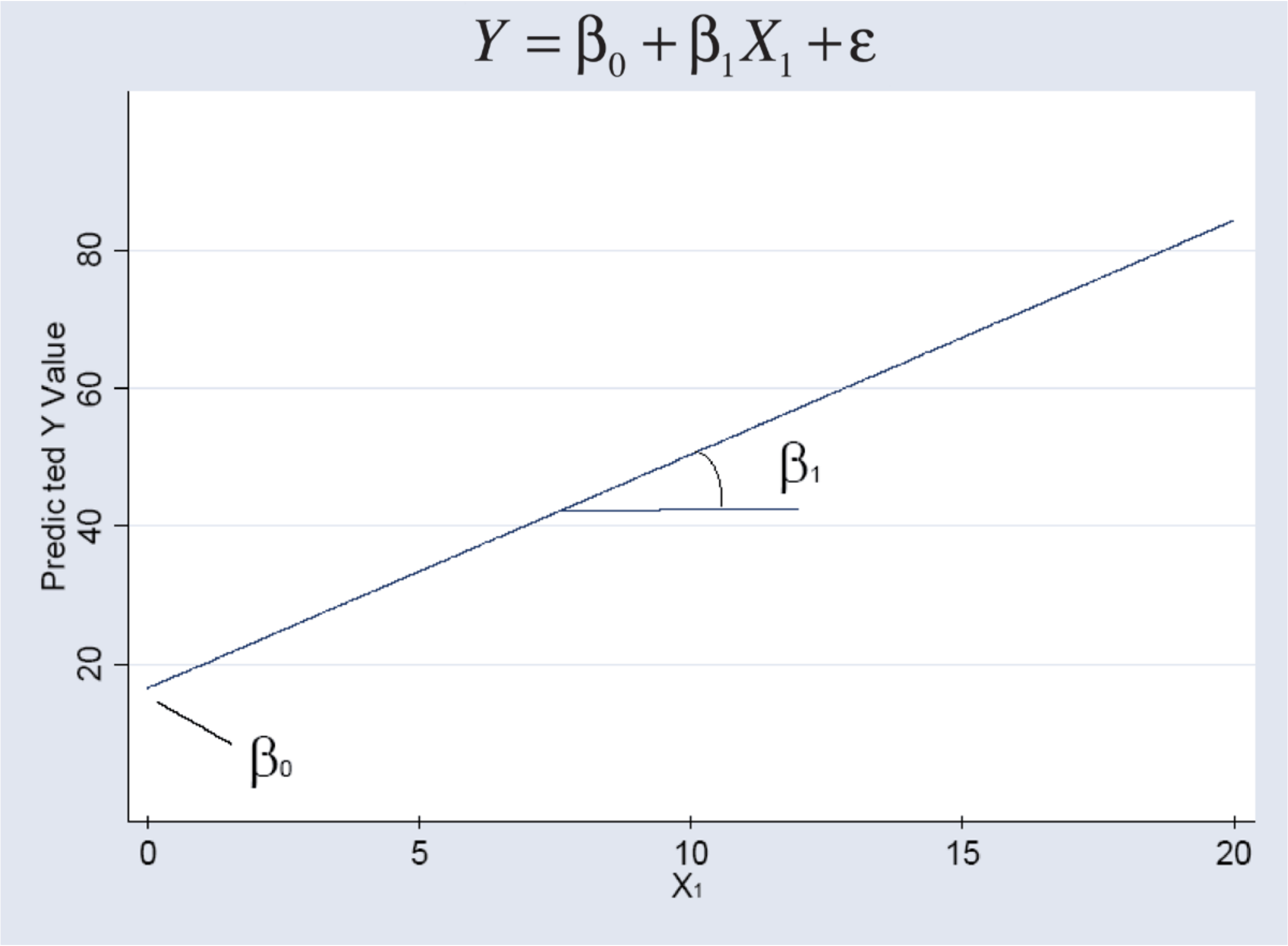
artinya, besarnya sumbangan pendapatan (X) terhadap variasi (naik turunnya) konsumsi (Y) sebesar 96%, sedangkan sisanya sebesar 4% merupakan sumbangan faktor lainnya.

Contoh



Garis Regresi Sampel

Single Variable Regression



QUIZ

1. …………
2. ……………
3. ……………
4. ………………
5. ………………

TERIMA KASIH

NEXT CHAPTER :

PENGUJIAN HIPOTESIS
DALAM REGRESI SEDERHANA