

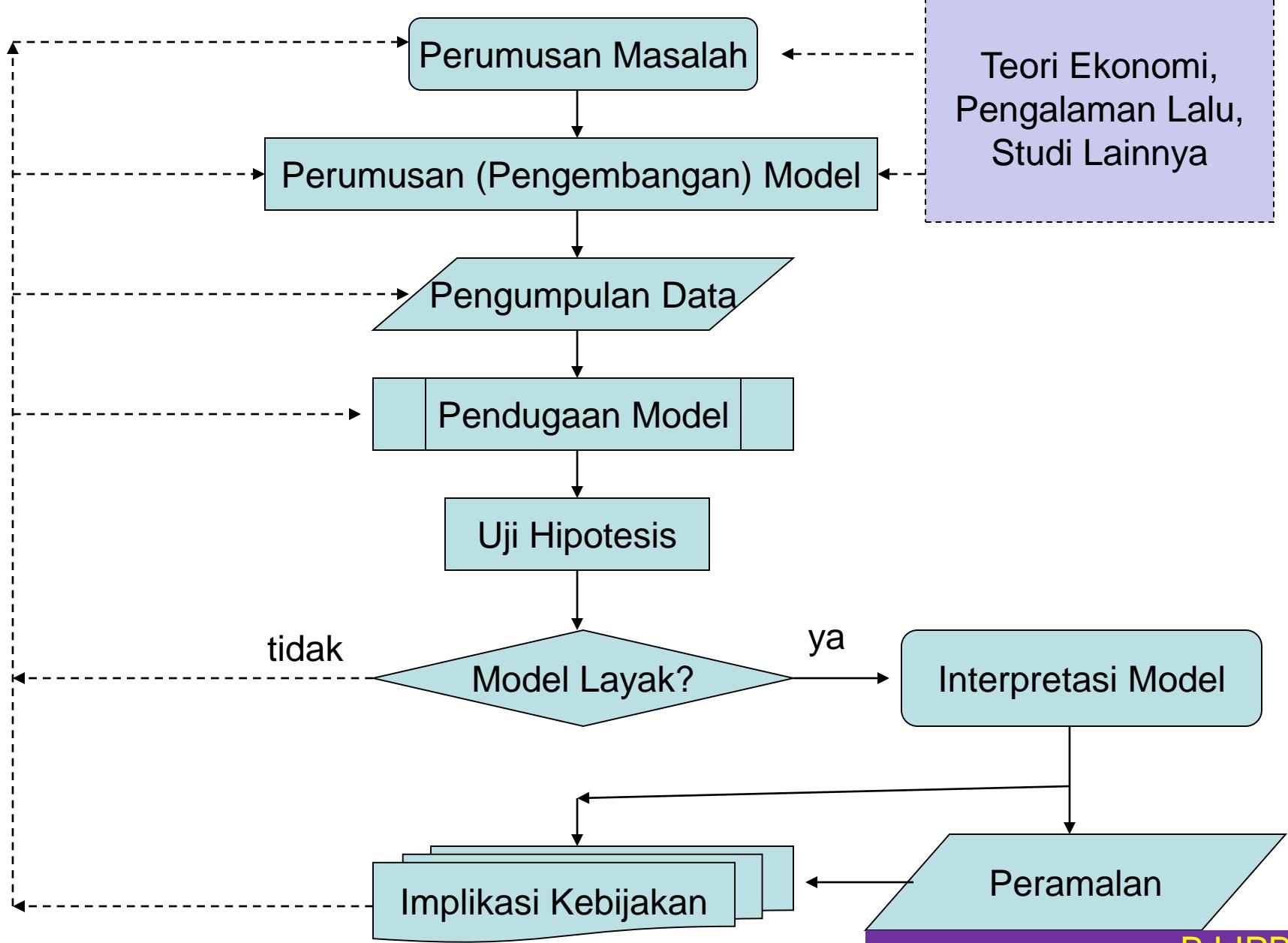
ANALISIS REGRESI

Oleh
Bambang Juanda
<https://bambangjuanda.com/>

Analisis (Model) Regresi

- **Analisis regresi** salah satu metode analisis untuk mengkaji hubungan (“sebab-akibat”) antara satu atau beberapa variabel/peubah bebas X dengan variabel tak bebas Y.
- Menggunakan pendekatan suatu model statistik (yang mencakup *error*)
- **Model**: Abstraksi atau penyederhanaan realitas.
- Hasil Dugaan Model Regresi dapat digunakan untuk peramalan dan memberikan beberapa implikasi kebijakan

Tahapan Studi Empiris



Ekonometrika Membahas (Juanda, 2012): :

- Penggunaan model statistik dalam menjelaskan sistem2 (masalah, perilaku, fenomena) ekonomi.

Tahapan Studi Empiris dgn Model Econometrik

- Formulasi Model
- Pengumpulan Data
- Pendugaan Model
- Uji Hipotesis Model
- Interpretasi Hasil

Tujuan Pemodelan

1. menduga **hubungan-hubungan ekonomi**. Misal, Bagaimana hubungan harga saham dgn karakteristik perusahaan yg mengeluarkan saham, dan juga dgn keadaan umum perekonomian?
2. mengkonfrontasi teori ekonomi dengan fakta, dan menguji hipotesis yg berkaitan dengan perilaku ekonomi. Misal Apakah pelaksanaan DesFis dan pemekaran wilayah dapat meningkatkan pendapatan per kapita dan pengeluaran pemerintah untuk belanja modal serta menurunkan tingkat pengangguran dan tingkat kemiskinan?
3. **peramalan** perilaku peubah-peubah ekonomi. Misal, **meramal** permintaan suatu produk, pendapatan dan belanja negara, inflasi, dan pengangguran, serta defisit anggaran.

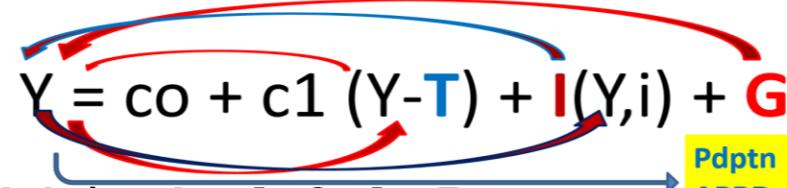
Karena ketiga aspek di atas umumnya berdasarkan data contoh (*sample*), bukan data sensus lengkap, maka akan ada ketidakpastian dalam kajian standar ini, misalnya:

1. Dugaan Hubungan tidak tepat
2. Kesimpulan dari pengujian hipotesis mengandung risiko kesalahan dalam menerima suatu hipotesis yang salah atau pun menolak suatu hipotesis yang benar.
3. Peramalan berdasarkan dugaan hubungan (model) jarang sekali tepat sama dengan nilai sebenarnya.

Untuk mengurangi tingkat ketidakpastian, ahli ekonometrik biasanya menduga beberapa hubungan yang berbeda diantara peubah-peubah yang dikaji. Kemudian melakukan beberapa pengujian untuk menentukan dugaan hubungan yang mana yang paling mendekati gambaran perilaku dari peubah-peubah yang dikaji.

Ketidakpastian ini membuat metodologi statistika sangat penting dalam ekonometrika.

Beberapa Pola Hubungan

1. Langsung (*direct relationship*): $P \rightarrow Q^d$
2. Tidak Langsung (*indirect relationship*):
 $\text{bunga}(i) \rightarrow \text{Investasi } (I) \rightarrow \text{GDP } (Y)$
3. Aditif: $\hat{Q} = \beta_0 + \beta_1 K + \beta_2 L$
4. Mutiplikatif: $\hat{Q} = \beta_0 K^{\beta_1} L^{\beta_2}$, β_2 : elastisitas
5. Saling ketergantungan
(*interdependent*):
 $AS (\text{Aggregate Supply}) = AD (\text{Aggregate Demand})$
 $Y = C + I + G + (EX - IM)$
Misal: $C = co + c1(Y - T)$ dan $EX - IM = 0$
$$Y = co + c1(Y - T) + I(Y, i) + G$$

6. Semu (*spurious relationship*): $A \rightarrow ? \leftarrow B$

Pdptn
APBD

Analisis Korelasi (ρ atau $r = \hat{\rho}$); Pearson

- Mengukur keeratan **hubungan linier** antara dua peubah metrik (berskala interval).
- Tidak harus menggambarkan hubungan sebab akibat
- $-1 \leq r \leq 1$
- r mendekati 1 atau -1 : semakin erat hubungan liniernya
- $r = 1$, data tepat pada garis lurus dgn kemiringan positif
- $r = -1$, data tepat pada garis lurus dgn slope negatif
- $r = 0$ Tidak ada hubungan linier.

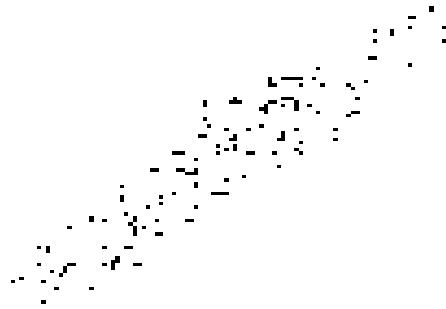
Kelemahan Koef Korelasi:

- Sangat peka terhadap nilai pencilan (*outlier*)
- Tidak dapat mendeteksi hubungan *non linear*

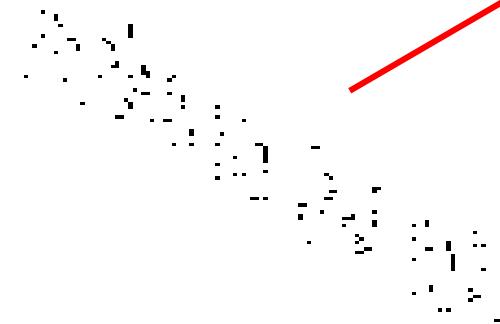
$$r = \frac{S_{xy}}{\sqrt{S_x^2 S_y^2}}$$

$$S_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$
$$S_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

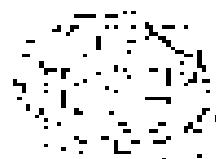
Beberapa Pola hubungan



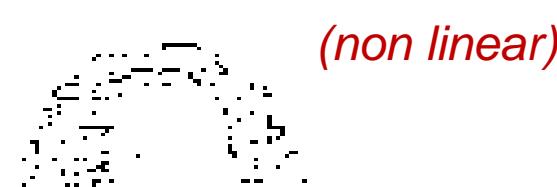
Gambar (a). $r > 0$



Gambar (b). $r < 0$



Gambar (c). $r = 0$



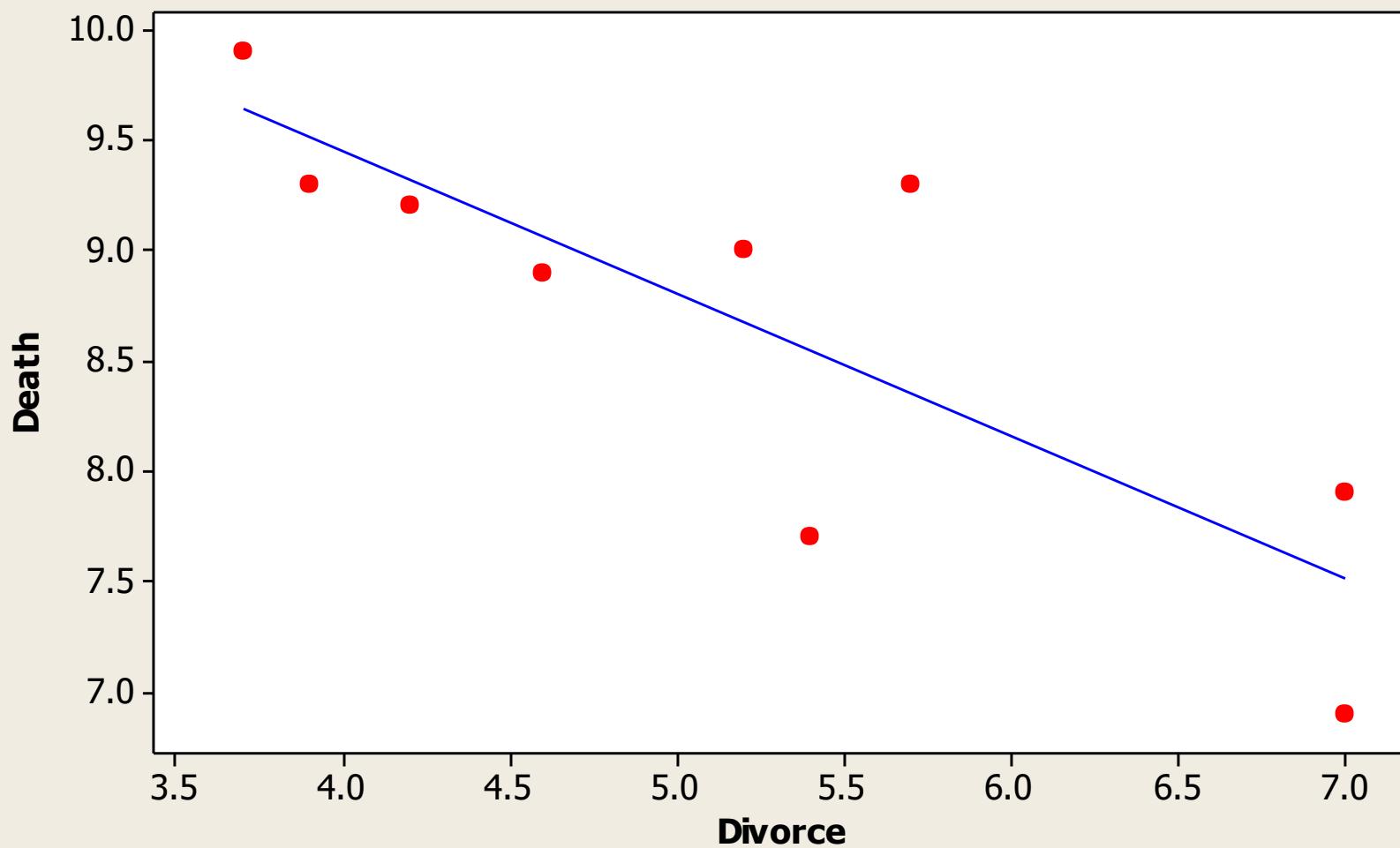
Gambar (c). $r = 0$

*
(outlier)

Perhitungan Koefisien Korelasi (r)

| Daerah di AS | Death (Xi) | Divorce (Yi) | $(x_i - \bar{x})$ | $(y_i - \bar{y})$ | $(x_i - \bar{x})(y_i - \bar{y})$ |
|--------------------|------------|--------------|-------------------|-------------------|----------------------------------|
| Mountain | 6.9 | 7.0 | -1.7780 | 1.8110 | -3.2200 |
| Pacific | 7.7 | 5.4 | -0.9780 | 0.2110 | -0.2064 |
| West South Central | 7.9 | 7.0 | -0.7780 | 1.8110 | -1.4090 |
| East North Central | 8.9 | 4.6 | 0.2220 | -0.5890 | -0.1308 |
| South Atlantic | 9.0 | 5.2 | 0.3220 | 0.0110 | 0.0035 |
| West North Central | 9.2 | 4.2 | 0.5220 | -0.9890 | -0.5163 |
| East South Central | 9.3 | 5.7 | 0.6220 | 0.5110 | 0.3178 |
| New England | 9.3 | 3.9 | 0.6220 | -1.2890 | -0.8018 |
| Mid Atlantic | 9.9 | 3.7 | 1.2220 | -1.4890 | -1.8196 |
| Rataan | 8.678 | 5.189 | | Jumlah | -7.7822 |
| StDev | 0.963 | 1.223 | | Korelasi | -0.8260 |

Plot Tebaran Data Tingkat Kematian dan Tingkat Perceraian ($r = -0.823$)



Apakah berimplikasi bahwa jika tk perceraian tinggi akan mengakibatkan tk kematian rendah?

Utk menentukan apakah korelasi antara 2 peubah merupakan hubungan sebab-akibat harus memenuhi kriteria (Mosteller dan Tukey, 1977) :

1. **Kekonsistenan**: apakah hubungan tsb berlaku pd kondisi yg lain juga?
2. **Mekanistik**: bagaimana menentukan suatu model yg menggambarkan proses hubungan sebab-akibat tsb? → **Model Regresi**

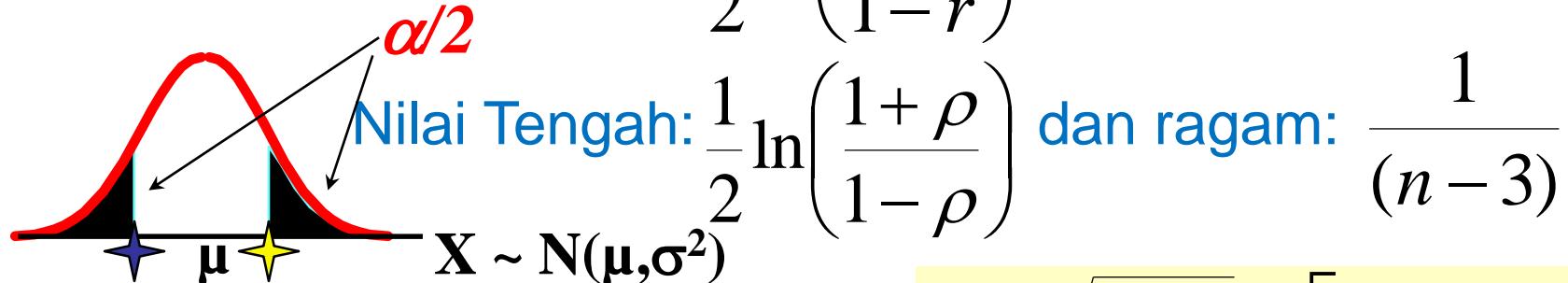
Cara lain utk menjelaskan perkiraan adanya hubungan sebab-akibat adalah mendefinisikan peubah lain, yg disebut **Lurking Variable**, yg layak menerangkan hubungan tersebut. Cara ini bermanfaat terutama jika ada kesan terdapat hubungan semu (**spurious relationship**)

Usia  **Usia Muda** : tk Kematian Rendah, tk Perceraian tinggi (-)
(Lurking variable) **Dewasa (Tua)** : tk Kematian Tinggi, tk Perceraian rendah (-)

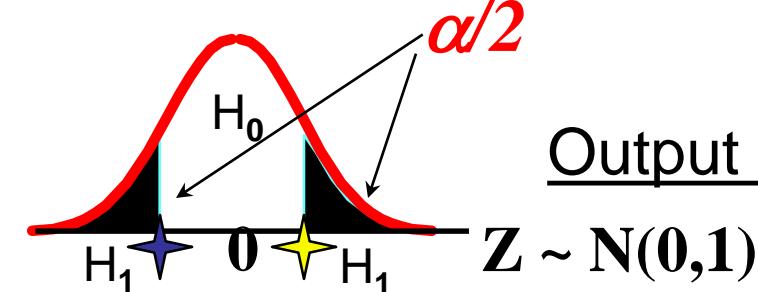
Uji Signifikansi Koef Korelasi

Hipotesis statistik: $H_0: \rho = 0$ (tdk ada korelasi)
 $H_1: \rho \neq 0$ (ada korelasi)

Teori: Peubah acak $\frac{1}{2} \ln \left(\frac{1+r}{1-r} \right) = x$, menyebarkan Normal dgn



Statistik Uji Normal Bakunya: $Z = \frac{\sqrt{n-3}}{2} \ln \left[\frac{(1+r)(1-\rho)}{(1-r)(1+\rho)} \right]$



Output Komputer: Pearson Correlation = -0.823,
P-Value = 0.006 < α (5%)

Nilai-p (*p-value*, *sign.*): **peluang (risiko) kesalahan dlm menyimpulkan H_1** . Artinya, meskipun kita menyimpulkan H_1 , tapi mungkin saja H_0 yg benar. Alternatif kriteria uji menggunakan taraf nyata α (mis 1%, 5%, 10%) adalah sbb:

- Jika $p > \alpha$ maka terima H_0 (kesalahannya melebihi batas taraf nyata jika terima H_1)
- Jika $p < \alpha$ maka terima H_1 (kesalahannya kurang dari taraf nyata jika terima H_1).

taraf nyata α : **peluang (risiko) kesalahan maks yg dpt ditolerir dlm menyimpulkan H_1** (berapa α ?)

α : nyata 5%(*), 1%(**), 0.1(***) ? → tergantung konteks permasalahan !

PR: Jelaskan korelasi antara tingkat kemiskinan dan dana utk penanggulangan kemiskinan, apakah positif atau negatif?

Korelasi Antara Berbagai Jenis Belanja dengan Indikator Prioritas Nasional

(Jelaskan hasil analisis korelasi ini, Data 2005-2017)

| | POV | UNEMP | IPM | BP | BB | BM | L.BM | BBS |
|--------|-------------------|-------------------|------------------|------------------|------------------|------------------|------------------|-----------------|
| UNEMP | 0.9847 0.0000 | 1.0000 | | | | | | |
| IPM | -0.9379 0.0000 | -0.9482 0.0000 | 1.0000 | | | | | |
| BP | -0.9386 0.0000 | -0.9493 0.0000 | 0.9981 0.0000 | 1.0000 | | | | |
| BB | -0.8994 0.0000 | -0.8988 0.0000 | 0.9884 0.0000 | 0.9876 0.0000 | 1.0000 | | | |
| BM | -0.8996 0.0000 | -0.9157 0.0000 | 0.9514 0.0000 | 0.9555 0.0000 | 0.9455 0.0000 | 1.0000 | | |
| L.BM | -0.8829 0.0001 | -0.8982 0.0001 | 0.9430 0.0000 | 0.9462 0.0000 | 0.9130 0.0000 | 0.8514 0.0004 | 1.0000 | |
| BBS | -0.6252 0.0223 | -0.6830 0.0101 | 0.5213 0.0677 | 0.5263 0.0647 | 0.4192 0.1539 | 0.6112 0.0264 | 0.3772 0.2267 | 1.000 |
| STotal | -0.5125 0.0733 | -0.5546 0.0492 | 0.3291 0.2722 | 0.3356 0.2623 | 0.2147 0.4812 | 0.3857 0.1931 | 0.3268 0.2998 | 0.687 0.0095 |

Hipotesis statistik: $H_0: \rho = 0$ (tdk ada korelasi)
 $H_1: \rho \neq 0$ (ada korelasi)

Pengertian Model & Tujuan Pemodelan

- Perumusan masalah → Model
- Model: Abstraksi realitas → dlm pers matematika
- Model ekonometrika: model statistik yg mencakup *error*

$$Y = f(X_1, X_2, \dots, X_p) + \text{error} \quad (2.1)$$

data aktual = dugaan + sisaan (simpangan)

data = komp. sistematik + komp. non-sistematik

$$\text{dugaan } Y = f(X_1, X_2, \dots, X_p) \quad (2.2)$$

diharapkan unsur-unsur ketidak-teraturan nilai Y dapat dijelaskan oleh nilai-nilai dari peubah X_1, X_2, \dots, X_p berdasarkan model dugaan dalam persamaan (2.2). Oleh karena itu, komponen sisaan diusahakan menjadi relatif kecil dibandingkan komponen dugaannya.

Deskripsi komponen error :

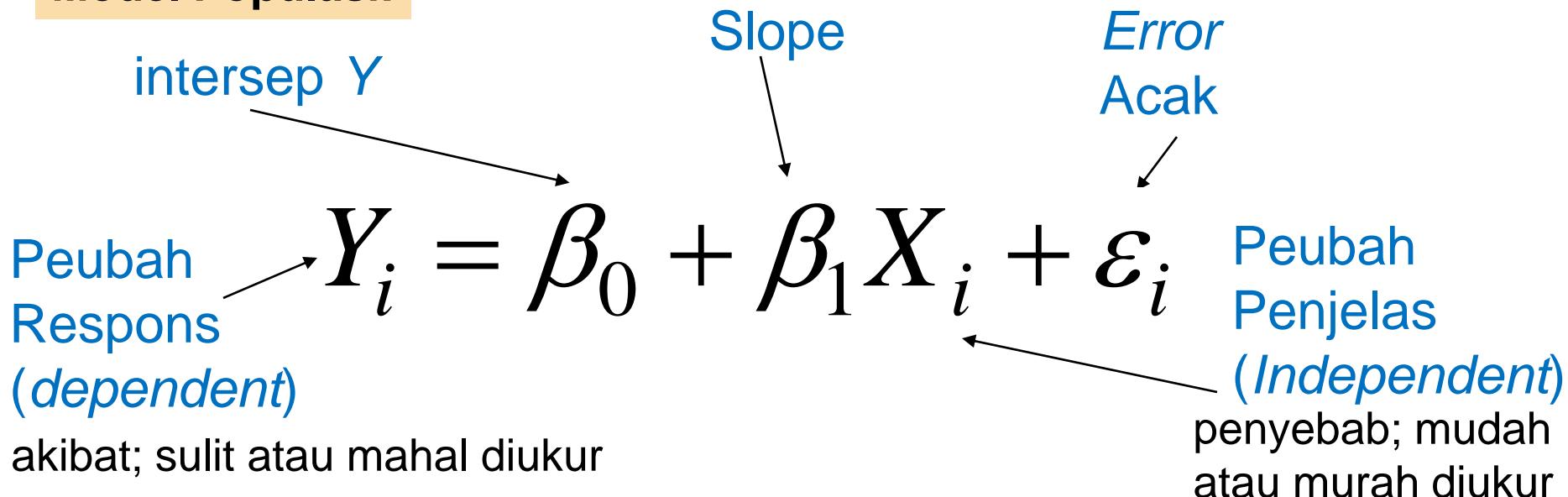
1. Kesalahan pengukuran dan *proxy* dari peubah respons Y maupun peubah penjelas X₁, X₂, ..., dan X_p.
2. Asumsi bentuk fungsi f yang salah. Mungkin ada bentuk fungsi lainnya yang lebih cocok, linear maupun non-linear.
3. *Omitted relevant variables*. Peubah (*variable*) yang seharusnya dimasukkan ke dalam model, dikeluarkan karena alasan-alasan tertentu (misalnya penyederhanaan, atau data sulit diperoleh dan lain-lain).
4. Pengaruh faktor-faktor lain yang belum terpikirkan atau tidak dapat diramalkan (*unpredictable effects*).

Digunakan → Seni dalam Memodifikasi (Mengembangkan) Model

Model Regresi Linear Sederhana

- Hubungan antar Peubah dlm Fungsi Linear dlm Parameter
- Garis Lurus yg Paling Cocok dgn Data

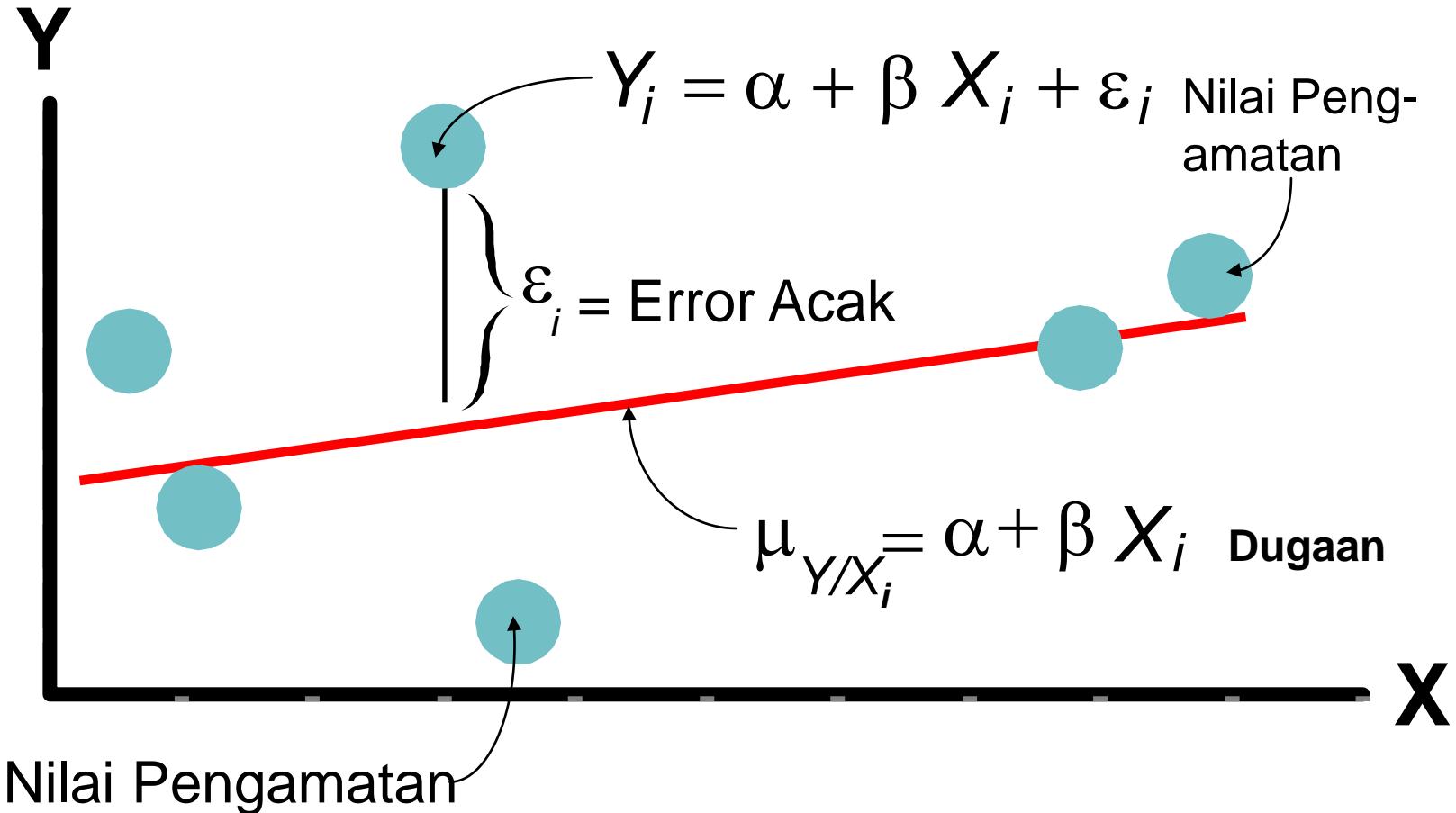
Model Populasi:



Model Regresi Contoh:

$$Y_i = b_0 + b_1 X_i + e_i$$

Model Regresi Linear Populasi



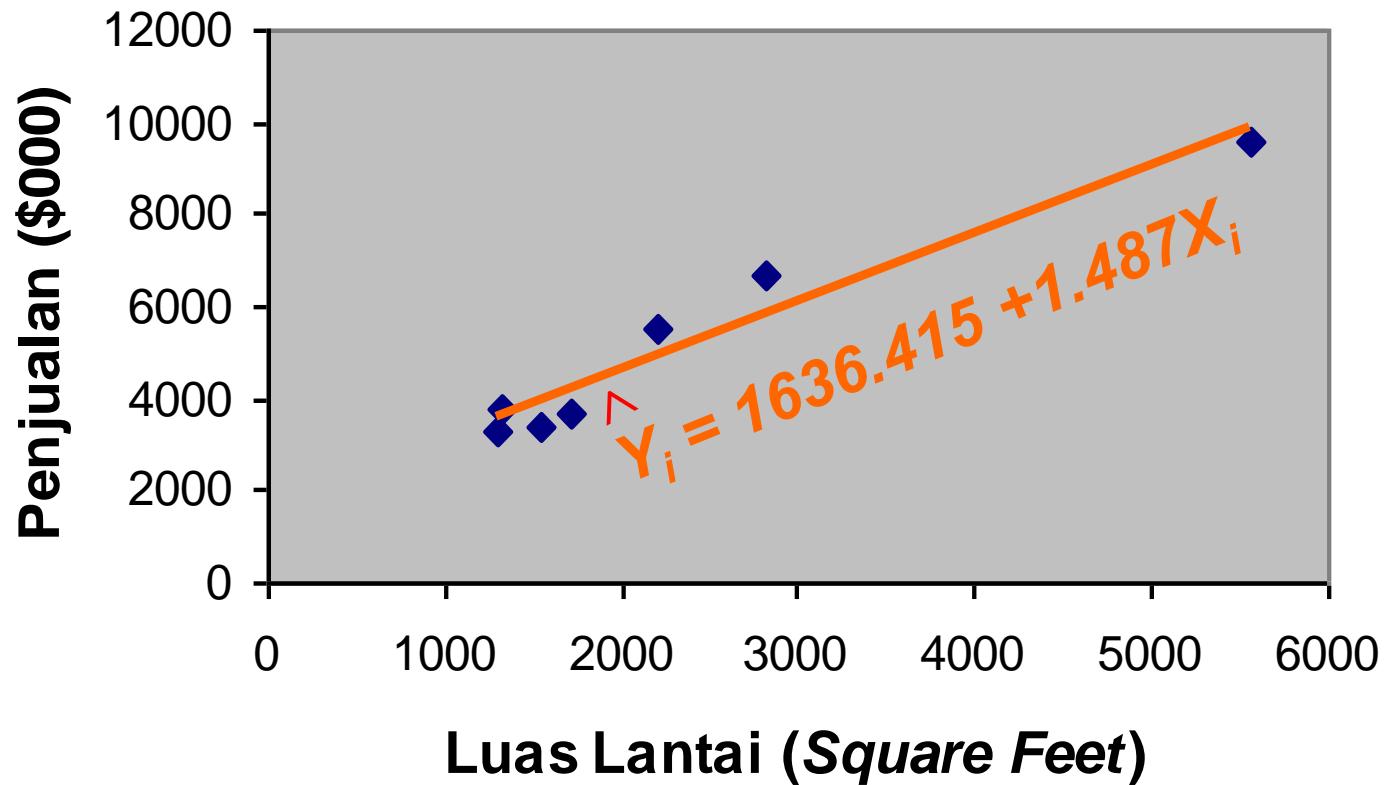
Persamaan Regresi Linear Sederhana (Teladan)



Ingin mengkaji hubungan antara luas lantai toko (hasil pertanian) dengan total penjualan tahunannya. Data contoh utk 7 toko telah diperoleh. **Tentukan persamaan garis lurus yg paling cocok dgn data tsb**

| Store | Square Feet | Annual Sales (\$000) |
|-------|-------------|----------------------|
| 1 | 1,726 | 3,681 |
| 2 | 1,542 | 3,395 |
| 3 | 2,816 | 6,653 |
| 4 | 5,555 | 9,543 |
| 5 | 1,292 | 3,318 |
| 6 | 2,208 | 5,563 |
| 7 | 1,313 | 3,760 |

Diagram Pencar (*Scatter Diagram*) & Dugaan Model Regresinya



Manakah peubah X dan manakah peubah Y? Mengapa?

Model Regresi Linear Contoh

$$Y_i = b_0 + b_1 X_i + e_i$$

$$\hat{Y}_i = b_0 + b_1 X_i$$

\hat{Y}_i = Nilai dugaan Y utk pengamatan ke- i

X_i = Nilai X utk pengamatan ke- i

b_0 = Dugaan bagi koef intersep populasi β_0 ;
rata-rata Y jika X=0

$\frac{\Delta Y}{\Delta X} = b_1$ = Dugaan bagi koef slope populasi β_1 ;
perbedaan rata-rata Y jika X berbeda 1 unit

Metode (Jumlah) Kuadrat (Sisaan) Terkecil: MKT atau *Ordinary Least Squares*.

Mencari dugaan koefisien yg menghasilkan jumlah kuadrat simpangan antara data aktual dgn data dugaan MINIMUM

$$e_i = \hat{e}_i = Y_i - \hat{Y}_i, \text{ dan } \text{minimumkan } q = \sum_{i=1}^n e_i^2 \text{ shg}$$

Simple Linear Model : $Y_i = a + b X_i + e_i$

$$\hat{\beta} = b = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{n \sum_{i=1}^n X_i Y_i - \sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{n \sum_{i=1}^n X_i^2 - \left(\sum_{i=1}^n X_i \right)^2}$$

$$\hat{\alpha} = a = \bar{Y} - b \bar{X}$$

Contoh Output STATA (Bgm Interpretasinya?)

Stata/MP 14.0 - C:\Users\BJ\Google Drive\GD\materi\Ekonometrika\Presentasi\buku eknmtrk-BJ+\ekonometrika BJ\Do File\Bab 2 Sales_Luas.dta

File Edit Data Graphics Statistics User Window Help

Review Filter commands here

Command _rc

1 use "C:\Users\BJ\G...
2 regr Penjualan Luas

Variables

| Name | Label |
|-----------|---------------------------|
| Luas | Luas Lantai (feet2) |
| Penjualan | Penjualan tahunan (\$000) |

History dari Command yang dipakai

Variabel yang Ada pada dataset

Source SS df MS Number of obs = 7
Model 30380456.1 1 30380456.1 F(1, 5) = 81.18
Residual 1871199.59 5 374239.919 Prob > F = 0.0003
Total 32251655.7 6 5375275.95 R-squared = 0.9420
Adj R-squared = 0.9304
Root MSE = 611.75

| Penjualan | Coef. | Std. Err. | t | P> t | [95% Conf. Interval] |
|-----------|----------|-----------|------|-------|----------------------|
| Luas | 1.486634 | .1649992 | 9.01 | 0.000 | 1.06249 1.910778 |
| _cons | 1636.415 | 451.4953 | 3.62 | 0.015 | 475.809 2797.02 |

$\hat{Y}_i = 1636.415 + 1.487 X_i$

Jika Selang Kepercayaan tidak mencakup 0, maka terima $H_1: \beta_j \neq 0$

Disini disajikan output dari Analisis

Menulis command disini

Interpretasi Koefisien

$$\hat{Y}_i = 1636.415 + 1.487X_i$$

Interpretasi Nilai slope 1.487 ('umumnya'): utk kenaikan 1 unit dlm X, diduga Y akan meningkat 1.487unit.

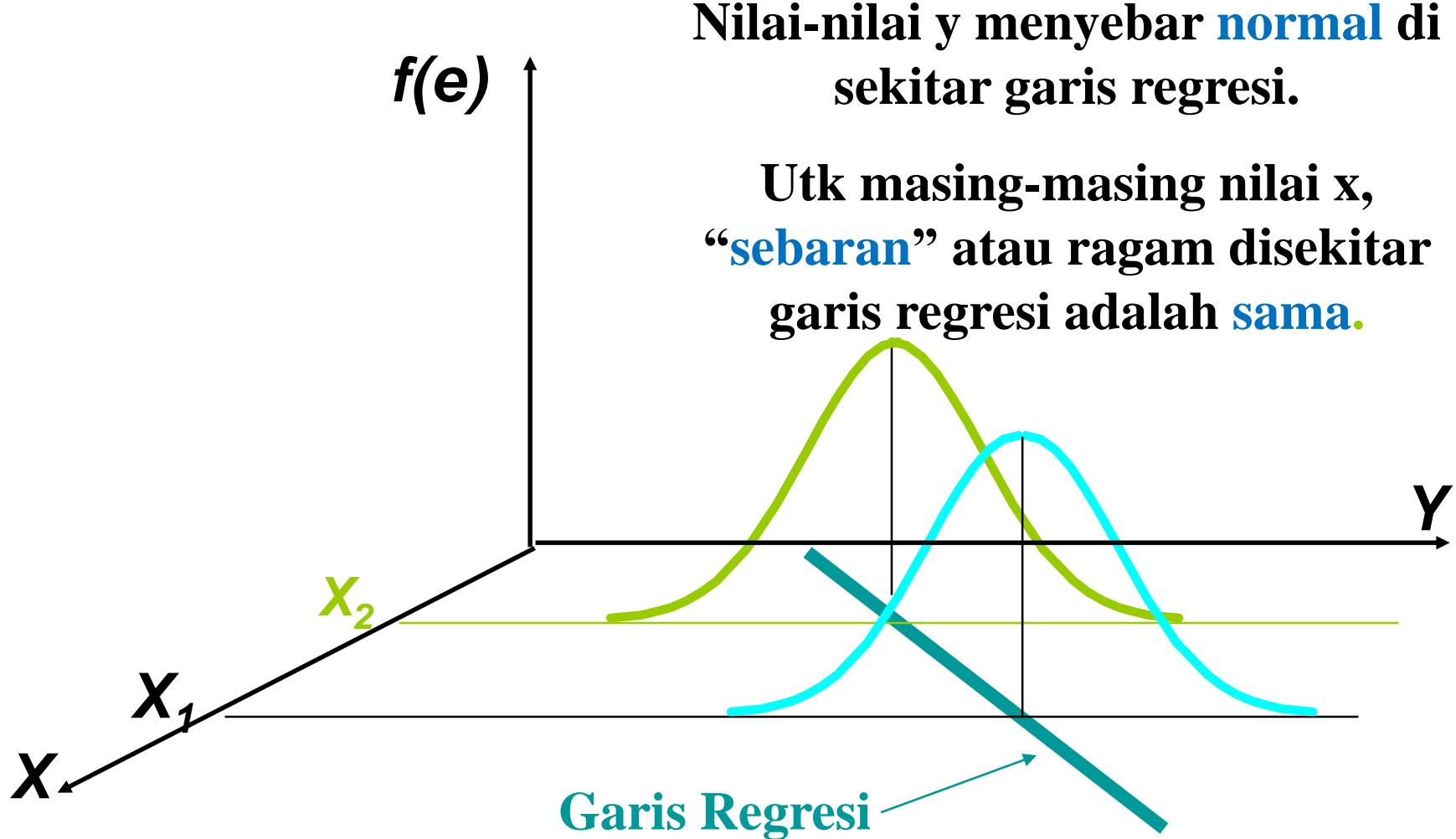
Interpretasi 'paling tepat' dlm kasus ini:

- *Rata2 perbedaan total penjualan antara toko yg luasnya berbeda 1 square feet adalah \$1487 per th*

Implikasi dari dugaan slope (dgn asumsi tertentu):

- *Jika ukuran lantai toko naik 1 square feet, model tsb memprediksi bahwa total penjualan yg diharapkan akan meningkat \$1487 per th.*
- *Bagaimana dengan interpretasi intersep?*

Ragam Error Sekitar Garis Regresi



Asumsi Model Regresi Linear

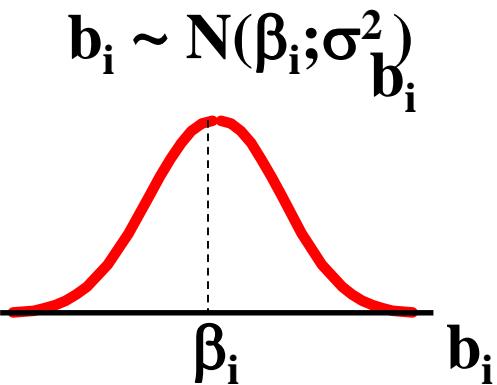
- **Kenormalan & Kebebasan**
 - Nilai-nilai **Y Menyebar Normal utk masing-masing nilai X**; dgn $E(Y_i) = \alpha + \beta X_i$ dan $\text{Var}(Y_i) = \sigma^2$ utk semua i.
 - (i) **Sebaran Peluang *Error adalah Normal, Bebas dan Identik*** dengan $E(\varepsilon_i) = 0$ dan $\text{Var}(\varepsilon_i) = \sigma^2$ untuk semua i.
 - (ii) Peubah X dan ε_i bebas
- **Homoskedastisitas (Ragam Konstan)**
- **Sisaan (*Error*) bebas**

Dugaan Galat baku (*Standard Error*)

$$\hat{\sigma} = S_{y/x} = \sqrt{\frac{JKS}{n-2}} = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n-2}}$$

Simpangan Baku pengamatan-pengamatan disekitar garis regresi

Jika **asumsi tentang peubah acak ε_i dipenuhi** maka masing-masing dugaan koefisien akan menyebar normal dgn $E(b_0) = \beta_0$ dan $E(b_1) = \beta_1$ serta dugaan ragam:



$$\sigma_{b_0}^2 = \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right] \sigma^2$$

$$S_{b_1} = \frac{S_{Y/X}}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}}$$

Teladan: Toko Hasil Pertanian

Model Regresi yg diperoleh:

$$\hat{Y}_i = 1636.415 + 1.487X_i$$

Data utk 7 Toko:

| Toko | Square Feet | Annual Sales (\$000) |
|------|-------------|----------------------|
| 1 | 1,726 | 3,681 |
| 2 | 1,542 | 3,395 |
| 3 | 2,816 | 6,653 |
| 4 | 5,555 | 9,543 |
| 5 | 1,292 | 3,318 |
| 6 | 2,208 | 5,563 |
| 7 | 1,313 | 3,760 |

Slope model ini adalah 1.487.

• Apakah ada hubungan linear antara *ukuran luas toko* dgn total penjualan tahunannya?

• Apakah total penjualan dpt diprediksi dari ukuran luas lantai tokonya?

Inferensi mengenai Slope: Uji-t

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

- **Uji-t utk Slope Populasi**

Ada Hubungan Linear antara X dgn Y ?

- Hipotesis Nol dan Alternatif

$H_0: \beta_1 = 0$ (X tidak dpt menjelaskan Y)

$H_1: \beta_1 \neq 0$ (X dapat menjelaskan Y)

- Statistik Uji:

$$t = \frac{b_1 - \beta_1}{S_{b_1}}$$

dimana

$$S_{b_1} = \sqrt{\frac{S_{Y/X}}{\sum_{i=1}^n (X_i - \bar{X})^2}}$$

dan $db = n - 2$

Inferensi ttg Slope: Contoh Uji-t

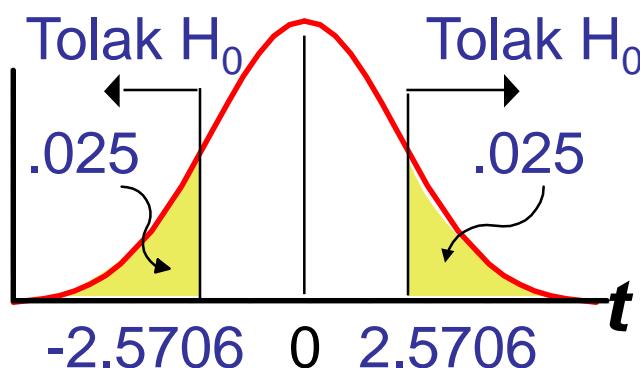
- $H_0: \beta_1 = 0$

- $H_1: \beta_1 \neq 0$

$$\alpha = .05$$

$$db = 7 - 2 = 5$$

Nilai-nilai kritis :



Statistik Uji-t :

| | <i>t Stat</i> | <i>P-value</i> |
|------------|---------------|----------------|
| Intercept | 3.6244333 | 0.0151488 |
| X Variable | 9.009944 | 0.0002812 |

Keputusan:

Tolak H_0

Kesimpulan:

Terbukti ada hubungan. Makin luas ukuran Toko, makin tinggi penjualannya

Selang Kepercayaan Slope

$$b_1 \pm t_{n-2} S_{b1}$$

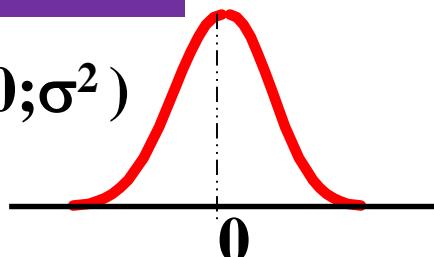
Output Excel masalah *Produce Stores*

| | <i>Lower 95%</i> | <i>Upper 95%</i> |
|------------|------------------|------------------|
| Intercept | 475.810926 | 2797.01853 |
| X Variable | 1.06249037 | 1.91077694 |

**95% yakin nilai slope antara 1.062 s/d 1.911.
(Selang Kepercayaan ini tdk mencakup nilai 0)**

Kesimpulan: Ada hubungan linear yg nyata antara penjualan tahunan dgn ukuran toko.

(i) $e_i \sim N(0; \sigma^2)$



Asumsi Model Regresi Linear: Peubah acak ε_i menyebar **Normal**, bebas dan **identik** utk $i=1, \dots, n$.

e_i

- **Bebas:** $Cov(\varepsilon_t, \varepsilon_s) = E(\varepsilon_t \varepsilon_s) = 0$ untuk $t \neq s$.
- **Homoskedastisitas:** $Var(\varepsilon_i) = E(\varepsilon_i^2) = \sigma^2$.

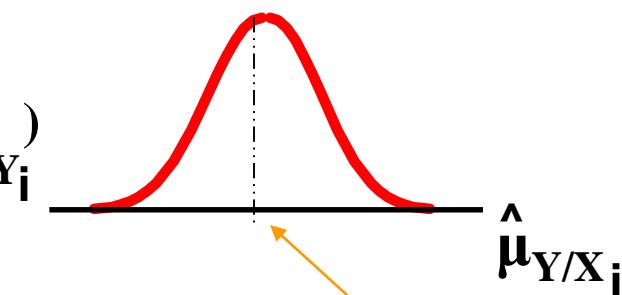
(ii) **X fixed variable**

➤ Dugaan Koefisien β_i dengan OLS bersifat **TAK BIAS dgn RAGAM MINIMUM (Best Linear Unbiased Estimator)**, dan menyebar **Normal**.

Simple Linear Model $Y_i = b_0 + b_1 X_i + e_i$

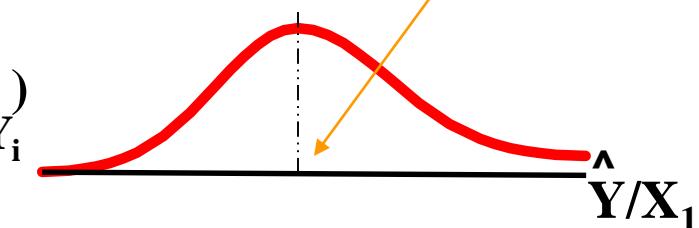
➤ Dugaan Rata2 Y utk X_i tertentu menyebar **Normal**

$$\hat{\mu}_{Y/X_i} \sim N(\beta_0 + \beta_1 X_i; \sigma_{\mu_{Y_i}}^2)$$



➤ Dugaan Individu Y utk X_i ttt sama dgn dugaan rata2nya, juga menyebar **Normal**, dgn ragam lebih besar

$$\hat{Y}/X_i \sim N(\beta_0 + \beta_1 X_i; \sigma_{Y_i}^2)$$



$$\beta_0 + \beta_1 X_i$$

Dugaan Selang Nilai-Nilai Ramalan

Selang kepercayaan bagi μ_{YX} ,
Rataan Y utk X_i tertentu

Standard error
dugaan

$$\hat{Y}_i \pm t_{n-2} \bullet S_{yx}$$

nilai t dari tabel dgn
 $db=n-2$

Selang bervariasi sesuai jaraknya
terhadap rataan, \bar{X} .

$$\sqrt{\frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2}}$$

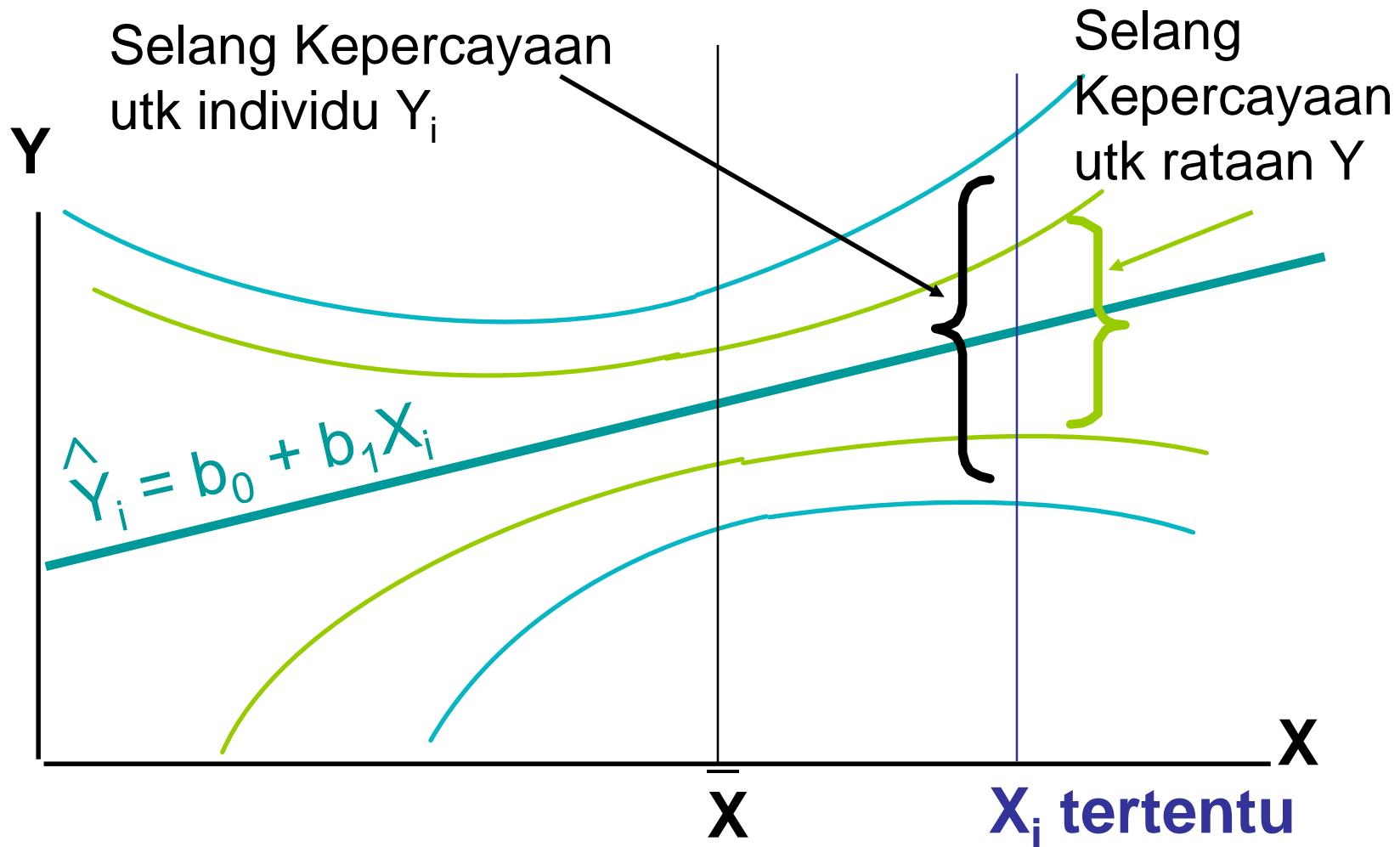
Dugaan Selang Nilai-Nilai Ramalan

Selang Kepercayaan bagi Dugaan
Respons individu Y_i utk X_i tertentu

Tambahan 1 ini membuat selangnya lebih lebar dari
SK bagi rataan Y , μ_{XY}

$$\hat{Y}_i \pm t_{n-2} \cdot S_{yx} \sqrt{1 + \frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2}}$$

Dugaan Selang utk Nilai-nilai X yang Berbeda



Data for 7 Toko:

| Toko | Square Feet | Annual Sales (\$000) |
|------|-------------|----------------------|
| 1 | 1,726 | 3,681 |
| 2 | 1,542 | 3,395 |
| 3 | 2,816 | 6,653 |
| 4 | 5,555 | 9,543 |
| 5 | 1,292 | 3,318 |
| 6 | 2,208 | 5,563 |
| 7 | 1,313 | 3,760 |

Dugalah penjualan tahunan utk suatu toko berukuran 2000 *square feet*.

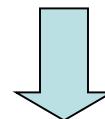
Model Regresi yg diperoleh:

$$\hat{Y}_i = 1636.415 + 1.487X_i$$

$$\hat{Y}_i = 1636.415 + 1.487 (2000)$$

Dugaan Penjualan = 4610.45 (\$000)

Seberapa besar kemungkinan kesalahan dari dugaan ini??



Tk Keyakinan bhw nilai sebenarnya berada dlm selang dugaan
→ Selang kepercayaan $(1-\alpha)100\%$ bagi nilai sebenarnya

Dugaan Selang Ramalan Rataan Y

Dugaan Selang Kepercayaan bagi μ_{XY}

Tentukan SK 95% bagi rata-rata penjualan tahunan utk toko berukuran 2,000 square feet

Dugaan Sales $\hat{Y}_i = 1636.415 + 1.487X_i = 4610.45 (\$000)$

$$\bar{X} = 2350.29 \quad S_{YX} = 611.75 \quad t_{n-2} = t_5 = 2.5706$$



$$\hat{Y}_i \pm t_{n-2} \cdot S_{yx} \sqrt{\frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2}} = 4610.45 \pm 980.97$$

SK bagi rataan Y

Dugaan Selang Ramalan Individu Y

Selang kepercayaan utk dugaan Individu Y

Tentukan SK 95% bagi penjualan tahunan utk suatu toko berukuran 2,000 *square feet*

$$\text{Ramalan Sales } \hat{Y}_i = 1636.415 + 1.487X_i = 4610.45 (\$000)$$

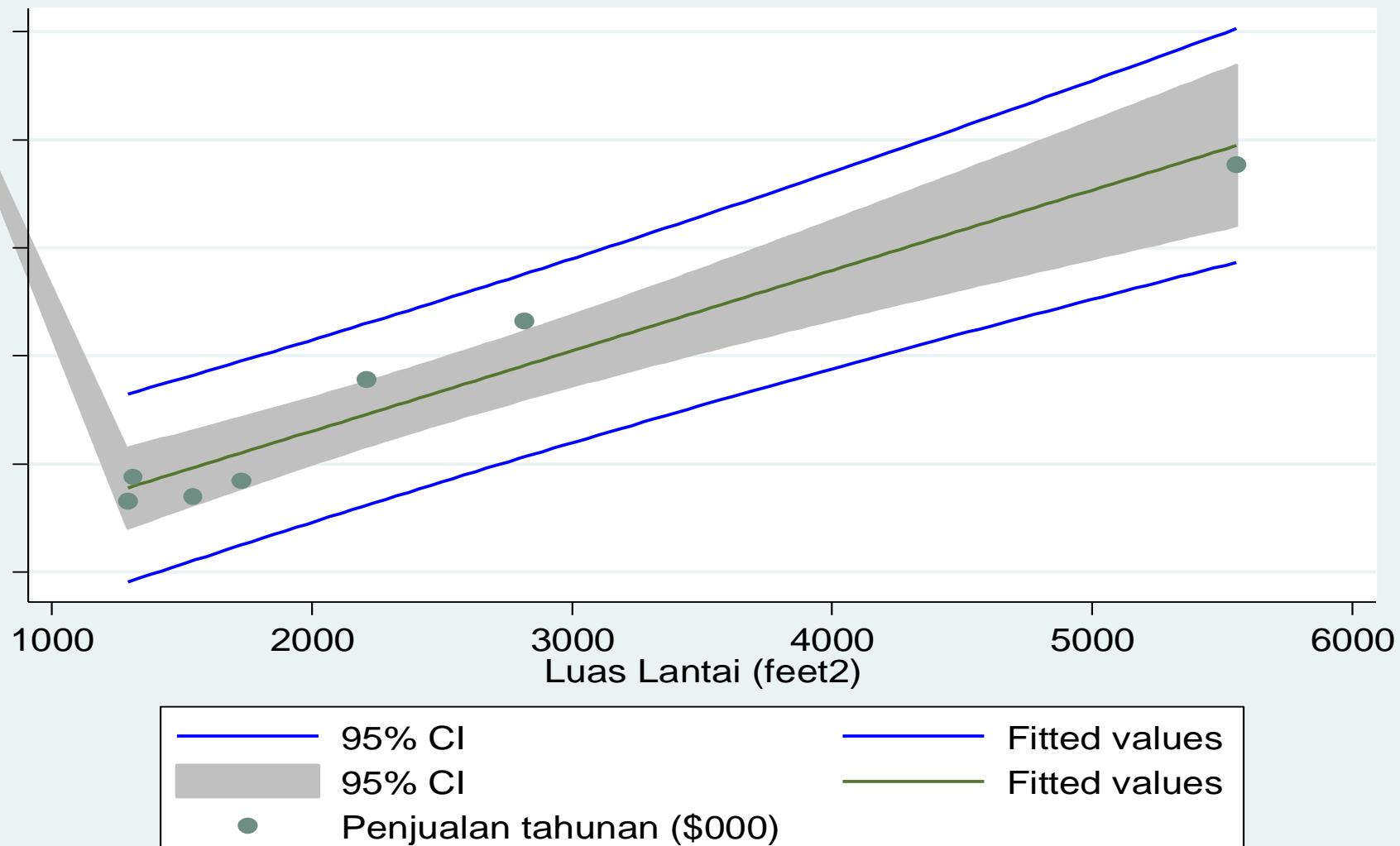
$$\bar{X} = 2350.29 \quad S_{YX} = 611.75 \quad t_{n-2} = t_5 = 2.5706$$



$$\hat{Y}_i \pm t_{n-2} \cdot S_{yx} \sqrt{1 + \frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2}} = 4610.45 \pm 1853.45$$

SK bagi individu Y

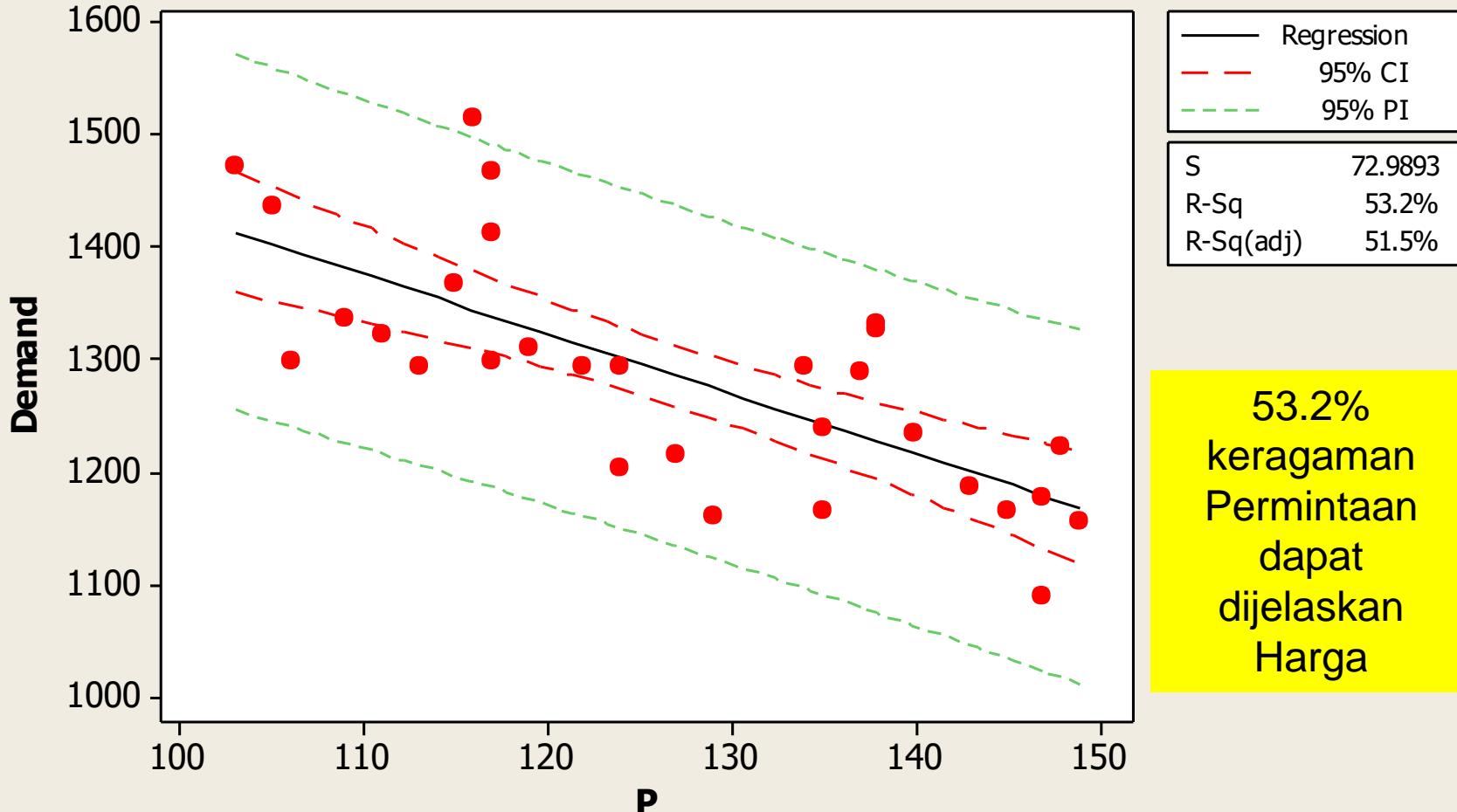
Dugaan Selang Ramalan Penjualan (rata-rata & Individu) (tanpa memperhatikan faktor-faktor lain)



Jika Harga (P) naik maka Permintaan akan turun
(tanpa memperhatikan faktor2 lain)

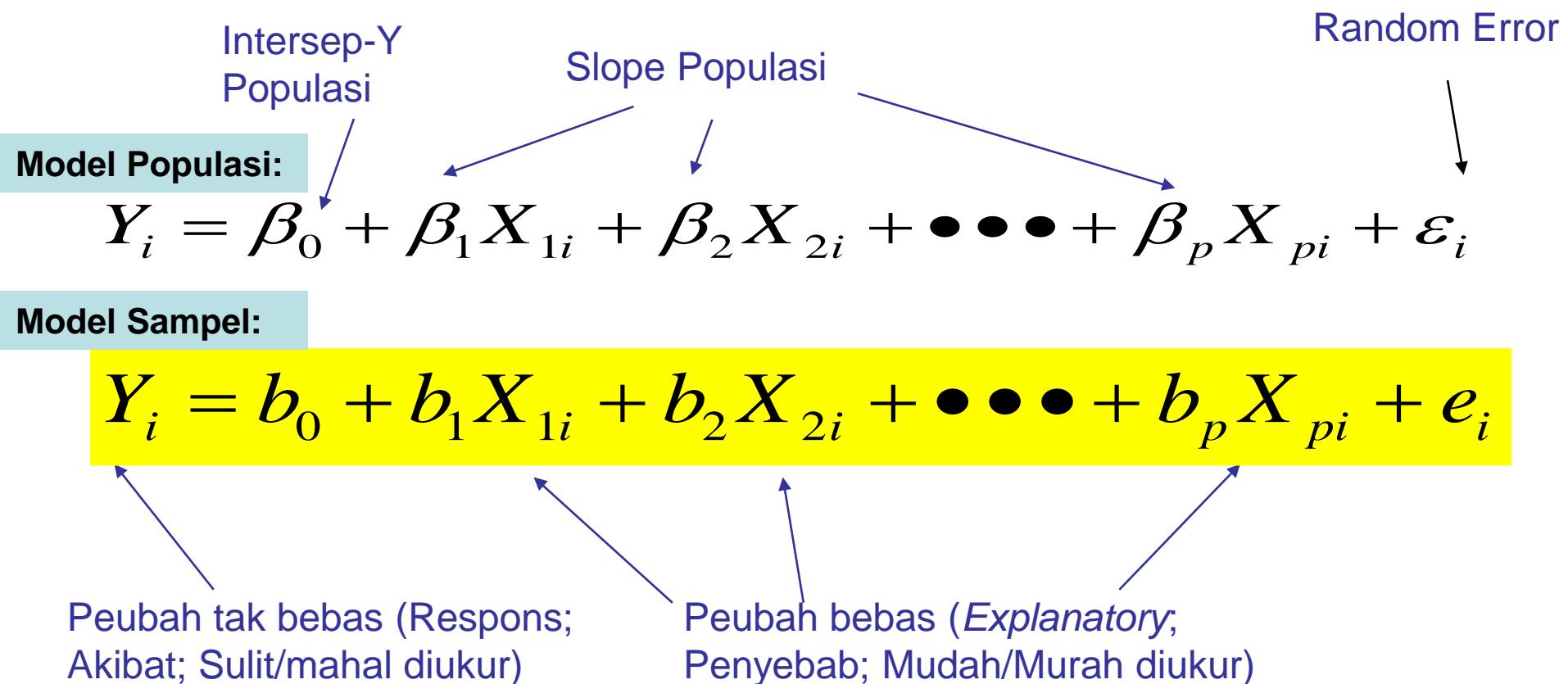
Fitted Line Plot

$$\text{Demand} = 1964 - 5.337 P$$



Model Regresi Berganda

Hubungan linear (dlm parameter) antara 1 peubah tak bebas dgn 2 atau lebih peubah bebas



Ilustrasi:

Data permintaan produk detergent baru selama 30 minggu terakhir setelah dipasarkan pertama kali, beserta faktor-faktor yang diperkirakan mempengaruhinya. Buatlah dugaan model permintaan

| Minggu Ke (i) | Demand (unit), Y | Harga (\$), X ₂ | Harga Kompetitor, X ₃ | Biaya Iklan, X ₄ | Income (\$), X ₅ | Periode (X ₆) |
|---------------|------------------|----------------------------|----------------------------------|-----------------------------|-----------------------------|---------------------------|
| 1 | 1290 | 137 | 94 | 814 | 42498 | 1 |
| 2 | 1177 | 147 | 81 | 896 | 41399 | 2 |
| 3 | 1155 | 149 | 89 | 852 | 39905 | 3 |
| 4 | 1299 | 117 | 92 | 854 | 34871 | 4 |
| 5 | 1166 | 135 | 86 | 810 | 34239 | 5 |
| 6 | 1186 | 143 | 79 | 768 | 44452 | 6 |
| 7 | 1293 | 113 | 91 | 978 | 30367 | 7 |
| 8 | 1322 | 111 | 82 | 821 | 37757 | 8 |
| 9 | 1338 | 109 | 81 | 843 | 40130 | 9 |
| 10 | 1160 | 129 | 82 | 849 | 31264 | 10 |
| 11 | 1293 | 124 | 91 | 797 | 34610 | 11 |
| 12 | 1413 | 117 | 76 | 988 | 41033 | 12 |
| 13 | 1299 | 106 | 90 | 914 | 30674 | 13 |
| 14 | 1238 | 135 | 88 | 913 | 31578 | 14 |
| 15 | 1467 | 117 | 99 | 867 | 41201 | 15 |
| 16 | 1089 | 147 | 76 | 785 | 30247 | 16 |
| 17 | 1203 | 124 | 83 | 817 | 33177 | 17 |
| 18 | 1474 | 103 | 98 | 846 | 37330 | 18 |
| 19 | 1235 | 140 | 78 | 768 | 44671 | 19 |
| 20 | 1367 | 115 | 83 | 856 | 37950 | 20 |
| 21 | 1310 | 119 | 76 | 771 | 43478 | 21 |
| 22 | 1331 | 138 | 100 | 947 | 36053 | 22 |
| 23 | 1293 | 122 | 90 | 831 | 35333 | 23 |
| 24 | 1437 | 105 | 86 | 905 | 44304 | 24 |
| 25 | 1165 | 145 | 96 | 996 | 30925 | 25 |
| 26 | 1328 | 138 | 97 | 929 | 36867 | 26 |
| 27 | 1515 | 116 | 97 | 1000 | 41799 | 27 |
| 28 | 1223 | 148 | 84 | 951 | 40684 | 28 |
| 29 | 1293 | 134 | 88 | 848 | 43637 | 29 |
| 30 | 1215 | 127 | 87 | 891 | 30468 | 30 |

Contoh Output STATA (Bgm Interpretasinya?)

. regress Demand P Pr Adv Income T

| Source | SS | df | MS | Number of obs | = | 30 |
|----------|------------|----|------------|---------------|---|--------|
| Model | 291363.592 | 5 | 58272.7183 | F(5, 24) | = | 51.49 |
| Residual | 27159.2084 | 24 | 1131.63368 | Prob > F | = | 0.0000 |
| Total | 318522.8 | 29 | 10983.5448 | R-squared | = | 0.9147 |
| | | | | Adj R-squared | = | 0.8970 |
| | | | | Root MSE | = | 33.64 |

| Demand | Coef. | Std. Err. | t | P> t | [95% Conf. Interval] |
|--------|----------|-----------|--------|-------|----------------------|
| P | -5.03264 | .4393547 | -11.45 | 0.000 | -5.939423 -4.125856 |
| Pr | 4.743503 | .9696136 | 4.89 | 0.000 | 2.742319 6.744687 |
| Adv | .2774435 | .104598 | 2.65 | 0.014 | .0615638 .4933232 |
| Income | .0106579 | .0013158 | 8.10 | 0.000 | .0079421 .0133736 |
| T | 1.309092 | .7533593 | 1.74 | 0.095 | -.2457651 2.863949 |
| _cons | 850.0416 | 134.7807 | 6.31 | 0.000 | 571.8678 1128.215 |

Jika Selang Kepercayaan tidak mencakup 0, maka terima $H_1: \beta_j \neq 0$

I. Uji Model Secara Keseluruhan

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon_i$$

• Apakah Model dapat menjelaskan keragaman Y?

• Hipotesis Statistik:

H_0 : Model tdk dpt menjelaskan keragaman Y

$$\sigma^2_{\text{regresi}} = \sigma^2_{\varepsilon} \text{ atau } \sigma^2_{\text{regresi}} / \sigma^2_{\varepsilon} = 1$$

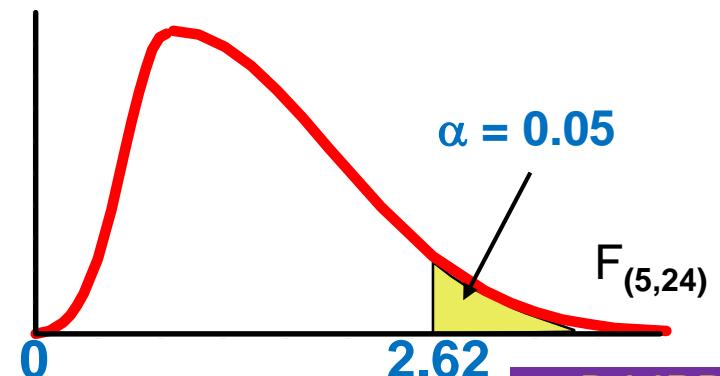
$\beta_1 = \beta_2 = \dots = \beta_p = 0$ (tidak dapat menjelaskan)

H_1 : Model dpt menjelaskan keragaman Y

$$\sigma^2_{\text{regresi}} > \sigma^2_{\varepsilon} \text{ atau } \sigma^2_{\text{regresi}} / \sigma^2_{\varepsilon} > 1$$

Minimal ada $\beta_i \neq 0$ (ada peubah bebas yg mempengaruhi Y)

• Statistik uji-F = KTR/KTS $\sim F_{(p, n-1-p)}$



Uji Model Secara Keseluruhan

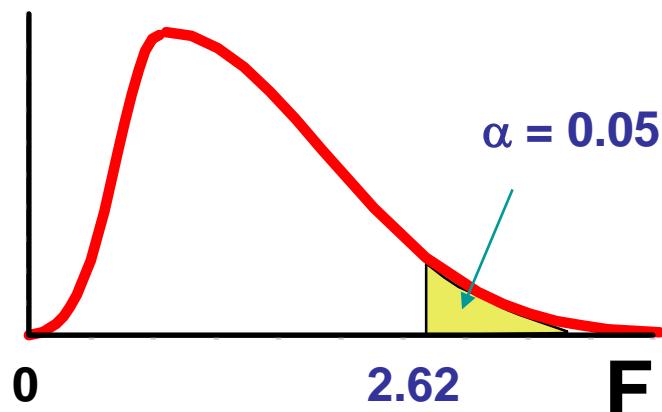
$H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$

$H_1: \text{Minimal ada } \beta_j \neq 0$

$\alpha = .05$

$db = 5 \text{ dan } 24$

Nilai Kritis:



Statistik Uji:

$F = 51.49 ; p=0.00$
(Output software)

Keputusan:

H_1 pada $\alpha = 0.05$

Kesimpulan:

- Model dpt menjelaskan keragaman Y
- Minimal ada satu peubah bebas yg mempengaruhi Y

Gunakan kriteria p-value !

JKT = Jumlah Kuadrat Total

$$JKT = \sum(Y_i - \bar{Y})^2$$

- mengukur keragaman nilai-nilai Y_i sekitar rataan \bar{Y}

JKR = Jumlah Kuadrat Regresi

$$JKR = \sum(\hat{Y}_i - \bar{Y})^2$$

- Menjelaskan keragaman yg dpt dianggap berasal dari hubungan antara X dgn Y (*model regresi*)

JKS = Jumlah Kuadrat Sisa (error)

$$JKS = \sum(Y_i - \hat{Y}_i)^2$$

- Keragaman yg dpt dianggap berasal dari faktor-faktor selain hubungan antara X dgn Y

Tabel
ANOVA

| | db | JK |
|---------|----|--------|
| Regresi | 5 | 291364 |
| Sisa | 24 | 27159 |
| Total | 29 | 318523 |

JKR

JKS

JKT

Koefisien

Determinasi:

$$r^2 = \frac{JKR}{JKT} = \frac{\text{Jumlah Kuadrat Regresi}}{\text{Jumlah Kuadrat Total}}$$

- Mengukur “proporsi keragaman” yg dijelaskan oleh (peubah bebas X dlm) model regresi
- Sering secara “informal” sbg ukuran *goodness-of-fit* utk membandingkan validitas bbrp spesifikasi model
- 91.5% keragaman permintaan dpt dijelaskan oleh model

$S = 33.6398$ $R-Sq = 91.5\%$ $R-Sq(\text{adj}) = 89.7\%$

Analysis of Variance

| Source | DF | SS | MS | F | P |
|----------------|----|--------|-------|-------|-------|
| Regression | 5 | 291364 | 58273 | 51.49 | 0.000 |
| Residual Error | 24 | 27159 | 1132 | | |
| Total | 29 | 318523 | | | |

II. Uji Signifikansi Masing² Peubah Bebas

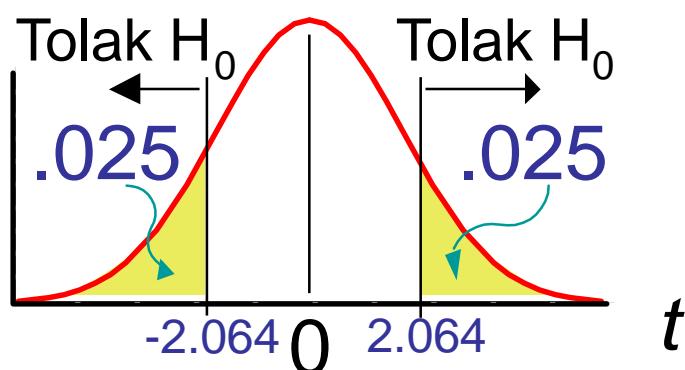
Apakah harga berpengaruh nyata terhadap permintaan Detergen? Ujilah pada $\alpha = 0.05$.

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0$$

$$db = 24$$

Nilai Kritis:



Statistik Uji: $t = -11.45$, $p=0.000$

Keputusan: Tolak H_0 pada $\alpha = 0.05$

Kesimpulan:

Terbukti nyata, pengaruh harga pada permintaan detergen.

Gunakan kriteria p-value !

Taraf Nyata, α dan Daerah Penolakan

Hipotesis Penelitian



Hipotesis statistika (yg mana?)

$$H_0: \beta_1 \geq k$$

$$H_1: \beta_1 < k$$

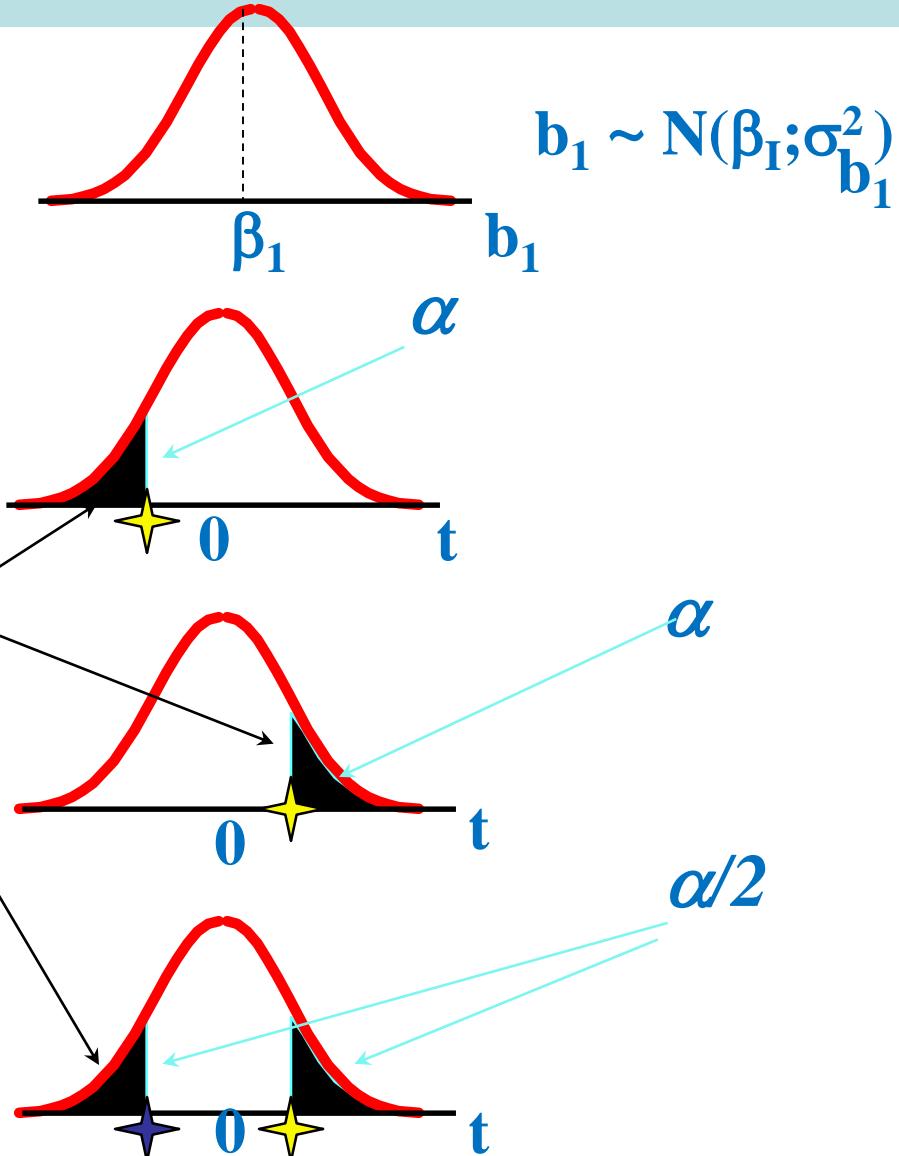
$$H_0: \beta_1 \leq k$$

$$H_1: \beta_1 > k$$

$$H_0: \beta_1 = k$$

$$H_1: \beta_1 \neq k$$

**Daerah
Penolakan
(ttk kritis)**



Untuk uji eka-arah, misalnya $H_1: \beta_1 > 0$,
nilai-p sebenarnya adalah $P(|t| > t_{hit}) = \frac{1}{2} p\text{-value}$ output komputer

II. Uji Eka Arah

Apakah ada kenaikan permintaan tiap minggu?
Ujilah pada $\alpha = 0.05$.

Hipotesis Penelitian: **Ada kenaikan permintaan tiap minggu**



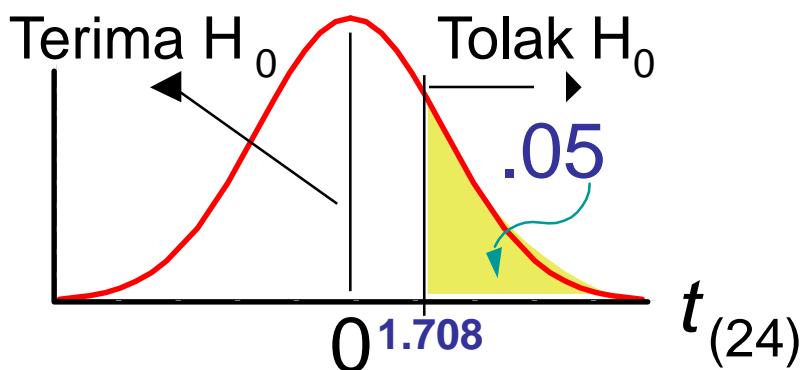
Hipotesis statistika (yg mana?)

$$H_0: \beta_T \leq 0 \text{ atau } \beta_T = 0$$

$$H_1: \beta_T > 0$$

$$db = 24$$

Nilai Kritis:



Gunakan kriteria p-value !

Statistik Uji: $t = 1.74$, $p=0.0475$

Keputusan: Tolak H_0 pada $\alpha = 0.05$

Kesimpulan:

Terbukti nyata, Ada kenaikan permintaan detergen tiap minggu.

Jika Kesimpulan H_0 : tdk ada alasan utk menolak H_0 atau belum cukup bukti menyimpulkan H_1

PR:

- Apakah jika harga deterjen naik 1 dolar maka permintaannya akan turun lebih dari 3 unit?.
- Apakah jika harga deterjen naik 1 dolar maka permintaannya akan turun minimal 4 unit?

Analisis Sisaan (*Residual*)

- Tujuan
 - Mengkaji Linearitas
 - Evaluasi pelanggaran asumsi
- Analisis Sisaan dgn Grafik
 - Plot sisaan Vs. nilai-nilai X_i atau \hat{Y}_i
 - (e_i, X_i) atau (e_i^2, X_i) atau $(e_i / s_e, X_i)$
 - *Standardized residuals*: $= e_i / s_e$
 - Memungkinkan mempertimbangkan besaran sisaan (sisaan-baku spt Normal baku) utk melihat adanya pencilan (*outlier*)

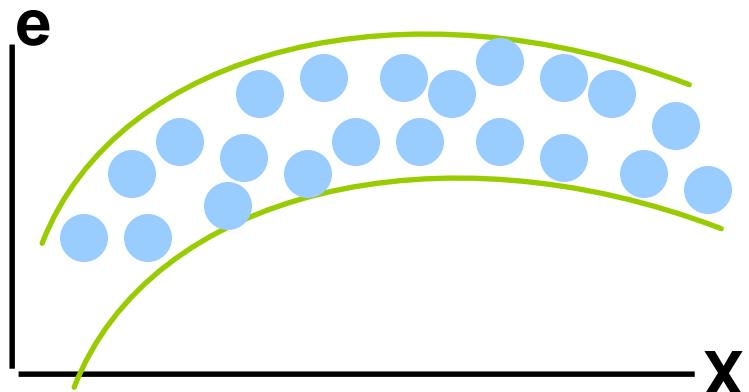
Perbedaan Pendekatan Regresi Klasik dengan Ekonometrika Deret Waktu

1. Pendekatan Regressi klasik: **Menduga model dulu, kemudian dilihat (diuji) apakah asumsi tentang error (ϵ) dipenuhi (ragam homogen/sama, dan tidak ada autokorelasi).** Dalam konteks data deret waktu, error tsb bersifat **Stasioner**.
 2. Pendekatan (terkini) Regressi Deret Waktu: **Data harus stasioner dulu, kemudian baru diduga modelnya.** Penentuan ordo/lag, juga dugaan parameternya, dari data yang sudah stasioner.
Jika dipaksakan pada data deret waktu yg belum stasioner, analisisnya “dapat” menyesatkan. Namun jika errornya memenuhi asumsi klasik atau stasioner, model tsb tetap valid.
- Faktanya, hampir semua data deret waktu bersifat tidak stasioner.
 - Ekonometrika menggunakan data deret waktu perlu ditangani dan dianalisis secara berbeda

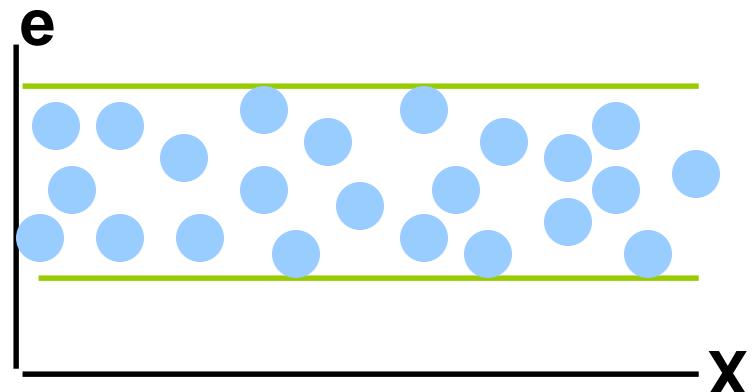
Analisis Sisaan utk Linearitas



Not Linear



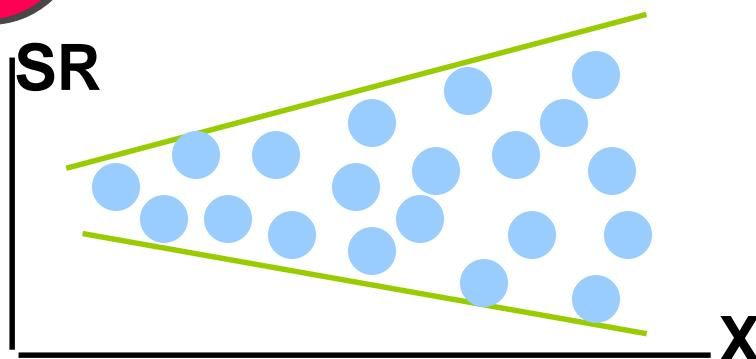
Linear



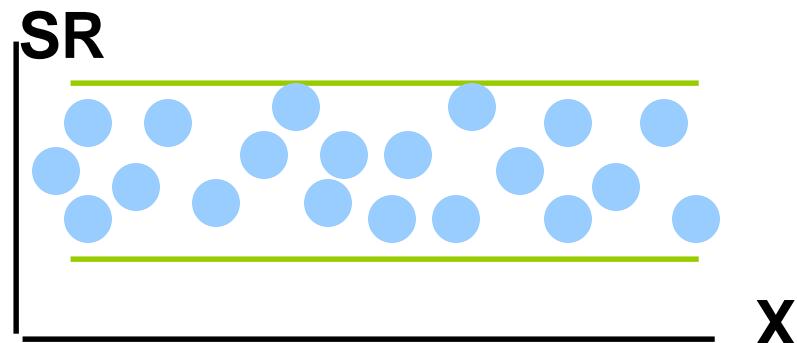
Analisis Sisaan utk Homoskedastisitas



Heteroskedastisitas



Homoscedasticity

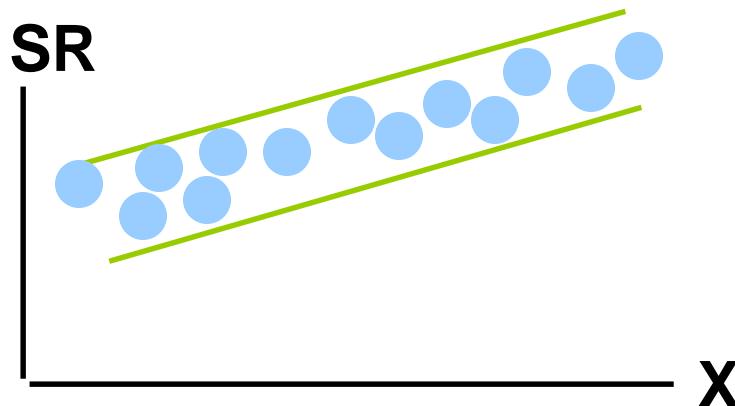


Menggunakan ***Standardized Residuals (SR)***

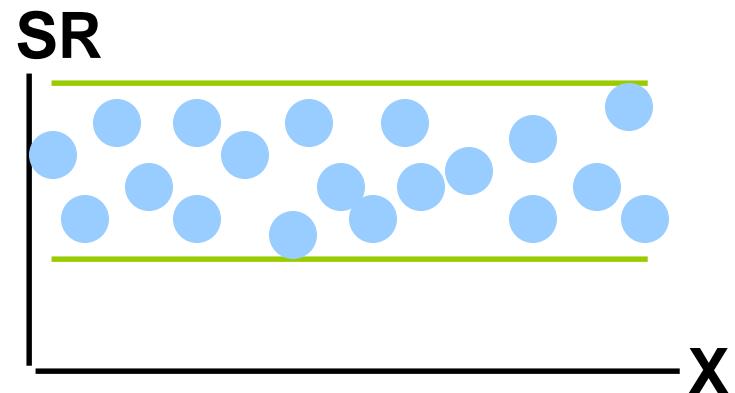
Analisis Sisaan utk Kebebasan e

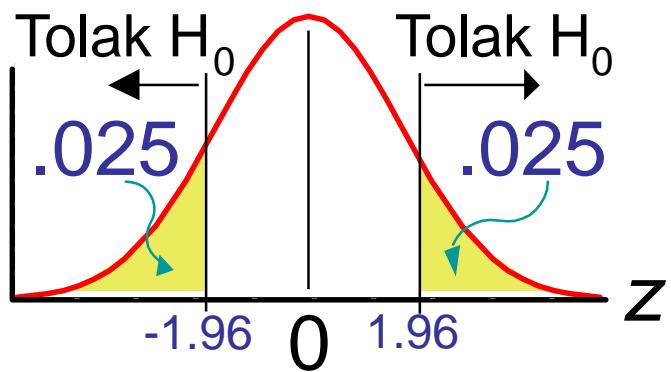


Tidak Bebas

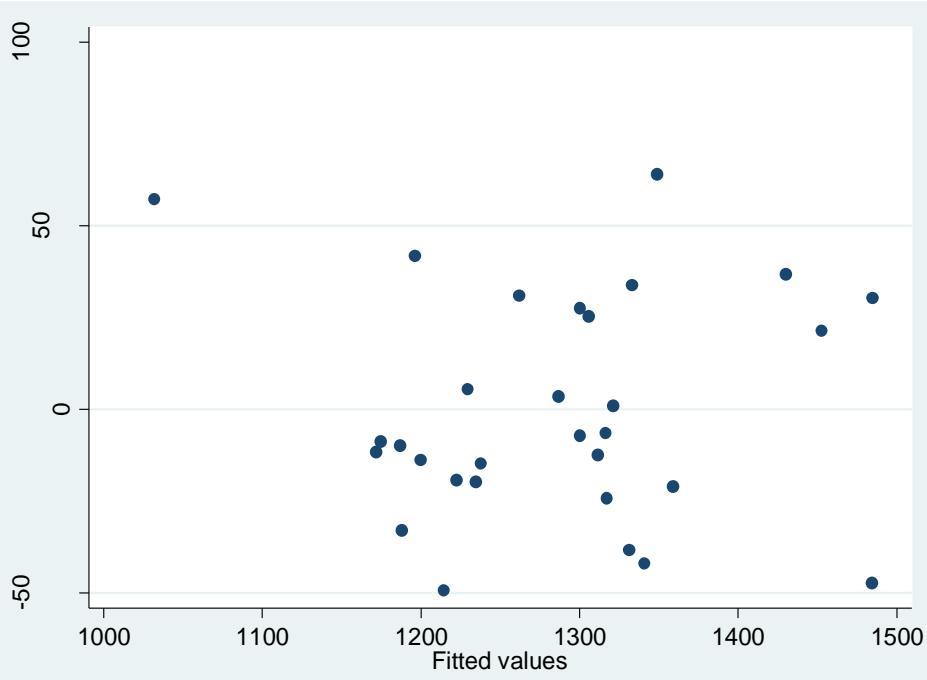
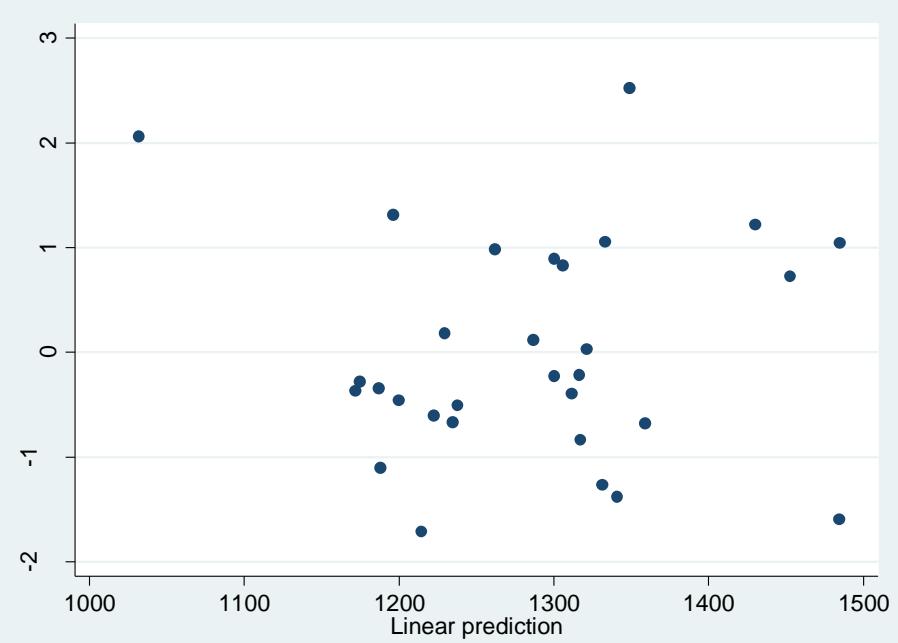


Bebas





Plot sisaan (juga sisaan baku)
dgn dugaan Y.
Apakah ada *outlier* (pencilan)?



Statistik Durbin-Watson

- Digunakan utk data *time series* guna mendekripsi autokorelasi (Sisaan dlm suatu periode berhubungan dgn sisaan dlm periode lain)
- Mengukur Pelanggaran asumsi kebebasan e

$$D = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2}$$

Seharusnya mendekati 2.

Jika tidak, kaji model utk autokorelasi.