

PEMODELAN EKONOMETRIKA

Oleh

Bambang Juanda

<https://bbjuanda.wordpress.com/>

<https://bambangjuanda.com/>

Tujuan Pemodelan

1. menduga **hubungan**-hubungan ekonomi. Misal, Bagaimana dampak DAU Terhadap Ketimpangan Pendapatan?
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 pendapatan dan belanja negara, inflasi, dan pengangguran, serta defisit anggaran.

Pengertian Model & Tujuan Pemodelan

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

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

$$\text{data aktual} = \text{dugaan} + \text{sisaan (simpangan)}$$

$$\text{data} = \text{komp. sistematis} + \text{komp. non-sistematis}$$

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

diharapkan unsur-unsur ketidak-teraturan nilai Y dapat dijelaskan oleh nilai-nilai dari peubah X_1, X_2, \dots , dan 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 X_1 , X_2 , ..., 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 Berganda

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

Intersep-Y
Populasi

Slope Populasi

Random Error

Model Populasi:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi} + \varepsilon_i$$

Model Sampel:

$$Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + \dots + b_p X_{pi} + e_i$$

Peubah tak bebas (Respons;
Akibat; Sulit/mahal diukur)

Peubah bebas (*Explanatory*;
Penyebab; Mudah/Murah diukur)

Model Regresi Linear Sederhana

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

Model Populasi:

intersep Y Slope Error Acak

Peubah Respons (*dependent*) Peubah Penjelas (*Independent*)

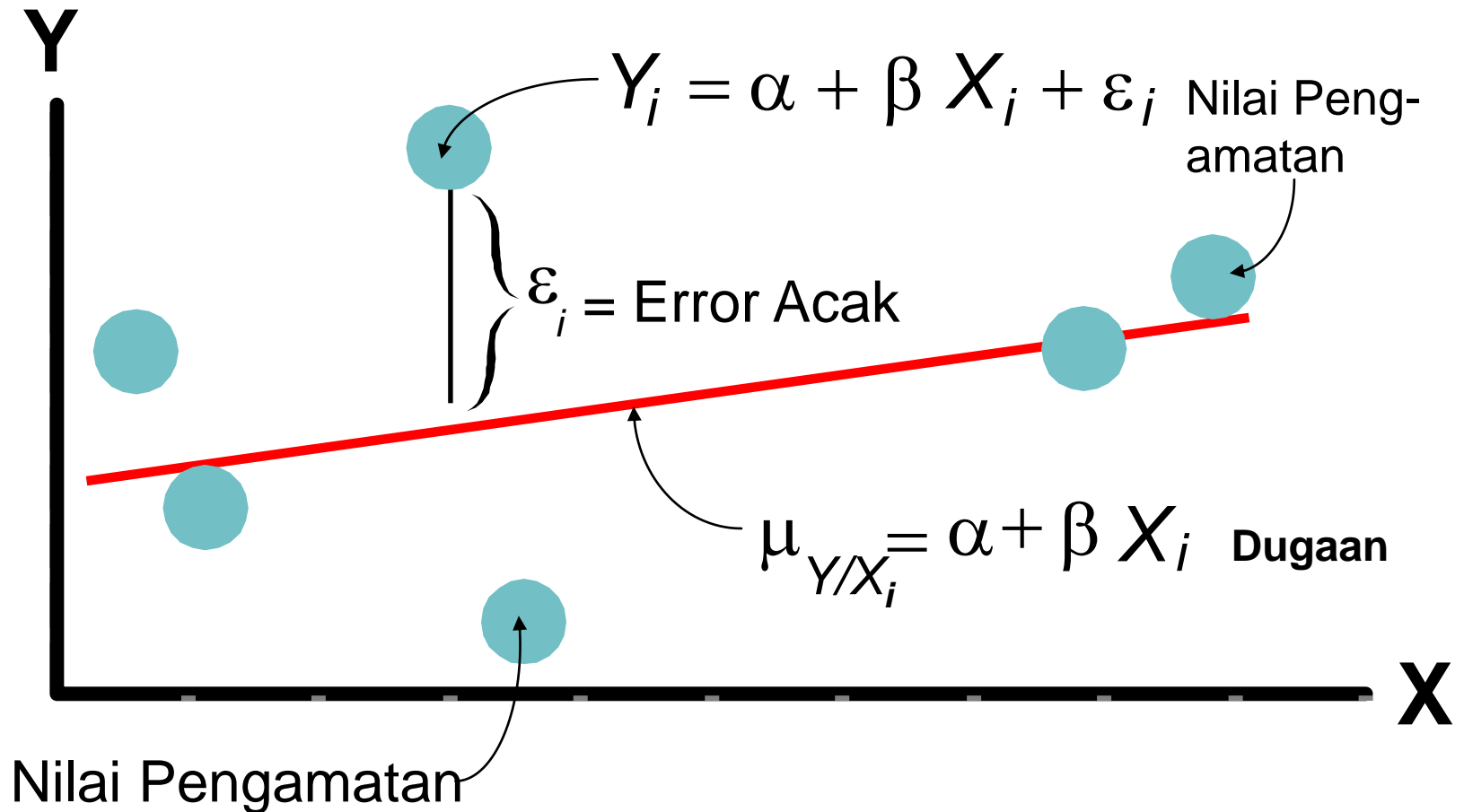
akibat; sulit atau mahal diukur penyebab; mudah atau murah diukur

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

Model Regresi Contoh:

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

Model Regresi Linear Populasi



Metode (Jumlah) Kuadrat (Sisaan) Terkecil: MKT atau *Ordinary Least Squares*. $Y=a+bX+e$

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

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

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

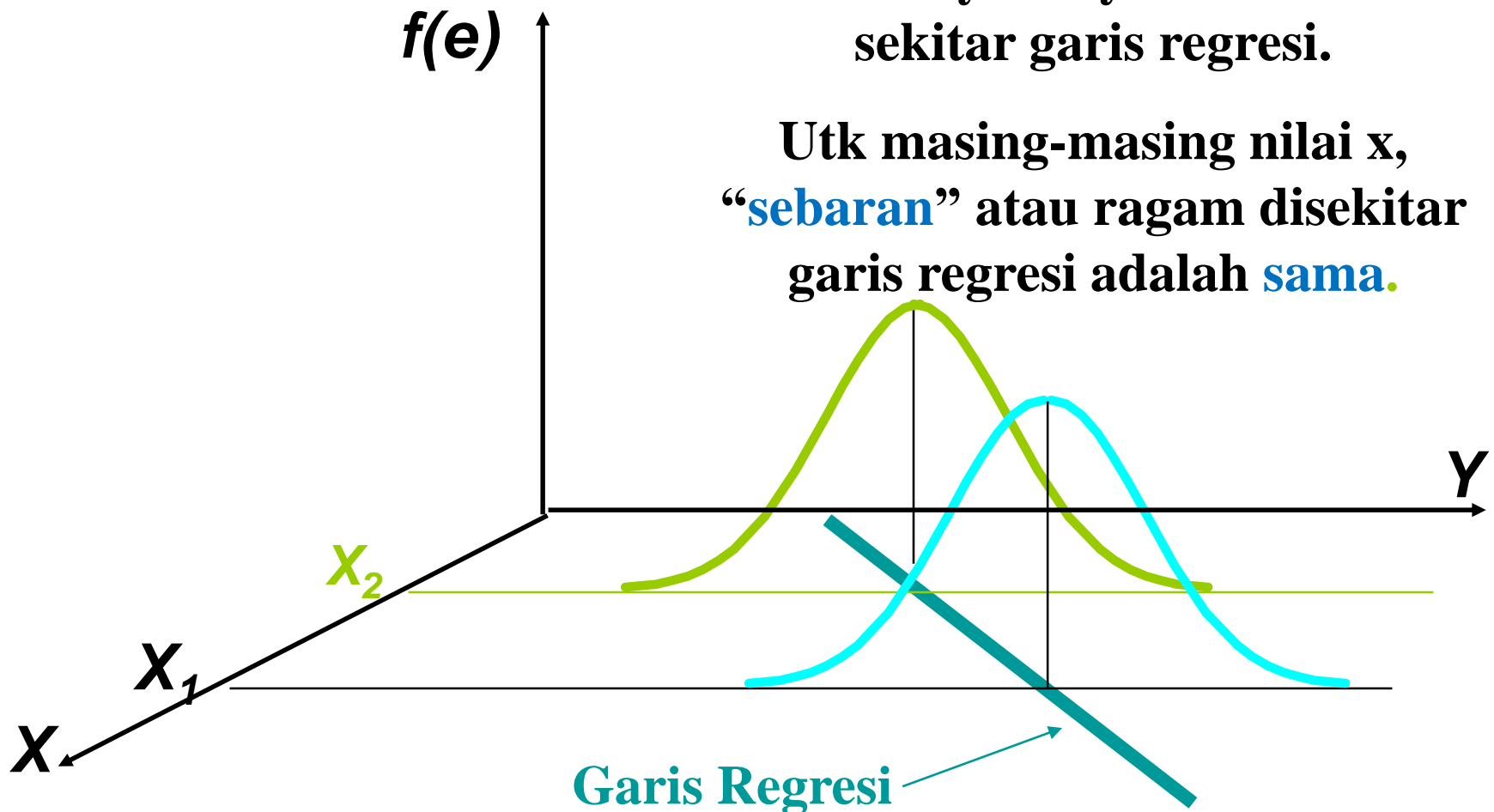
$$\hat{\beta} = b = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{n \sum_{i=1}^n X_i Y_i - \sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{n \sum_{i=1}^n X_i^2 - \left(\sum_{i=1}^n X_i \right)^2}$$

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

Ragam *Error* Sekitar Garis Regresi

Nilai-nilai y menyebar **normal** di sekitar garis regresi.

Utk masing-masing nilai x ,
“**sebaran**” atau ragam disekitar garis regresi adalah **sama**.



Asumsi Model Regresi Linear

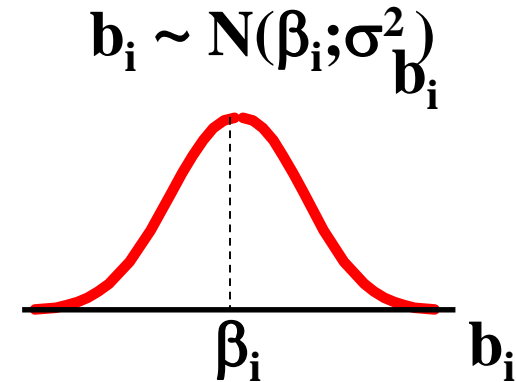
- **Kenormalan & Kebebasan**
 - Nilai-nilai Y **Menyebar Normal** utk masing-masing nilai X ; dgn $E(Y_i) = \alpha + \beta X_i$ dan $\text{Var}(Y_i) = \sigma^2$ utk semua i .
 - (i) **Sebaran Peluang *Error* adalah Normal, Bebas dan Identik** dengan $E(\varepsilon_i) = 0$ dan $\text{Var}(\varepsilon_i)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

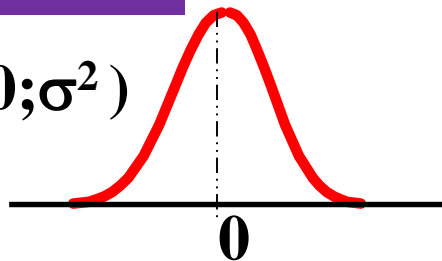


$$\sigma_{b_0}^2 = \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right] \sigma^2$$

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

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

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



□ **Bebas**: $Cov(\varepsilon_t, \varepsilon_s) = E(\varepsilon_t \varepsilon_s) = 0$ untuk $t \neq s$.

□ **Homoskedastisitas**: $Var(\varepsilon_i) = E(\varepsilon_i^2) = \sigma^2$.

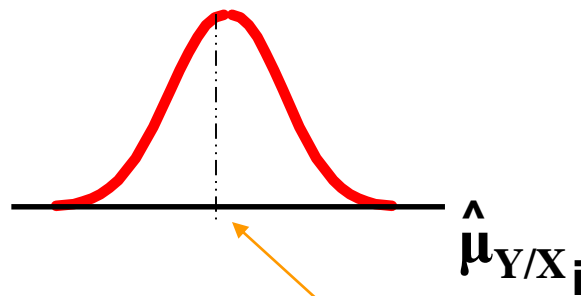
(ii) X fixed variable

➤ 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**

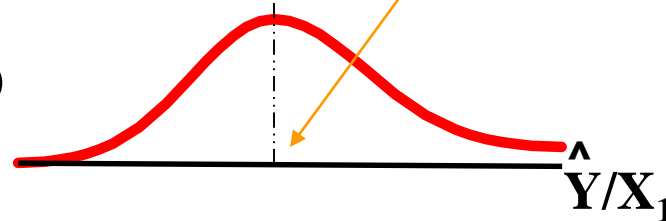
$\hat{\mu}_{Y/X_i} \sim N(\beta_0 + \beta_1 X_i; \sigma_{\mu_{Y_i}}^2)$



$\beta_0 + \beta_1 X_i$

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

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



Dugaan Selang Nilai-Nilai Ramalan

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

*Standard error
dugaan*

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

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

nilai t dari tabel dgn db=n-2

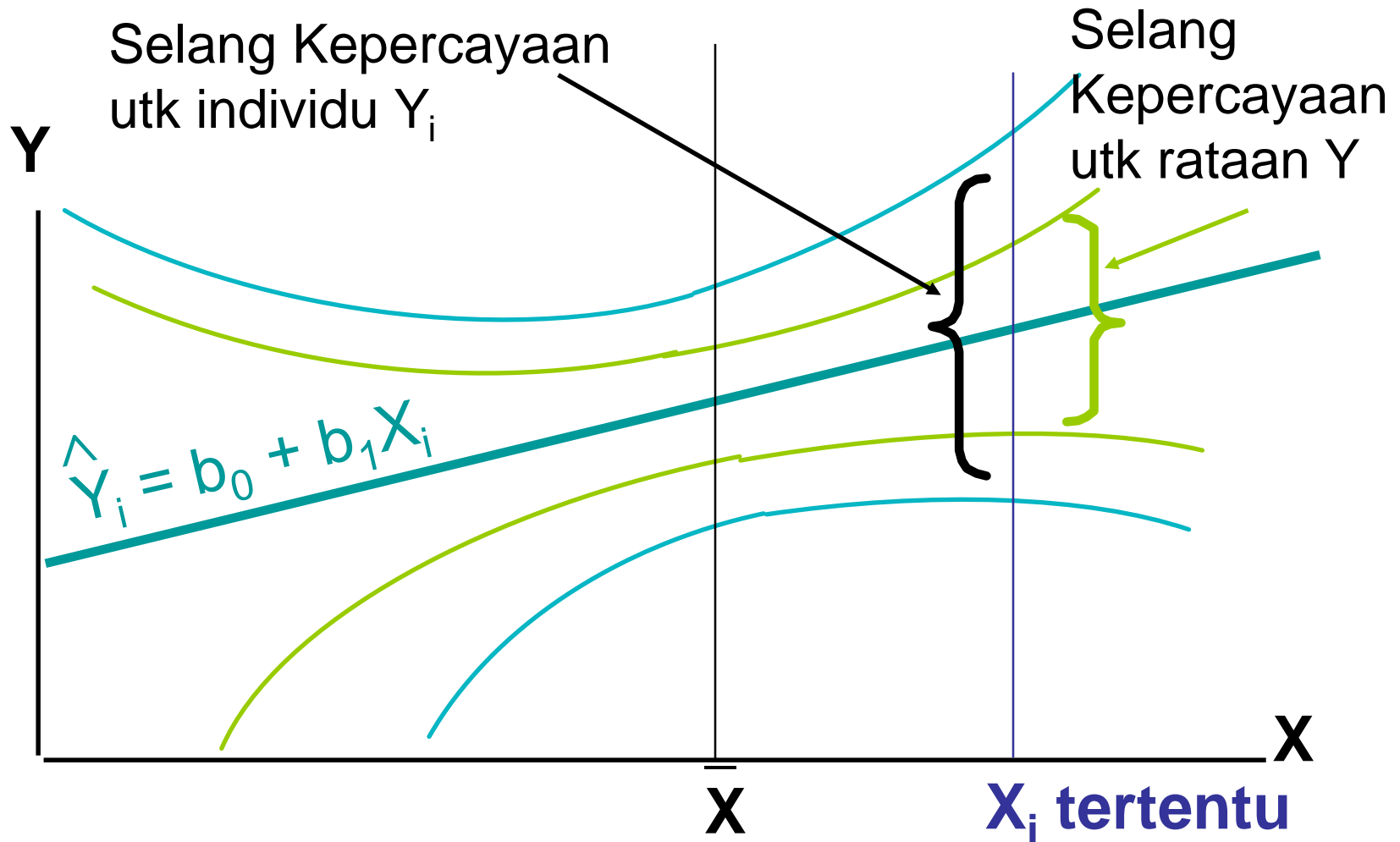
Dugaan Selang Nilai-Nilai Ramalan

Selang Kepercayaan bagi Dugaan
Respons individu Y_i utk X_i tertentu

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

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

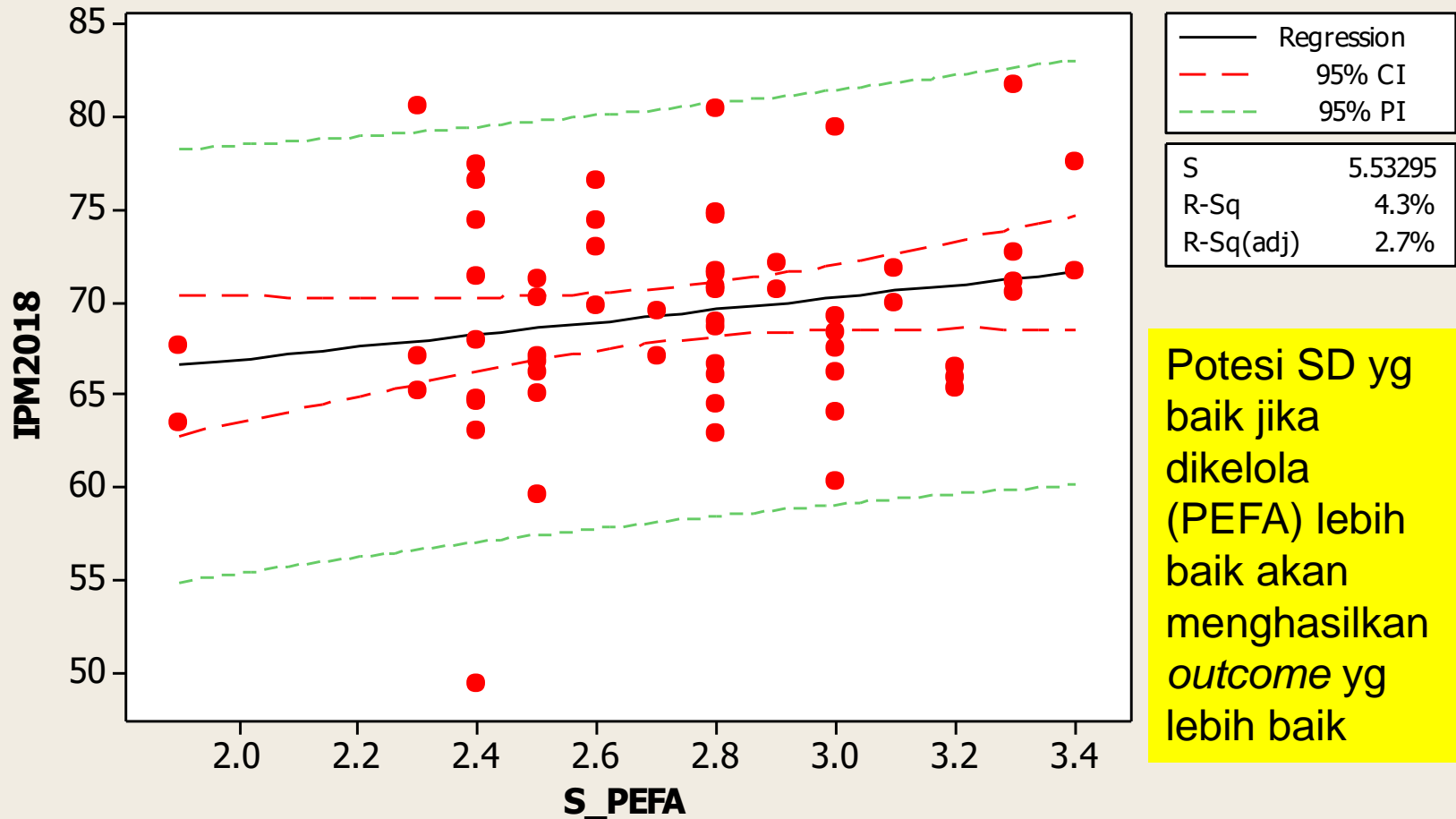
Dugaan Selang utk Nilai-nilai X yang Berbeda



Jika PEFA naik maka IPM akan naik (tanpa memperhatikan faktor2 lain)
 4.3% keragaman IPM dapat dijelaskan PEFA

Fitted Line Plot

$IPM_{2018} = 60.22 + 3.363 S_PEFA$



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

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

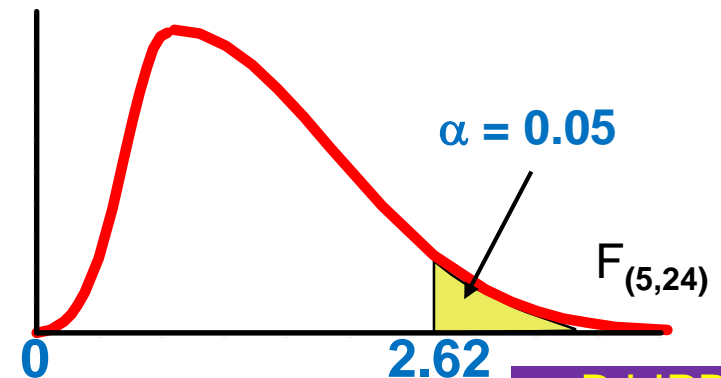
$$\beta_1 = \beta_2 = \dots = \beta_p = 0 \text{ (tidak dapat menjelaskan)}$$

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

Statistik Uji:

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

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

$$\alpha = .05$$

$$db = 5 \text{ dan } 24$$

Nilai Kritis:

$$F = 51.49 ; p=0.00$$

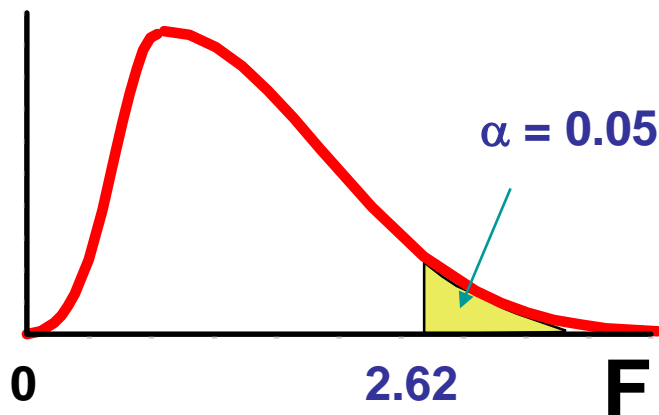
(Output software)

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 rata-rata \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

	<i>db</i>	<i>JK</i>
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(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

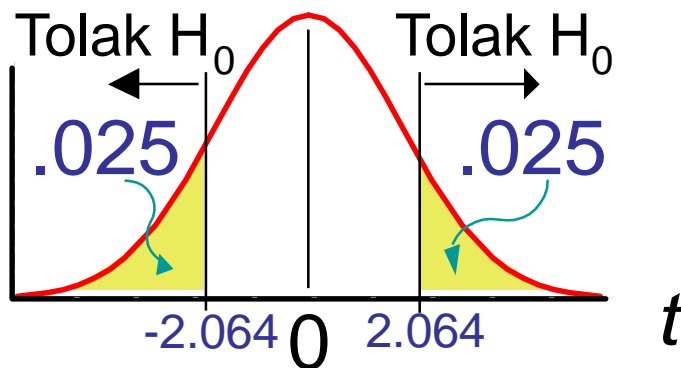
Apakah harga berpengaruh nyata terhadap permintaan Detergen? Ujilah pada $\alpha = 0.05$.

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0$$

$$db = 24$$

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 !

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** .

Interpretasi Koefisien

$$\hat{\beta}_1 = -5.0326$$

Interpretasi Nilai slope -5.03 ('umumnya'): utk kenaikan 1 unit dlm X, diduga Y akan menurun 5.03 unit, *ceteris paribus*

Implikasi dari dugaan slope (data time series):

- *Jika harga **naik 1 dolar**, model tsb memprediksi bahwa permintaan yg diharapkan akan **turun 5.03 unit**, jika faktor lainnya tetap*

Hati2 Interpretasi untuk Data **Cross Section**:

- ***Perbedaan rata2 permintaan** antara entitas (misal toko) yg harganya berbeda 1 dolar adalah **5.03 unit***

Taraf Nyata, α dan Daerah Penolakan

Hipotesis Penelitian



Hipotesis statistika (yg mana?)

$$H_0: \beta_1 \geq k$$

$$H_1: \beta_1 < k$$

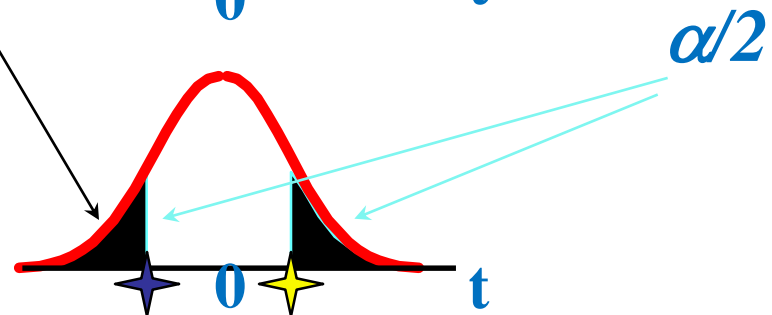
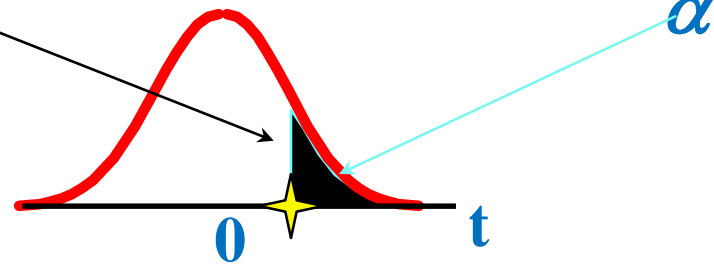
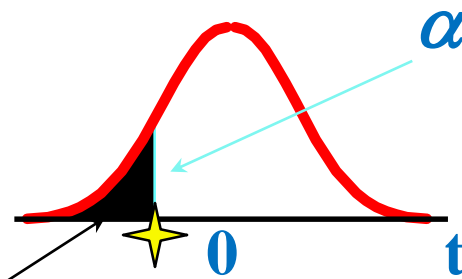
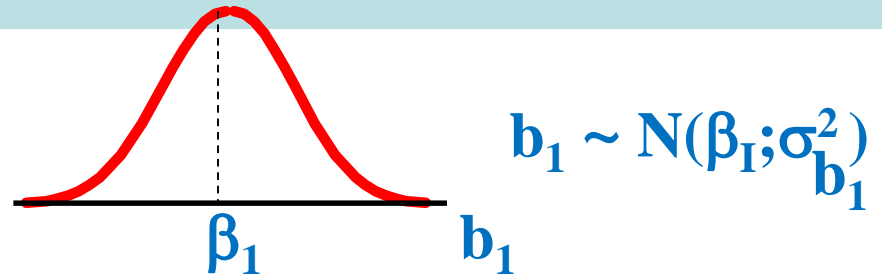
$$H_0: \beta_1 \leq k$$

$$H_1: \beta_1 > k$$

$$H_0: \beta_1 = k$$

$$H_1: \beta_1 \neq k$$

Daerah
Penolakan
(ttk kritis)



Untuk uji eka-arah, misalnya $H_1: \beta > 0$,
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**



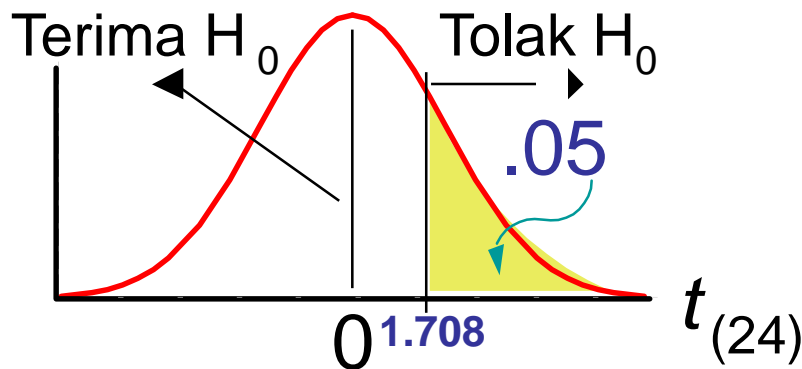
Hipotesis statistika (yg mana?)

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

$$H_1: \beta_T > 0$$

$$db = 24$$

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 deterjen tiap minggu.

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 $(e_i/s_e, X_i)$
 - *Studentized residuals*: $= e_i/s_e$
 - Memungkinkan mempertimbangkan besaran sisaan (sisaan-baku spt Normal baku)

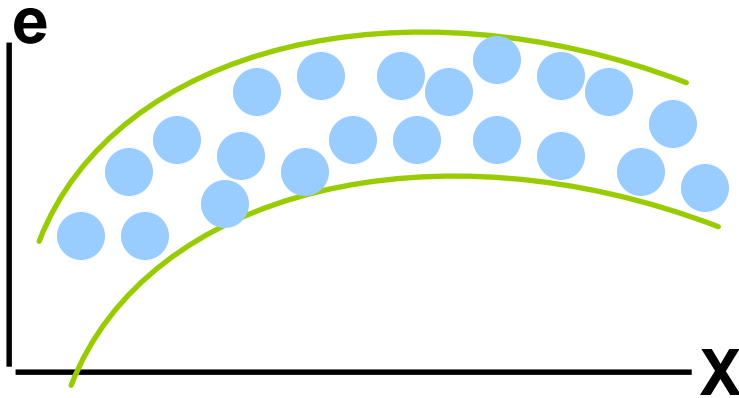
Perbedaan Pendekatan Regresi Klasik dengan Ekonometrika Deret Waktu

1. Pendekatan Regresi 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) Regresi 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

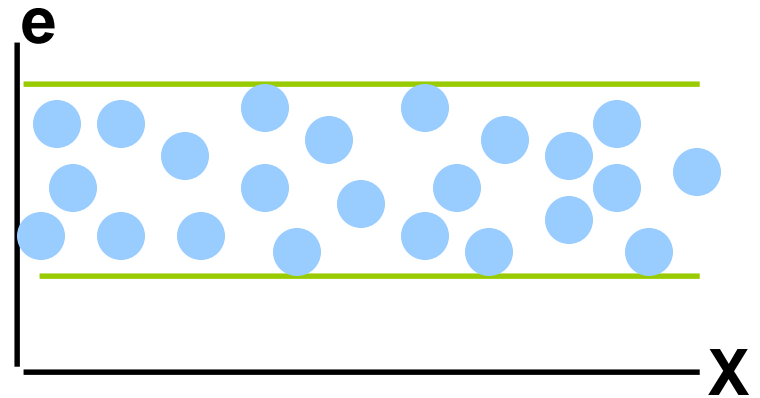
Analisis Sisaan utk Linearitas



Not Linear



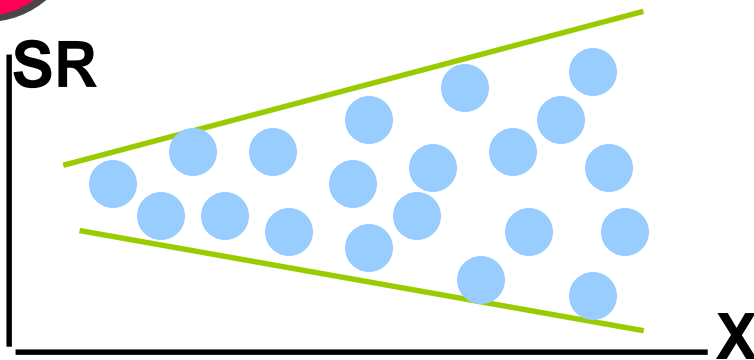
Linear



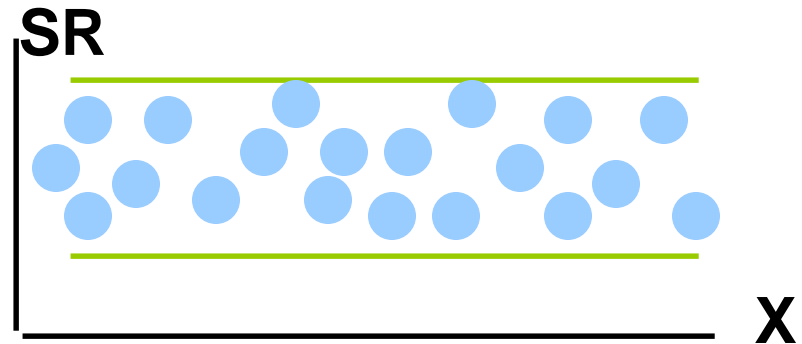
Analisis Sisaan utk Homoskedastisitas



Heteroskedastisitas



Homoscedasticity

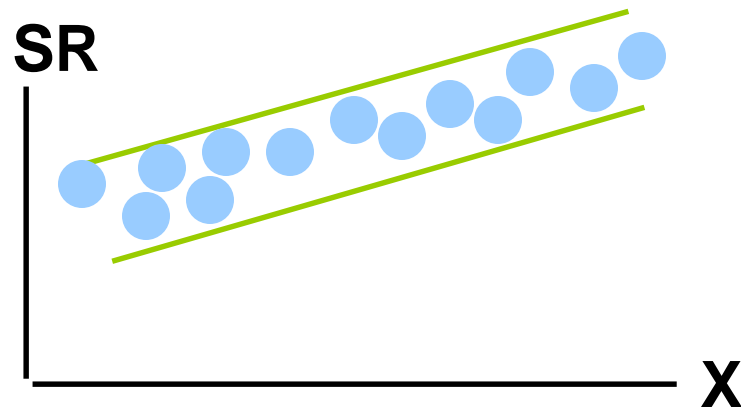


Menggunakan ***Standardized Residuals (SR)***

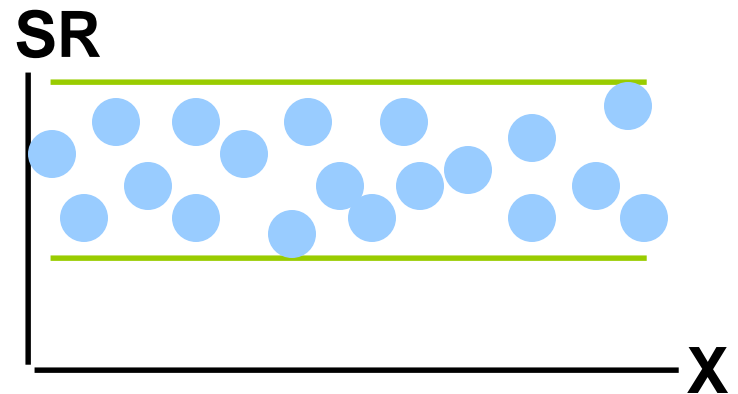
Analisis Sisaan utk Kebebasan e



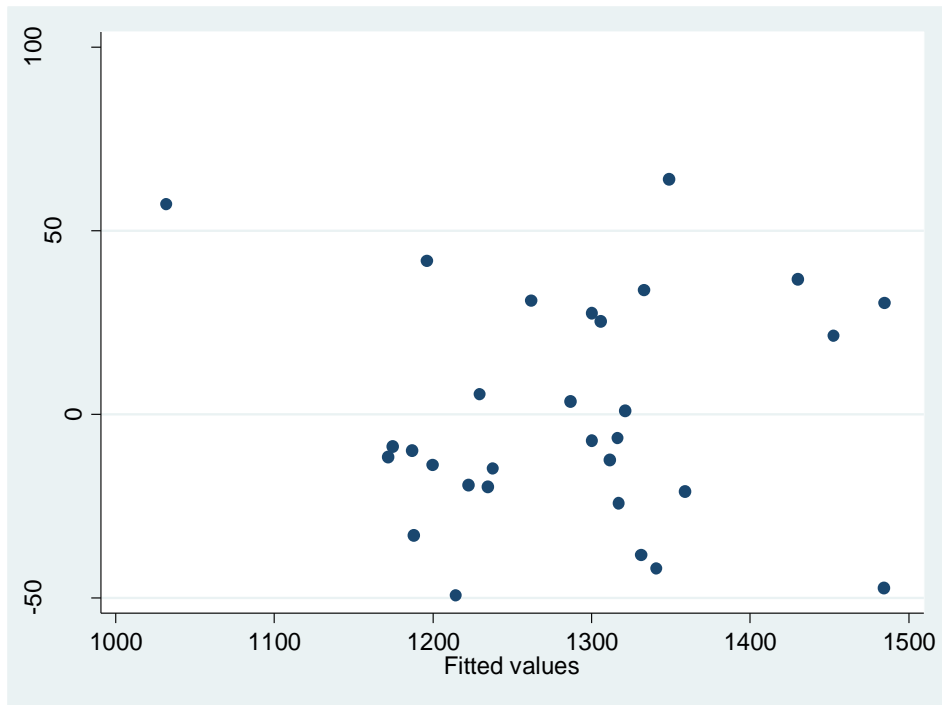
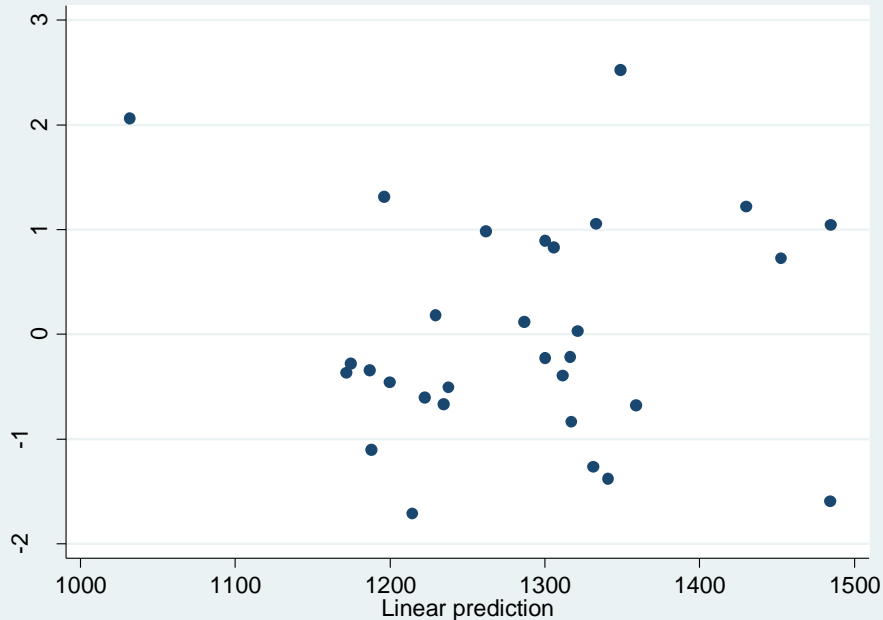
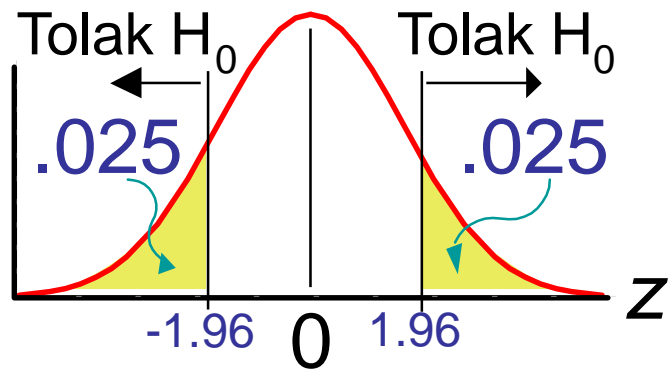
Tidak Bebas



Bebas



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



Statistik Durbin-Watson

- Digunakan utk data *time series* guna mendeteksi 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.