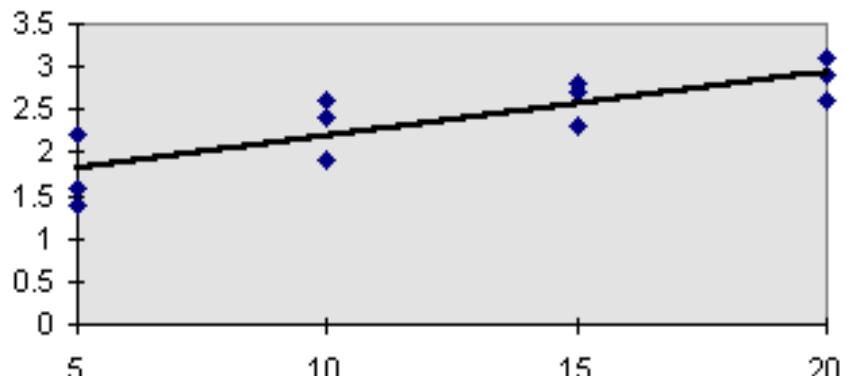


# **APLIKASI MODEL REGRESI**

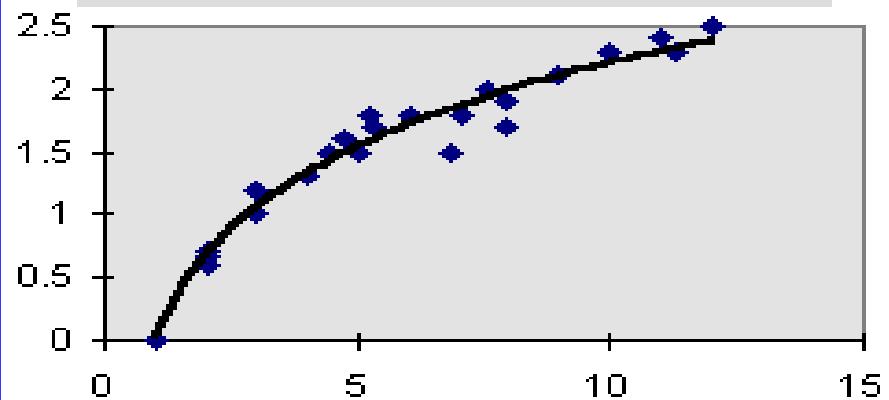
Oleh  
**Bambang Juanda**  
<https://bambangjuanda.com/>

# Tipe Model Regresi

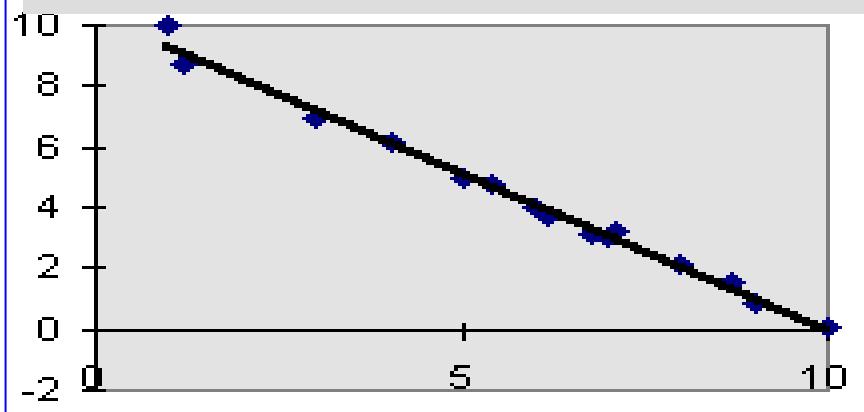
Hubungan Linear Positif



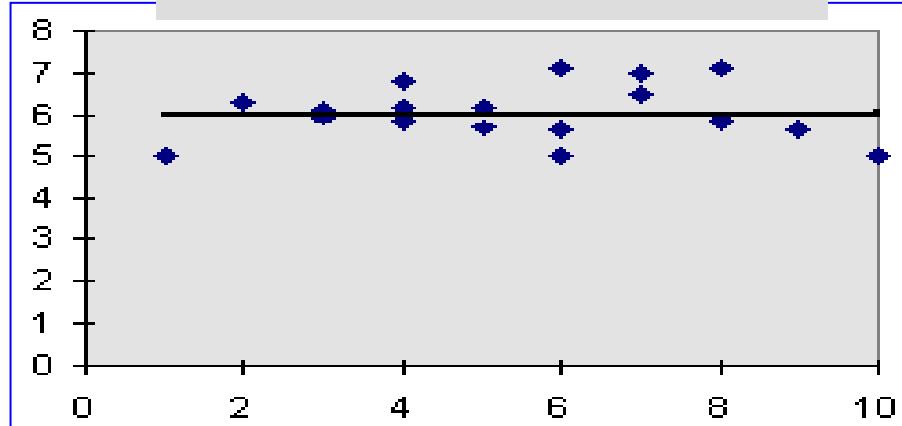
Hubungan Tidak Linear



Hubungan Linear Negatif



Tidak Ada Hubungan



# Faktor Mana yang Paling Berperan dalam Mempengaruhi Permintaan?

. 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

# Koefisien Baku dan Elastisitas

Koefisien regresi  $\beta_j$  dlm model regresi ganda menggambarkan berapa perubahan Y jika peubah bebas ke-j berubah 1 unit.

Koefisien  $\beta_j$  yg paling besar bukan berarti pengaruh peubah bebas ke-j paling besar karena satuan koefisien regresi tergantung satuan peubah respons Y dan satuan peubah bebas ke-j.

Untuk mengkaji relatif pentingnya masing-masing peubah bebas, dapat menggunakan koefisien baku (*standardized coefficients*), elastisitas atau korelasi parsial.

**Koefisien baku** menggambarkan relatif pentingnya peubah bebas  $X_j$  dlm model regresi berganda.

Utk menghitung koefisien baku, kita hanya melakukan suatu regresi linear, yg mana didalamnya, masing-masing peubah di-*normal*-kan dgn cara dikurangi dengan rata-ratanya dan dibagi dengan simpangan bakuinya.

Model regresi yg di-*normal*-kan adalah sbb:

$$\frac{Y_i - \bar{Y}}{S_Y} = \beta_2^* \frac{X_{2i} - \bar{X}_2}{S_{X_2}} + \beta_3^* \frac{X_{3i} - \bar{X}_3}{S_{X_3}} + \dots + \beta_k^* \frac{X_{ki} - \bar{X}_k}{S_{X_k}} + \varepsilon_i^*$$

$$\hat{\beta}_j = \hat{\beta}_j^* \frac{S_Y}{S_{X_j}} \text{ atau } \hat{\beta}_j^* = \hat{\beta}_j \frac{S_{X_j}}{S_Y}; j = 2, 3, \dots, k$$

$\hat{\beta}_j^* = 0.8$  perubahan 1 simpangan (skor) baku dlm peubah bebas  $X$  akan menyebabkan perubahan 0.8 simpangan baku dalam peubah respons  $Y$ .

**Elastisitas** ( $E_j$ ) mengukur pengaruh 1 persen perubahan dlm peubah bebas  $X$  terhadap persentase perubahan peubah respons  $Y$ .

Secara umum, nilai elastisitas tidak konstan tapi berubah jika diukur pada titik yang berbeda sepanjang jaris regresi.

Elastisitas kadangkala dikeluarkan oleh paket program komputer yg dihitung pd titik rata-rata masing-masing peubah.

Untuk koefisien ke- $j$ , elastisitas dihitung sbb :

$$E_j = \eta_j = \frac{\Delta Y}{Y} / \frac{\Delta X}{X} = \frac{\Delta Y}{\Delta X} \frac{X}{Y} \cong \beta_j \frac{\bar{X}_j}{\bar{Y}}$$

# Koef Regresi, Koefisien Baku, dan Elastisitas

Peubah	Koefisien $\beta_j$	Koef baku $\beta_j^*$	Elastisitas $E_j = \eta_j$	Mean	Std. Deviation
Y				1285.8000	104.80241
X2_P	-5.033	-.688	-0.49712	127.0000	14.31903
X3_Pr	4.744	.327	0.32222	87.3333	7.22702
X4_Advertise	.277	.184	0.18746	870.1667	69.50717
X5_Income	.011	.498	0.32021	37430.0333	4898.18285
X6_T	1.309	.110	0.01578	15.5000	8.80341

Peubah bebas yg kontribusinya paling besar dalam mempengaruhi permintaan produk, secara berurutan, adalah harga produk tsb ( $X_2$ ), daya beli masyarakat ( $X_5$ ), dan harga produk pesaing ( $X_3$ )

# **R<sup>2</sup> dan R<sup>2</sup> Terkoreksi**

R<sup>2</sup> sensitif terhadap jumlah peubah bebas dlm model. Penambahan peubah bebas baru ke dalam persamaan regresi tidak pernah mengurangi R<sup>2</sup>, bahkan cenderung menaikkan R<sup>2</sup> (karena tidak menambah JKT tapi cenderung menaikkan JKR). Jadi, kita tambahkan lebih banyak peubah bebas saja jika ingin memaksimumkan R<sup>2</sup>.

R<sup>2</sup> terkoreksi mempunyai karakteristik yang diinginkan sebagai ukuran goodness of fit daripada R<sup>2</sup>. Jika peubah baru ditambahkan, R<sup>2</sup> selalu naik, tapi R<sup>2</sup>-terkoreksi dapat naik atau turun. Penggunaan R<sup>2</sup> menghindari dorongan peneliti untuk memasukkan sebanyak mungkin peubah bebas tanpa pertimbangan yang logis.

# R<sup>2</sup> Terkoreksi

$$R^2_{terkoreksi} = \bar{R}^2 = 1 - \frac{\hat{Var}(\varepsilon)}{\hat{Var}(Y)} = 1 - \frac{\sum ei^2}{\sum(Y_i - \bar{Y})^2} \left( \frac{n-1}{n-k} \right)$$

$$\bar{R}^2 = 1 - (1 - R^2) \left( \frac{n-1}{n-k} \right)$$

Dari persamaan di atas terlihat bahwa :

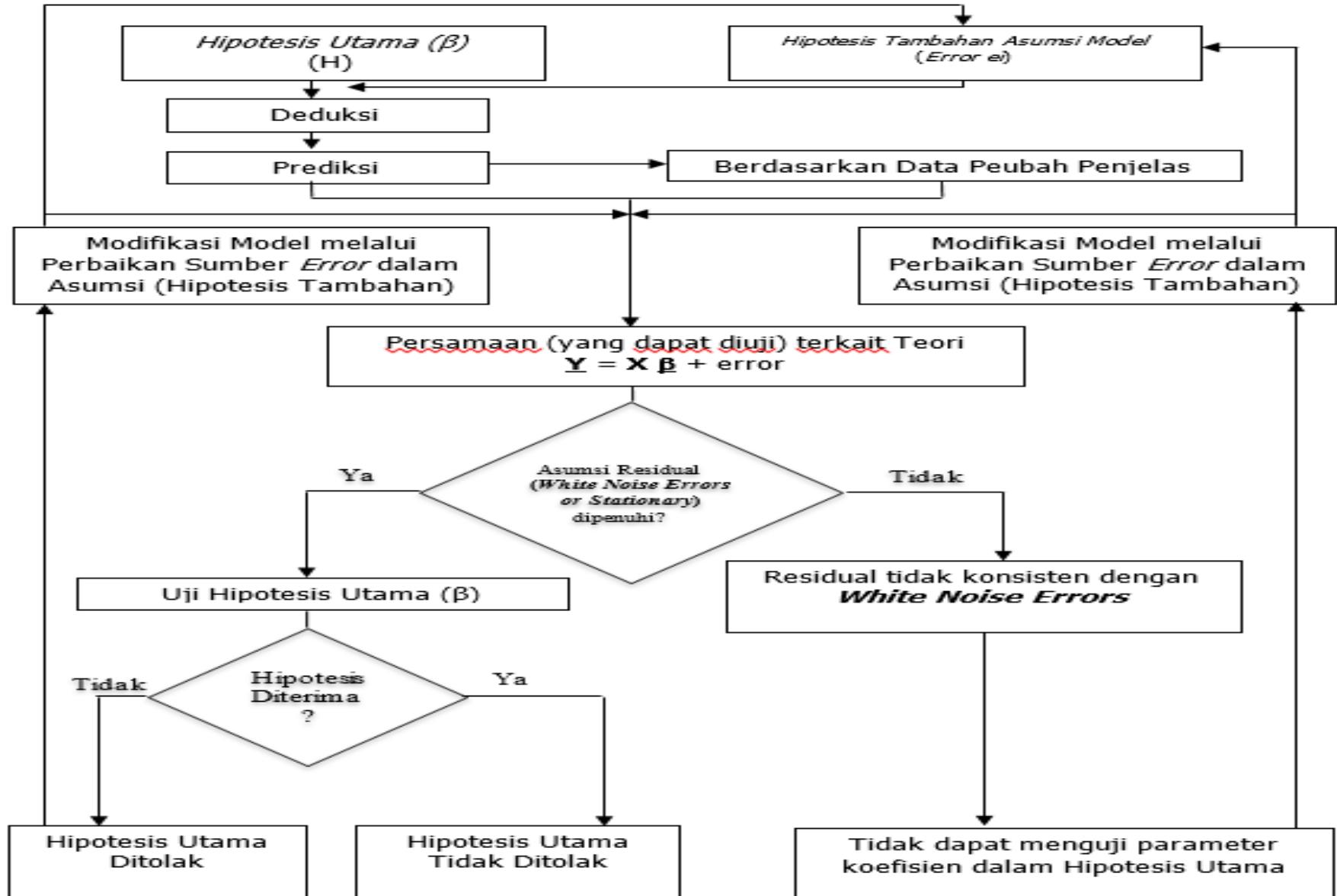
- Jika  $k=1$  maka  $R^2$  sama dengan  $R^2$  terkoreksi.
- Jika  $k>1$  maka  $R^2 \geq R^2$  terkoreksi.

# Pengaruh Marginal (Parsial) dan Elastisitas dari Berbagai Bentuk Fungsi Model

Nama	Bentuk Fungsi Model	<i>Marginal Effect = dy/dX</i>	Elastisitas [(X/Y)(dY/dX)]
Linear	$Y = \beta_1 + \beta_2 X$	$\beta_2$	$\beta_2 X/Y$
Linear-Log	$Y = \beta_1 + \beta_2 \ln X$	$\beta_2/X$	$\beta_2/Y$
Reciprocal	$Y = \beta_1 + \beta_2 (1/X)$	$-\beta_2/X^2$	$-\beta_2/(XY)$
Quadratic	$Y = \beta_1 + \beta_2 X + \beta_3 X^2$	$\beta_2 + 2\beta_3 X$	$(\beta_2 + 2\beta_3 X)X/Y$
Interaction	$Y = \beta_1 + \beta_2 X + \beta_3 XZ$	$\beta_2 + \beta_3 Z$	$(\beta_2 + \beta_3 Z)X/Y$
Log-Linear	$\ln Y = \beta_1 + \beta_2 X$	$\beta_2 Y$	$\beta_2 X$
Log-reciprocal	$\ln Y = \beta_1 + \beta_2 (1/X)$	$-\beta_2 Y/X^2$	$-\beta_2 / X$
Log-quadratic	$\ln Y = \beta_1 + \beta_2 X - \beta_3 X^2$	$Y(\beta_2 + 2\beta_3 X)$	$X(\beta_2 + 2\beta_3 X)$
Double-log	$\ln Y = \beta_1 + \beta_2 \ln X$	$\beta_2 Y - X$	$\beta_2$
Logistic	$\ln \left[ \frac{Y}{1-Y} \right] = \beta_1 - \beta_2 X$	$\beta_2 Y(1-Y)$	$\beta_2(1-Y)X$

Variabel dgn satuan %, tidak disarankan dlm bentuk Log

# Tahapan Pemodelan Empiris



untuk menguji hipotesis, perlu diperiksa dulu apakah modelnya sudah “terspesifikasi dengan benar dengan melihat asumsi error”

Perlu diingat bahwa, jika hasil statistik-uji menunjukkan bahwa hipotesis utama ditolak, ini belum cukup bukti untuk menyimpulkan bahwa hipotesis tersebut benar-benar ditolak, karena kerangka pengujian hipotesis tersebut tergantung dari cara bagaimana peneliti memformulasikan hipotesis tersebut ke dalam koefisien parameter (model). Jadi kurang layak kalau menyimpulkan penolakan hipotesis pada pengujian pertama terhadap hipotesis tersebut

## Pemilihan Model “Terbaik”:

1. Asumsi Model regresi dipenuhi
2. R<sup>2</sup> tinggi dan signifikan;
3. Banyak koefisien signifikan;
4. Interpretasi (arah) Koefisien logis

Berikut ini data nilai penjualan dari perusahaan Microsoft dari tahun 1984 sampai 1998:

Year	Sales Revenue (\$ Million)	Year	Sales Revenue (\$ Million)	Year	Sales Revenue (\$ Million)
1998	14000	1993	3753	1988	590.8
1997	11358	1992	2758.7	1987	345.9
1996	8671	1991	1843.4	1986	202.1
1995	5937	1990	1183.4	1985	139.5
1994	4649	1989	803.5	1984	99.5

- Buatlah plot deret waktunya, dan interpretasikan polanya.
- Buatlah dugaan model permintaan  $S_t = a + b t$
- Buatlah dugaan model permintaan  $S_t = S_0 (1+ g)^t$
- Buatlah dugaan model permintaan  $S_t = S_0 e^{gt}$
- Interpretasikan koefisien dari 3 model tsb, dan ujilah apakah signifikan.
- Model permintaan mana yang lebih cocok (misalnya utk peramalan) menurut saudara?**

# Contoh Output STATA (Bgm Interpretasinya?)

. regr Ln\_D Ln\_P Ln\_Pr Ln\_Ad Ln\_Inc Ln\_T

Source	SS	df	MS	Number of obs	=	30
Model	.174106945	5	.034821389	F(5, 24)	=	54.33
Residual	.015383242	24	.000640968	Prob > F	=	0.0000
Total	.189490187	29	.006534144	R-squared	=	0.9188
				Adj R-squared	=	0.9019
				Root MSE	=	.02532

Ln_D	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
Ln_P	-.4683947	.0423134	-11.07	0.000	-.5557253 - .3810641
Ln_Pr	.3210997	.0638573	5.03	0.000	.1893047 .4528947
Ln_Ad	.1805561	.0687546	2.63	0.015	.0386535 .3224586
Ln_Inc	.3156583	.0363698	8.68	0.000	.2405947 .3907218
Ln_T	.0137286	.0058208	2.36	0.027	.0017152 .0257421
_cons	3.410851	.6761202	5.04	0.000	2.015407 4.806294

Model mana yang paling baik dibandingkan sebelumnya? Jelaskan

Seorang peneliti mempunyai hipotesis bahwa produksi sektor industri atau non-pertanian (Q) dipengaruhi oleh banyaknya modal (K), tenaga kerja (L), dan bahan baku impor (M) yang digunakan. Berdasarkan data berkala selama 40 tahun terakhir, diperoleh dugaan model regresinya di bawah ini. Angka di dalam tanda kurung adalah nilai-p dari statistik-uji t yang digunakan untuk menguji apakah masing-masing faktor berpengaruh nyata. Selain itu dapat dilihat juga nilai R<sup>2</sup> dan statistik Durbin-Watson dari model dugaan tersebut.

$$\log Q_t = 12.103 + 0.5284 \log K_t + 0.9382 \log L_t + 0.0691 \log M_t + \varepsilon_t$$
$$(0.091) \quad (0.064) \quad (0.034) \quad (0.597)$$
$$R^2 = 0.9583; \quad \text{statistik-DW}=2.024$$

### Pertanyaan:

- Dari hasil analisis regresi diatas, apakah hipotesis peneliti tsb didukung oleh dugaan model empiris tersebut? Jelaskan.
- Jelaskan, jika Anda menghadapi masalah ini, Apakah Anda puas dengan hasil dugaan model ini? Jika tidak puas, apa yang akan Anda lakukan untuk memperbaiki model yang menggambarkan faktor-faktor apa saja yang berpengaruh terhadap produksi sektor industri tersebut.

Bagaimana memasukkan *technological improvement (omitted relevant variable)*?

# Model dgn Peubah-Dummy *(tanpa/dengan interaksi)*

- Peubah Kategori memerlukan (peubah dummy) dgn 2 Taraf:
  - Ya atau Tidak, Pria atau Wanita,
  - Dikodekan 1 atau 0
- Perbedaan Intersep
- Asumsi Slope Sama
- Model Regresi punya Bentuk Sama:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_p X_{pi} + \varepsilon_i$$

# Model untuk menduga Harga Rumah

Data Editor (Edit) - [Bab 5 Nilai Rumah]

File Edit View Data Tools

Nilai\_Rumah[1] 3869

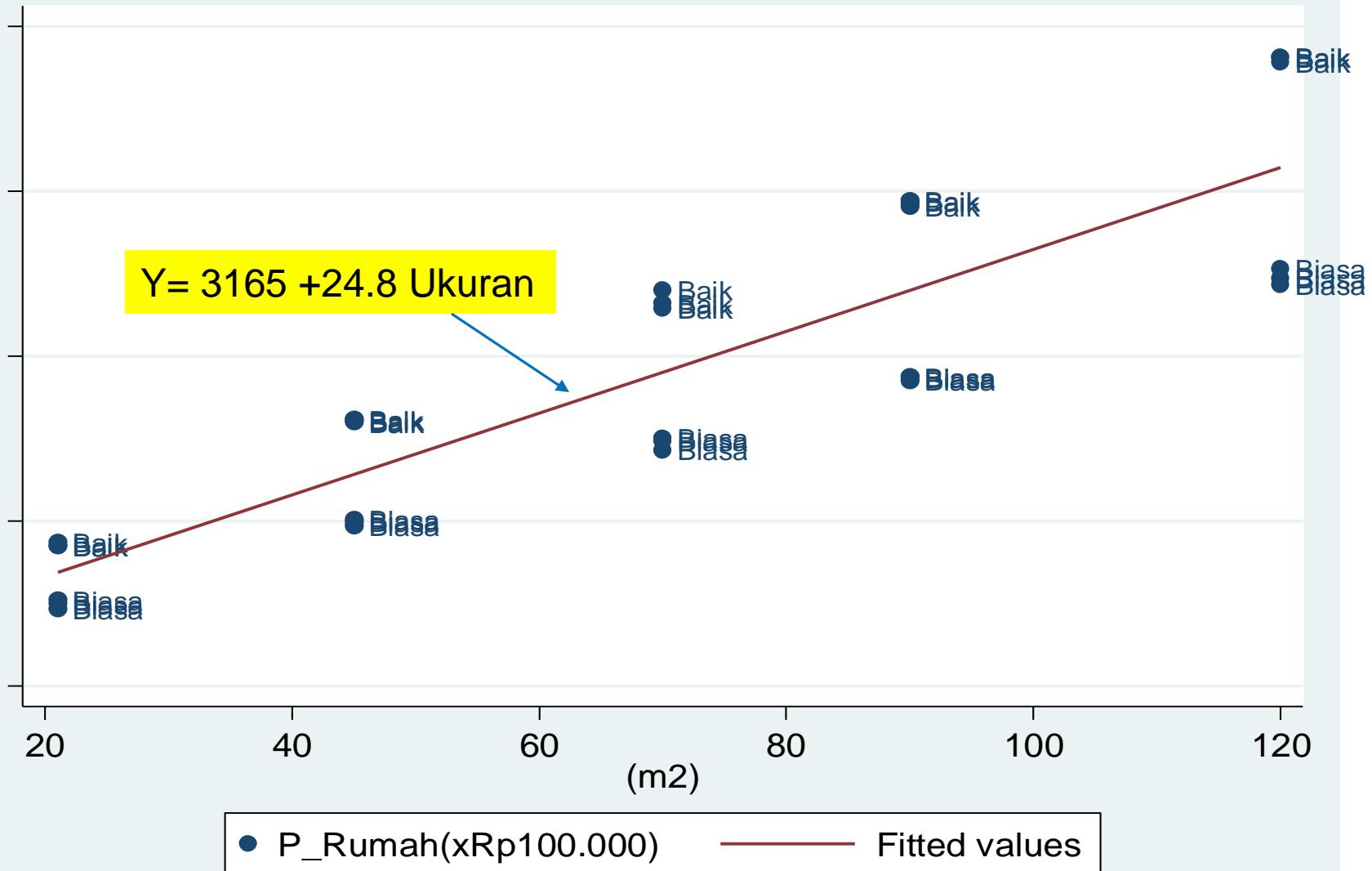
	Nilai_Rumah	Ukuran	Lok	UkxLok	Lokasi
1	3869	21	1	21	Baik
2	3845	21	1	21	Baik
3	3846	21	1	21	Baik
4	4605	45	1	45	Baik
5	4618	45	1	45	Baik
6	4598	45	1	45	Baik
7	5287	70	1	70	Baik
8	5403	70	1	70	Baik
9	5319	70	1	70	Baik
10	5943	90	1	90	Baik
11	5936	90	1	90	Baik
12	5911	90	1	90	Baik
13	6809	120	1	120	Baik
14	6782	120	1	120	Baik
15	6816	120	1	120	Baik
16	3496	21	0	0	Biasa
17	3464	21	0	0	Biasa
18	3519	21	0	0	Biasa
19	4010	45	0	0	Biasa
20	3995	45	0	0	Biasa
21	3971	45	0	0	Biasa
22	4504	70	0	0	Biasa
23	4429	70	0	0	Biasa
24	4489	70	0	0	Biasa

Variables

Filter variables here

Name	Label
Nilai_Rumah	P_Rumah(xRp100.000)
Ukuran	(m <sup>2</sup> )
Lok	kode lokasi (1:Baik, 0:Biasa)
UkxLok	Interaksi Ukuran & Lokasi
Lokasi	Lokasi

# Scatter Plot & Dugaan Model



# Contoh Output STATA (Bgm Interpretasinya?)

. regress Nilai\_Rumah Ukuran

Source	SS	df	MS	Number of obs	=	30
Model	21914731.3	1	21914731.3	F(1, 28)	=	98.63
Residual	6221561.7	28	222198.632	Prob > F	=	0.0000
Total	28136293	29	970216.999	R-squared	=	0.7789
				Adj R-squared	=	0.7710
				Root MSE	=	471.38

Nilai_Rumah	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
Ukuran	24.83298	2.500527	9.93	0.000	19.71089 29.95508
_cons	3164.924	193.2569	16.38	0.000	2769.055 3560.793

. regress Nilai\_Rumah Ukuran Lok

Source	SS	df	MS	Number of obs	=	30
Model	27268228.9	2	13634114.5	F(2, 27)	=	424.07
Residual	868064.062	27	32150.5208	Prob > F	=	0.0000
Total	28136293	29	970216.999	R-squared	=	0.9691
				Adj R-squared	=	0.9669
				Root MSE	=	179.31

Nilai_Rumah	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
Ukuran	24.83298	.9511628	26.11	0.000	22.88136 26.78461
Lok	844.8667	65.47317	12.90	0.000	710.5268 979.2065
_cons	2742.491	80.47174	34.08	0.000	2577.376 2907.605

# Asumsi Model dgn Peubah-Dummy

Model:

$$\hat{Y}_i = b_1 + b_2 X_{2i} + b_3 X_{3i}$$

Y = Perkiraan Nilai Rumah

X<sub>2</sub> = Ukuran Rumah (Square footage)

X<sub>3</sub> = Lingkungan Rumah = D

{ 0 jika “biasa”  
1 jika “baik”

“Baik” (X<sub>3</sub> = 1)

$$\hat{Y}_i = b_1 + b_2 X_{2i} + b_3(1) = (b_1 + b_3) + b_2 X_{2i}$$

Slope  
Sama?

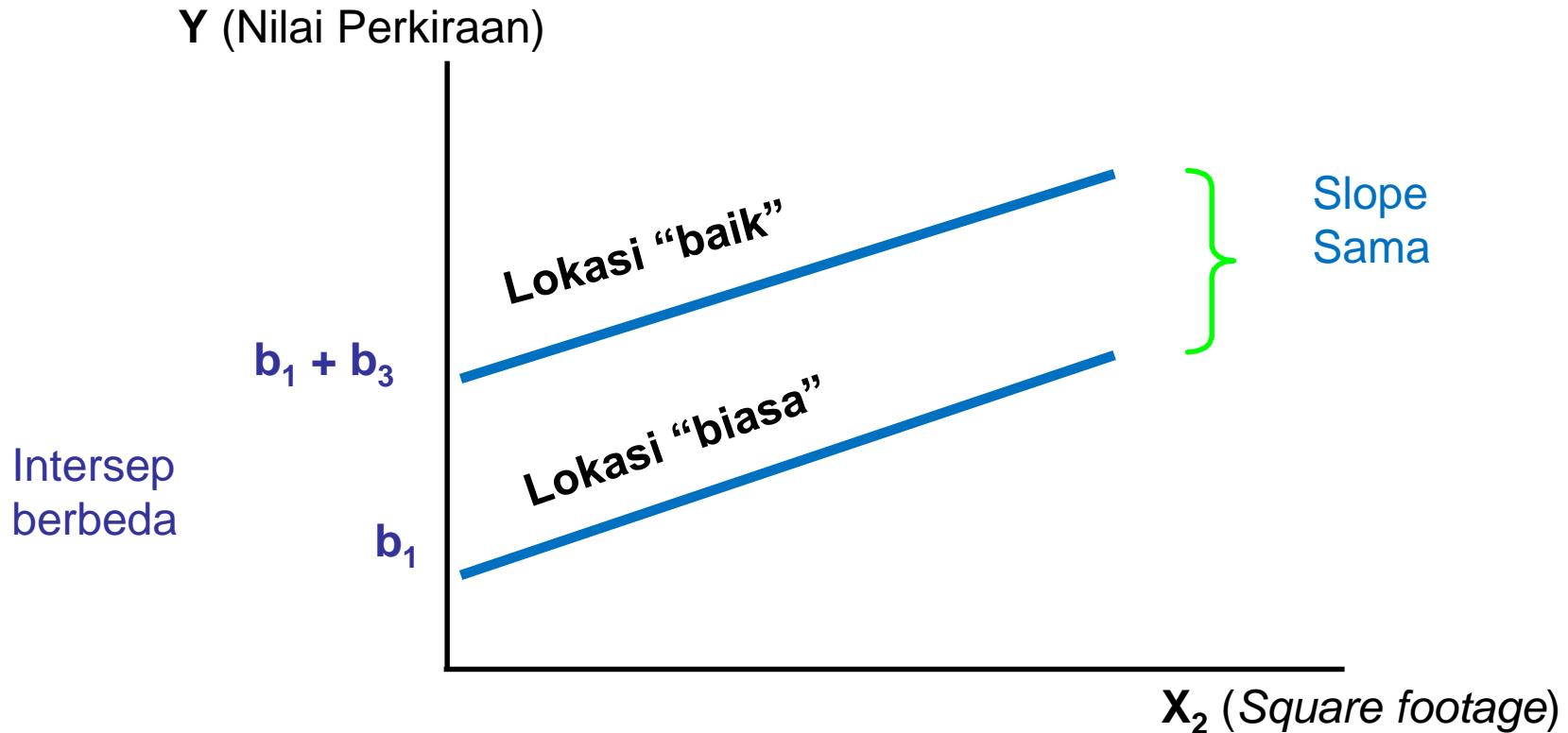
“Biasa” (X<sub>3</sub> = 0)

$$\hat{Y}_i = b_1 + b_2 X_{2i} + b_3(0) = b_1 + b_2 X_{2i}$$

b<sub>3</sub>: rata-rata perbedaan nilai rumah antara kelompok rumah di lingkungan “baik” dgn di lingkungan “biasa”, yg ukurannya sama.

- Pengaruh perubahan ukuran rumah sama saja utk kedua lingkungan tsb (slope sama)
- Tdk ada pengaruh interaksi antara ukuran dgn lingkungan thd nilai rumah

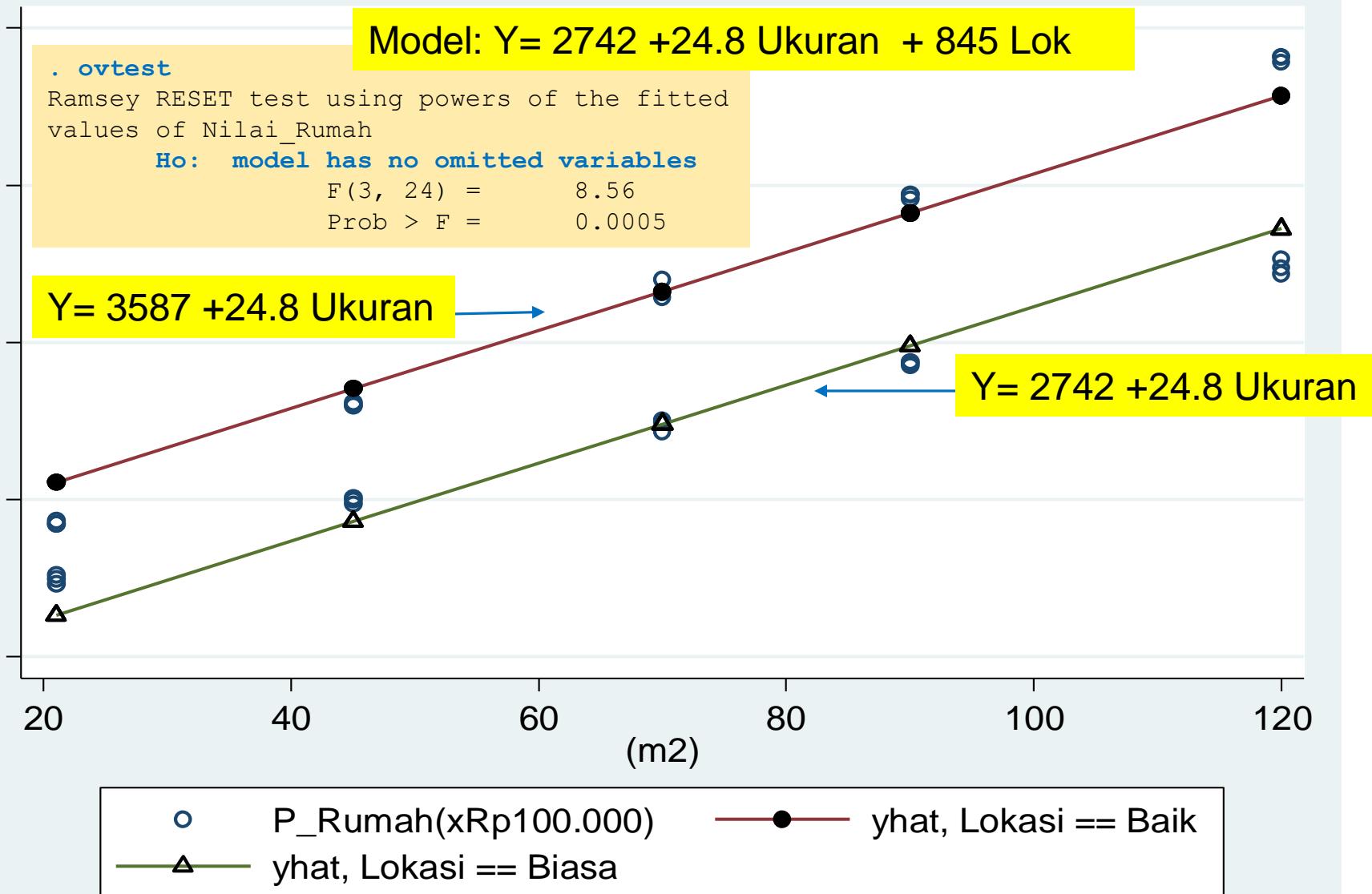
# Asumsi Model dgn Peubah-Dummy



$b_3$ : rata-rata perbedaan nilai rumah antara kelompok rumah di lingkungan “baik” dgn di lingkungan “biasa”, yg ukurannya sama.

- Pengaruh perubahan ukuran rumah sama saja utk kedua lingkungan tsb (slope sama)
- Tdk ada pengaruh interaksi antara ukuran dgn lingkungan thd nilai rumah

# Dugaan Model dg Dummy



# Evaluasi Adanya Interaksi (Perbedaan Slope)

- Interaksi: Pengaruh  $X_2$  thd Y tergantung Lokasi (D)
- Berisi Perkalian 2 Peubah

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 D_i + \beta_4 (X_{2i} D_i) + \varepsilon_i$$

- Hipotesis:

$H_0: \beta_4 = 0$  (Tdk ada interaksi antara  $X_2$  dgn  $X_3$ )

$H_1: \beta_4 \neq 0$  ( $X_2$  berinteraksi dgn  $X_3$ )

Lingkungan “baik” (D=1)

$$\hat{Y}_i = b_1 + b_2 X_{2i} + b_3(1) + b_4 X_{2i}(1) = (b_1 + b_3) + (b_2 + b_4) X_{2i}$$

Lingkungan “biasa” ( $X_3=0$ )

$$\hat{Y}_i = b_1 + b_2 X_{2i} + b_3(0) + b_4 X_{2i}(0) = b_1 + b_2 X_{2i}$$

- Pengaruh perubahan ukuran rumah tidak sama utk kedua lingkungan tsb
- Ada pengaruh interaksi antara ukuran dgn lingkungan thd nilai rumah

Slope  
tdkSama

Lingkungan “baik” ( $X_3=1$ )

$$\hat{Y}_i = b_1 + b_2 X_{2i} + b_3(1) + b_4 X_{2i}(1) = (b_1 + b_3) + (b_2 + b_4)X_{2i}$$

Slope  
tdkSama

Lingkungan “biasa” ( $X_3=0$ )

$$\hat{Y}_i = b_1 + b_2 X_{2i} + b_3(0) + b_4 X_{2i}(0) = b_1 + b_2 X_{2i}$$

$b_3$  : rata-rata perbedaan nilai rumah lingkungan “baik” dgn “biasa”, jika  $X_2= 0$  ??? (tdk perlu diinterpretasi)

$b_2$  : rata-rata perbedaan nilai rumah di lingkungan “biasa” yg ukurannya berbeda 1 m<sup>2</sup>. (*pengaruh ukuran thd nilai rumah utk lingkungan “biasa”*)

( $b_2 + b_4$ ): rata-rata perbedaan nilai rumah di lingkungan “baik” yg ukurannya berbeda 1 m<sup>2</sup>. (*pengaruh ukuran thd nilai rumah utk lingkungan “baik”*)

$b_4$  : perbedaan pengaruh ukuran terhadap nilai rumah antara kedua lingkungan.

*Pengaruh ukuran thd nilai rumah tergantung lingkungannya*

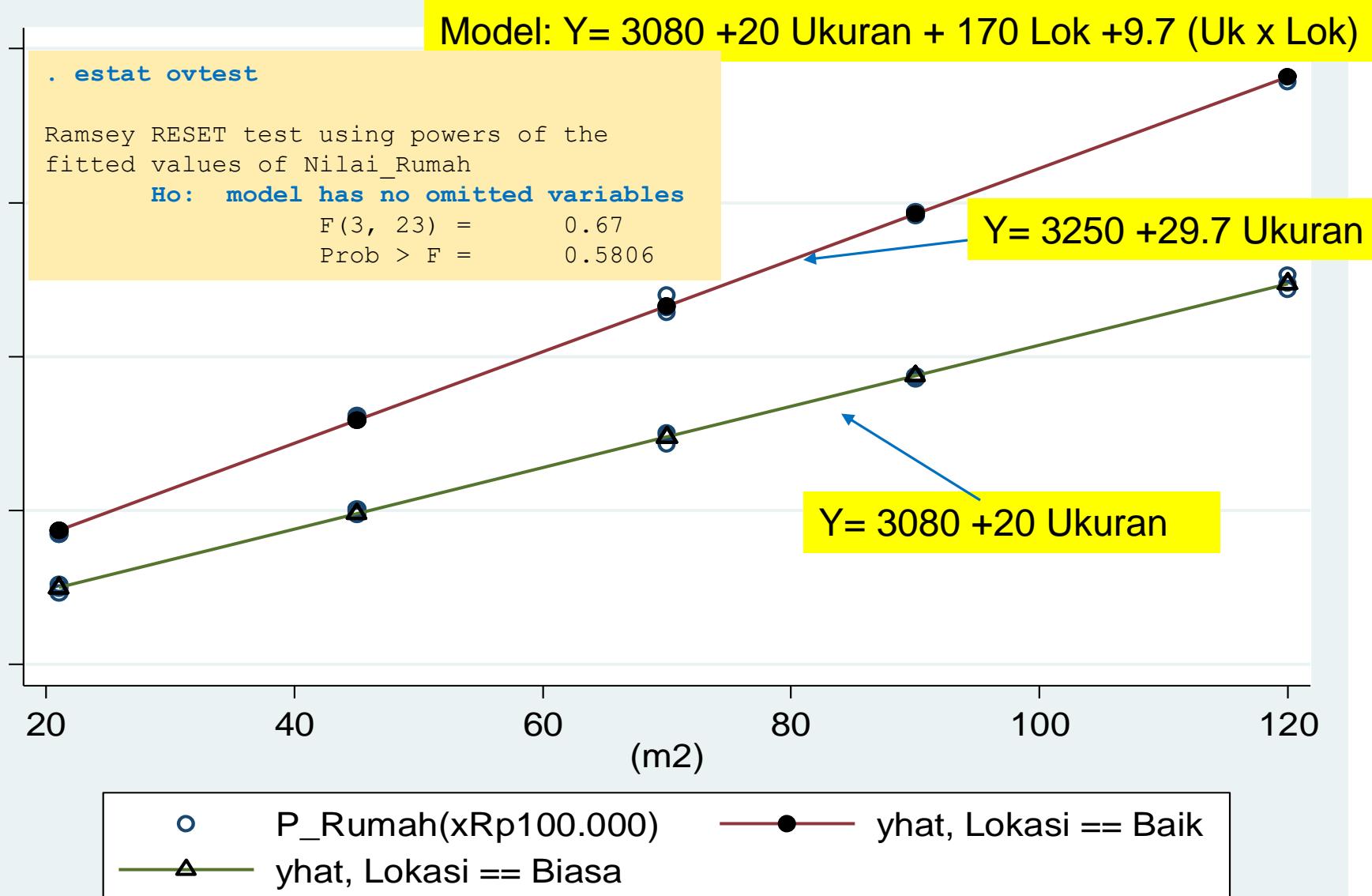
# Contoh Output STATA (Bgm Interpretasinya?)

. regress Nilai\_RatRib Ukuran\_m2 Lokasi UkwLok

Source	SS	df	MS	Number of obs	=	30
Model	28112612.6	3	9370870.85	F(3, 26)	=	10288.78
Residual	23680.4162	26	910.785239	Prob > F	=	0.0000
Total	28136293	29	970216.999	R-squared	=	0.9992
				Adj R-squared	=	0.9991
				Root MSE	=	30.179

Nilai_RatRib	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Ukuran_m2	19.95848	.2264039	88.15	0.000	19.4931	20.42386
Lokasi	170.2348	24.74585	6.88	0.000	119.369	221.1006
UkwLok	9.749015	.3201834	30.45	0.000	9.090869	10.40716
_cons	3079.807	17.49796	176.01	0.000	3043.839	3115.774

# Dugaan Model dg Dummy Interaksi



# Model untuk menduga Gaji Karyawan

Data Editor (Edit) - [Bab 5 Gaji]

File Edit View Data Tools



var10[1]

	Pddk	Pddk_id	D_SL	D_PT	ThxD_SL	ThxD_PT	D_SD	Gaji	Penglmn	Variables	
1	SD	SD	0	0	0	0	1	3390	2	<input type="checkbox"/> Filter variables here	
2	SD	SD	0	0	0	0	1	3182	2	<input checked="" type="checkbox"/> Name	Label
3	SD	SD	0	0	0	0	1	4916	5	<input checked="" type="checkbox"/> Pddk	
4	SD	SD	0	0	0	0	1	4164	3	<input checked="" type="checkbox"/> Pddk_id	
5	SD	SD	0	0	0	0	1	5047	6	<input checked="" type="checkbox"/> D_SL	
6	SD	SD	0	0	0	0	1	4285	4	<input checked="" type="checkbox"/> D_PT	
7	SD	SD	0	0	0	0	1	6565	8	<input checked="" type="checkbox"/> ThxD_SL	
8	SD	SD	0	0	0	0	1	7524	10	<input checked="" type="checkbox"/> ThxD_PT	
9	SD	SD	0	0	0	0	1	4865	5	<input checked="" type="checkbox"/> D_SD	
10	SD	SD	0	0	0	0	1	3620	3	<input checked="" type="checkbox"/> Gaji	(Ribuan Rp)
11	SL	SL	1	0	7	0	0	9620	7	<input checked="" type="checkbox"/> Penglmn	Tahun
12	SL	SL	1	0	8	0	0	10942	8		
13	SL	SL	1	0	5	0	0	7777	5		
14	SL	SL	1	0	6	0	0	8735	6		
15	SL	SL	1	0	7	0	0	9428	7		
16	SL	SL	1	0	2	0	0	4588	2		
17	SL	SL	1	0	4	0	0	6537	4		
18	SL	SL	1	0	9	0	0	11684	9		
19	SL	SL	1	0	8	0	0	10573	8		
20	SL	SL	1	0	5	0	0	7874	5		
21	PT	PT	0	1	0	2	0	7147	2		
22	PT	PT	0	1	0	7	0	14610	7		
23	PT	PT	0	1	0	9	0	17776	9		
24	PT	PT	0	1	0	10	0	18756	10		
25	PT	PT	0	1	0	11	0	20098	11		
26	PT	PT	0	1	0	6	0	13124	6		
27	PT	PT	0	1	0	5	0	11075	5		
28	PT	PT	0	1	0	8	0	15851	8		

# Peubah Kualitatif dgn 3 Kategori

Model:  $\hat{Y}_i = b_1 + b_2 X_{2i} + b_3 X_{3i}$

Jika ada k kategori →  
ada k-1 peubah *Dummy*

$Y_i$  = Gaji karyawan ke-i (ribu rupiah)

$X_{2i}$  = Pengalaman kerja karyawan ke-i (th)

$X_{3i}$  = Pendidikan karyawan ke-i = D<sub>3i</sub> =

$\begin{cases} 0 & \text{jika "SD"} \\ 1 & \text{jika "SL"} \\ 2 & \text{Jika "PT"} \end{cases}$

"PT" ( $X_3 = 2$ )

$$\hat{Y}_i = b_1 + b_2 X_{2i} + b_3(2) = (b_1 + 2b_3) + b_2 X_{2i}$$

"SL" ( $X_3 = 1$ )

$$\hat{Y}_i = b_1 + b_2 X_{2i} + b_3(1) = (b_1 + b_3) + b_2 X_{2i}$$

"SD" ( $X_3 = 0$ )

$$\hat{Y}_i = b_1 + b_2 X_{2i} + b_3(0) = b_1 + b_2 X_{2i}$$

Slope  
Sama

Note: Perbedaan gaji tiap jenjang pendidikan **SAMA ( $b_3$ )**, jika pengalamannya sama

# Jika ada k kategori $\rightarrow$ ada k-1 peubah Dummy

Model:

$$\hat{Y}_i = b_1 + b_2 X_{2i} + b_3 D_{1i} + b_4 D_{2i}$$

$Y_i$  = Gaji karyawan ke-i (ribu rupiah)

$X_{2i}$  = Pengalaman kerja karyawan ke-i (th)

$X_{3i}$  = Pendidikan karyawan ke-i (D)

Pendidikan	D1_sl	D2_pt
SD	0	0
SL	1	0
PT	0	1

“PT” ( $D_1=0; D_2 = 1$ )

$$\hat{Y}_i = b_1 + b_2 X_{2i} + b_3(0) + b_4(1) = (b_1 + b_4) + b_2 X_{2i}$$

“SL” ( $D_1=1; D_2=0$ )

$$\hat{Y}_i = b_1 + b_2 X_{2i} + b_3(1) + b_4(0) = (b_1 + b_3) + b_2 X_{2i}$$

“SD” ( $D_1=0; D_2 = 0$ )

$$\hat{Y}_i = b_1 + b_2 X_{2i} + b_3(0) + b_4(0) = b_1 + b_2 X_{2i}$$

Slope  
Sama

Note: Perbedaan gaji tiap jenjang pendidikan dapat BERBEDA,  
jika pengalamannya sama

Model utk "PT" ( $D_1=0$ ;  $D_2 = 1$ )

$$\hat{Y}_i = b_1 + b_2 X_{2i} + b_3(0) + b_4(1) = (b_1 + b_4) + b_2 X_{1i}$$

"SL" ( $D_1=1$ ;  $D_2=0$ )

$$\hat{Y}_i = b_1 + b_2 X_{2i} + b_3(1) + b_4(0) = (b_1 + b_3) + b_2 X_{2i}$$

Slope  
Sama !

"SD" ( $D_1=0$ ;  $D_2 = 0$ )

$$\hat{Y}_i = b_1 + b_2 X_{2i} + b_3(0) + b_4(0) = b_1 + b_2 X_{2i}$$

$b_3$  : rata-rata perbedaan gaji karyawan lulusan SL dgn lulusan SD, jika pengalamannya ( $X_2$ ) sama

$b_4$  : rata-rata perbedaan gaji karyawan lulusan PT dgn lulusan SD, jika pengalamannya ( $X_2$ ) sama

$b_4 - b_3$  : rata-rata perbedaan gaji karyawan lulusan PT dgn lulusan SL, jika pengalamannya ( $X_2$ ) sama

$b_1$  : rata-rata gaji karyawan berpendidikan SD yg belum berpengalaman

$b_2$  : rata-rata kenaikan gaji karyawan jika pengalamannya bertambah 1 th,  
*cet-par*

# Contoh Output STATA (Bgm Interpretasinya?)

```
. regress Gaji Penglmn D_SL D_PT ThxD_SL ThxD_PT
```

Source	SS	df	MS	Number of obs	=	30
				F(5, 24)	=	2995.24
Model	642465562	5	128493112	Prob > F	=	0.0000
Residual	1029579.29	24	42899.1372	R-squared	=	0.9984
				Adj R-squared	=	0.9981
Total	643495141	29	22189487.6	Root MSE	=	207.12

Gaji	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Penglmn	524.4253	26.38967	19.87	0.000	469.9597	578.8909
D_SL	380.9843	252.2974	1.51	0.144	-139.732	901.7007
D_PT	1792.72	209.9112	8.54	0.000	1359.485	2225.956
ThxD_SL	484.7972	41.77671	11.60	0.000	398.5743	571.0201
ThxD_PT	958.453	34.21933	28.01	0.000	887.8278	1029.078
_cons	2238.558	142.6019	15.70	0.000	1944.243	2532.874

# Contoh Output STATA (Bgm Interpretasinya?)

. regr Gaji Penglmn D\_PT ThxD\_SL ThxD\_PT

Source	SS	df	MS	Number of obs	=	30
Model	642367740	4	160591935	F(4, 25)	=	3561.11
Residual	1127401.43	25	45096.0574	Prob > F	=	0.0000
Total	643495141	29	22189487.6	R-squared	=	0.9982
				Adj R-squared	=	0.9980
				Root MSE	=	212.36

Gaji	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Penglmn	504.4179	23.39962	21.56	0.000	456.2255	552.6104
D_PT	1671.009	198.7211	8.41	0.000	1261.735	2080.282
ThxD_SL	543.0991	16.36091	33.19	0.000	509.4032	576.795
ThxD_PT	978.4604	32.34799	30.25	0.000	911.8385	1045.082
_cons	2360.27	120.6133	19.57	0.000	2111.862	2608.678

# Dugaan Model dg Dummy Interaksi

Model:  $Y = 2360 + 504 \text{ Penglmn} + 1671 D_{PT} + 543 \text{ Peng} * D_{SL} + 978 \text{ Peng} * D_{PT}$

. estat ovtest

Ramsey RESET test using powers of the fitted values of Gaji

**H<sub>0</sub>:** model has no omitted variables

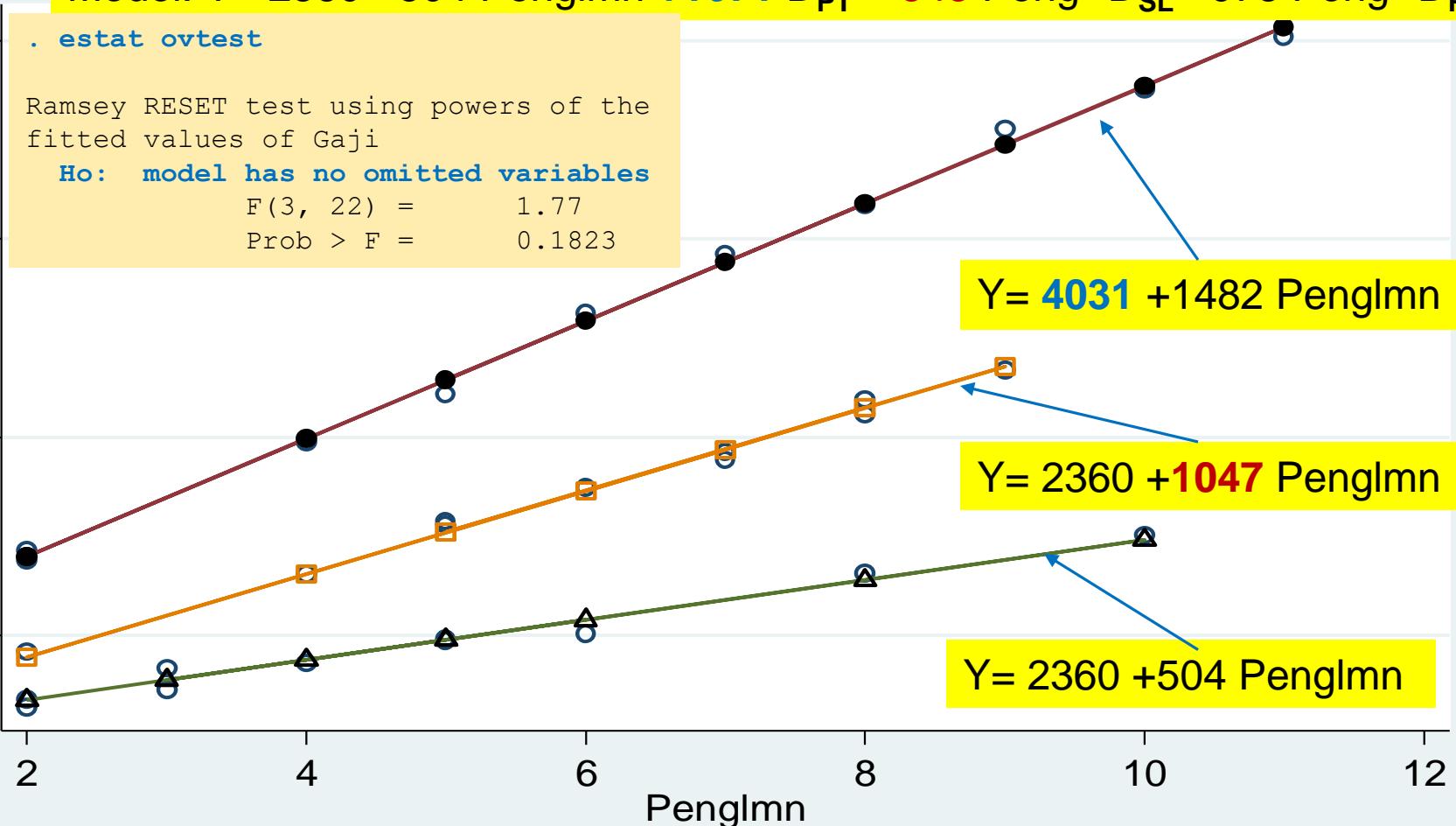
F(3, 22) = 1.77

Prob > F = 0.1823

$$Y = 4031 + 1482 \text{ Penglmn}$$

$$Y = 2360 + 1047 \text{ Penglmn}$$

$$Y = 2360 + 504 \text{ Penglmn}$$



○ Gaji

△ yhat, Pddk == SD

—●— yhat, Pddk == PT

—□— yhat, Pddk == SL