

Model dgn Peubah-*Dummy*

Oleh

Bambang Juanda

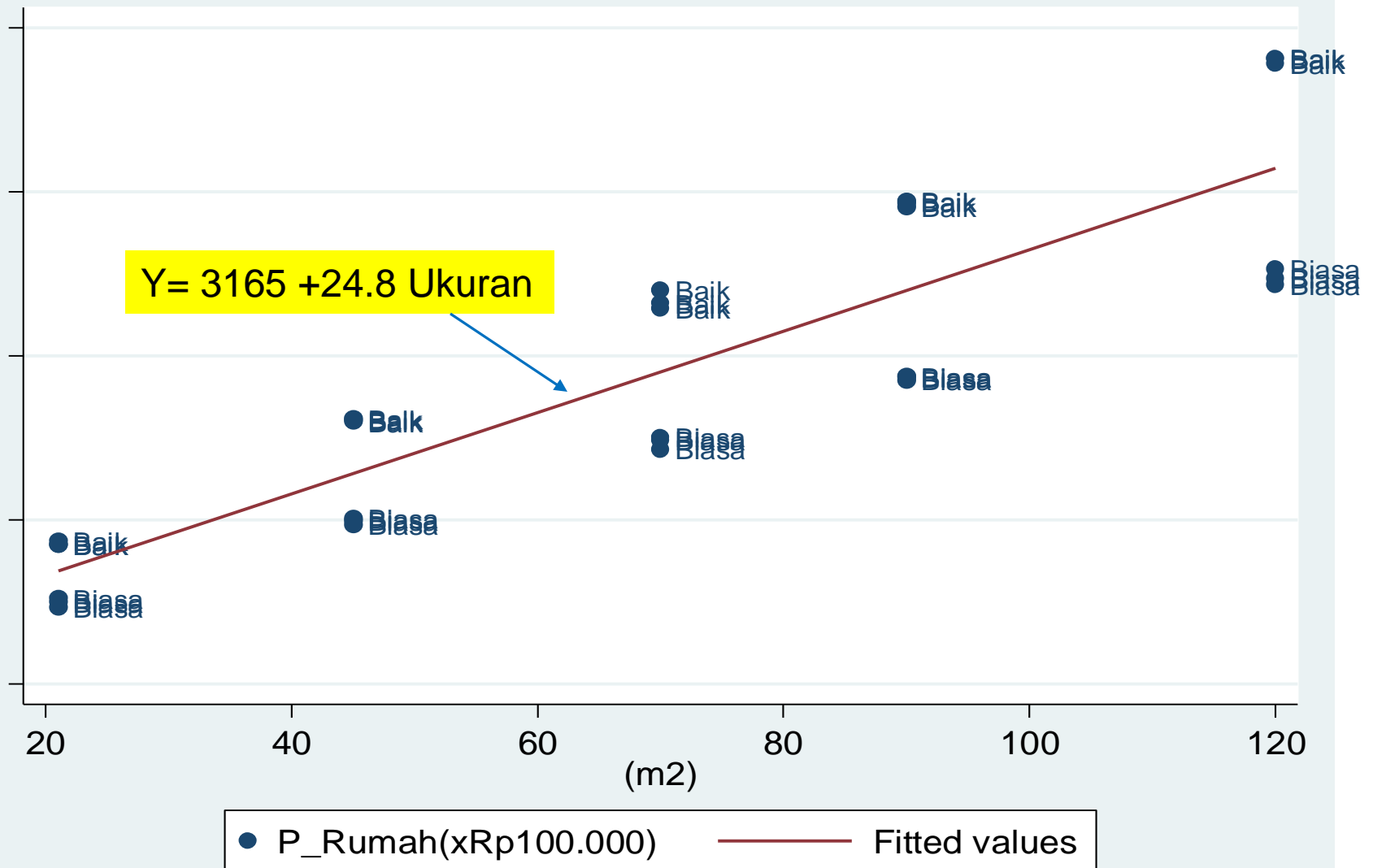
<https://bambangjuanda.com/>

Model dgn Peubah-*Dummy*

- 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:

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

Scatter Plot & Dugaan Model



Asumsi Model dgn Peubah-*Dummy*

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

Y = Perkiraan Nilai Rumah

X₂ = Ukuran Rumah (Square footage)

X₃ = Lingkungan Rumah = D

$\left\{ \begin{array}{l} 0 \text{ jika "biasa"} \\ 1 \text{ jika "baik"} \end{array} \right.$

“Baik” (X₃ = 1)

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

“Biasa” (X₃ = 0)

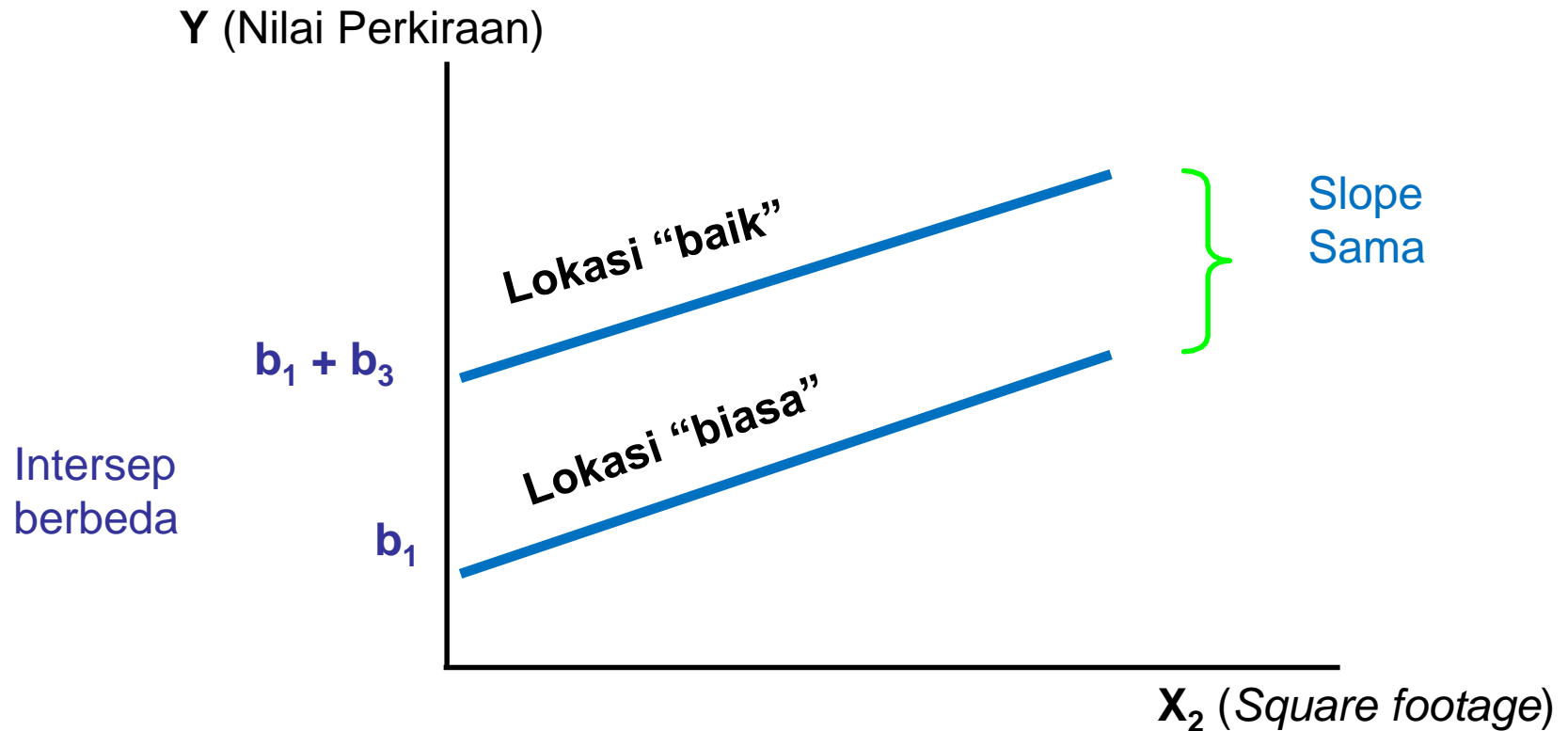
$$\hat{Y}_i = b_1 + b_2 X_{2i} + b_3 (0) = b_1 + b_2 X_{2i}$$

Slope
Sama?

b₃: 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

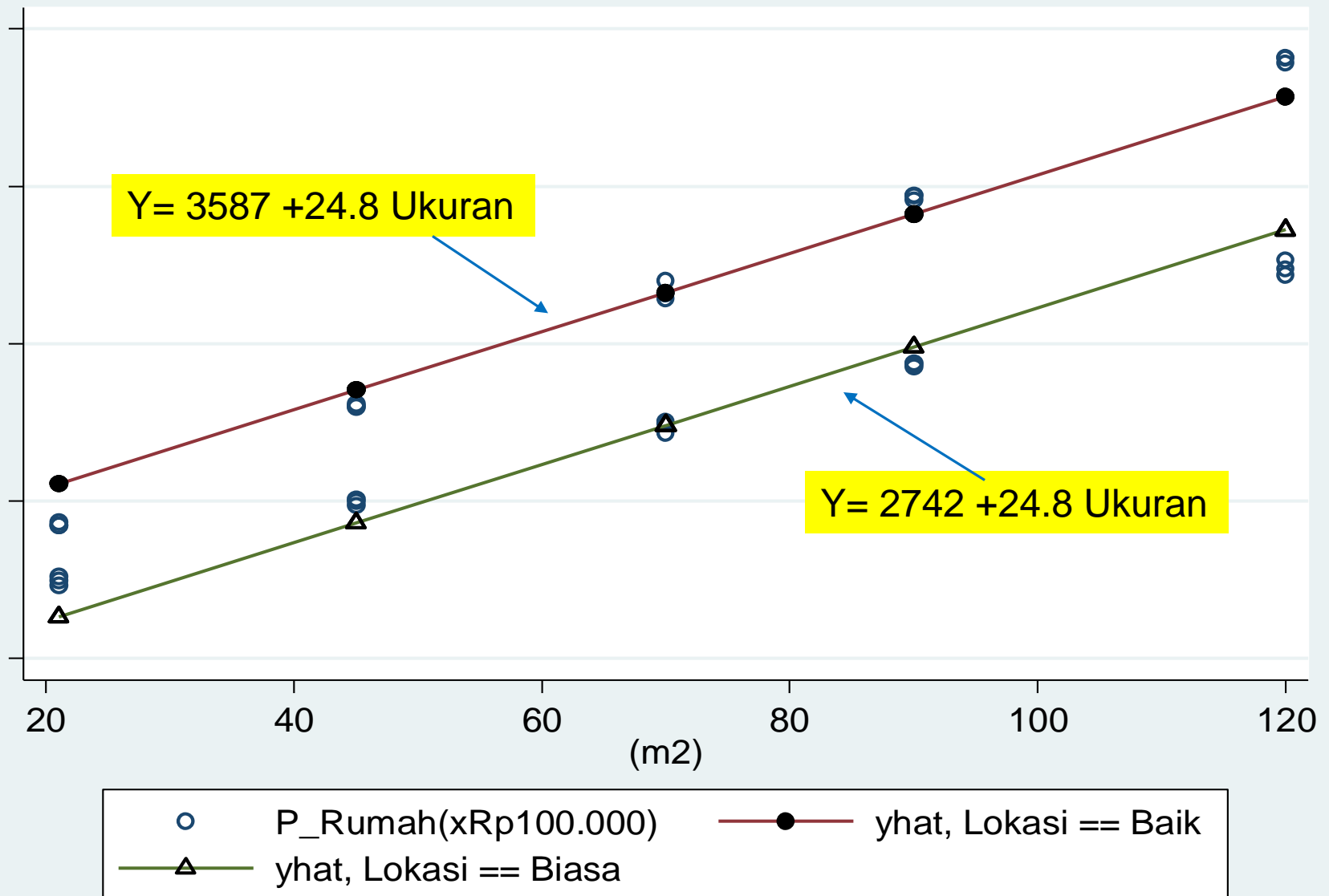
Asumsi Model dgn Peubah-*Dummy*



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 (X_3)
- Berisi Perkalian 2 Peubah

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{2i} X_{3i} + \varepsilon_i$$

- 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” ($X_3=1$)

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

Lingkungan “biasa” ($X_3=0$)

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

Slope
tdk Sama

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

Lingkungan “baik” ($X_3=1$)

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

Lingkungan “biasa” ($X_3=0$)

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

Slope
tdk Sama

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

b_2 : rata-rata perbedaan nilai rumah di lingkungan “biasa” yg ukurannya berbeda 1 feet². (*pengaruh ukuran thd nilai rumah utk lingkungan “biasa”*)

$(b_2 + b_4)$: rata-rata perbedaan nilai rumah di lingkungan “baik” yg ukurannya berbeda 1 feet². (*pengaruh ukuran thd nilai rumah utk lingkungan “baik”*)

b_4 : perbedaan pengaruh ukuran thd nilai rumah antara kedua lingkungan.

Pengaruh ukuran thd nilai rumah tergantung lingkungannya

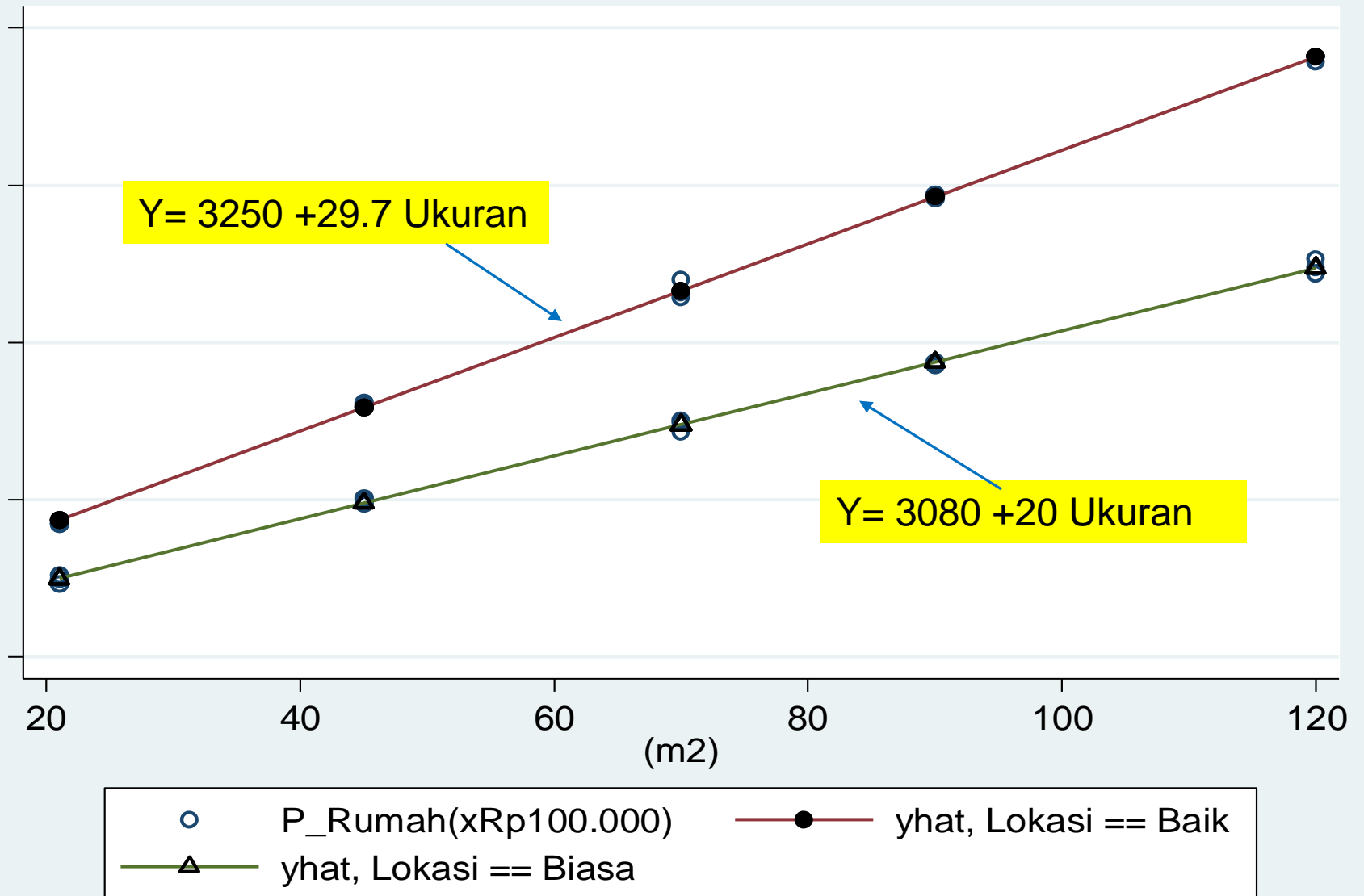
Contoh Output STATA (Bgm Interpretasinya?)

```
. regress Nilai_RatRib Ukuran_m2 Lokasi UkxLok
```

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
UkxLok	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



Peubah Kualitatif dgn 3 Kategori

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

Y_i = Gaji karyawan ke-i (ribu rupiah)

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

X_{3i} = Pendidikan karyawan ke-i = D3i =

0 jika "SD"

1 jika "SL"

2 Jika "PT"

"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 → 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" ($D_1=1$; $D_2=0$)

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

"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 !

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
Model	642465562	5	128493112	F(5, 24)	=	2995.24
Residual	1029579.29	24	42899.1372	Prob > F	=	0.0000
Total	643495141	29	22189487.6	R-squared	=	0.9984
				Adj R-squared	=	0.9981
				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