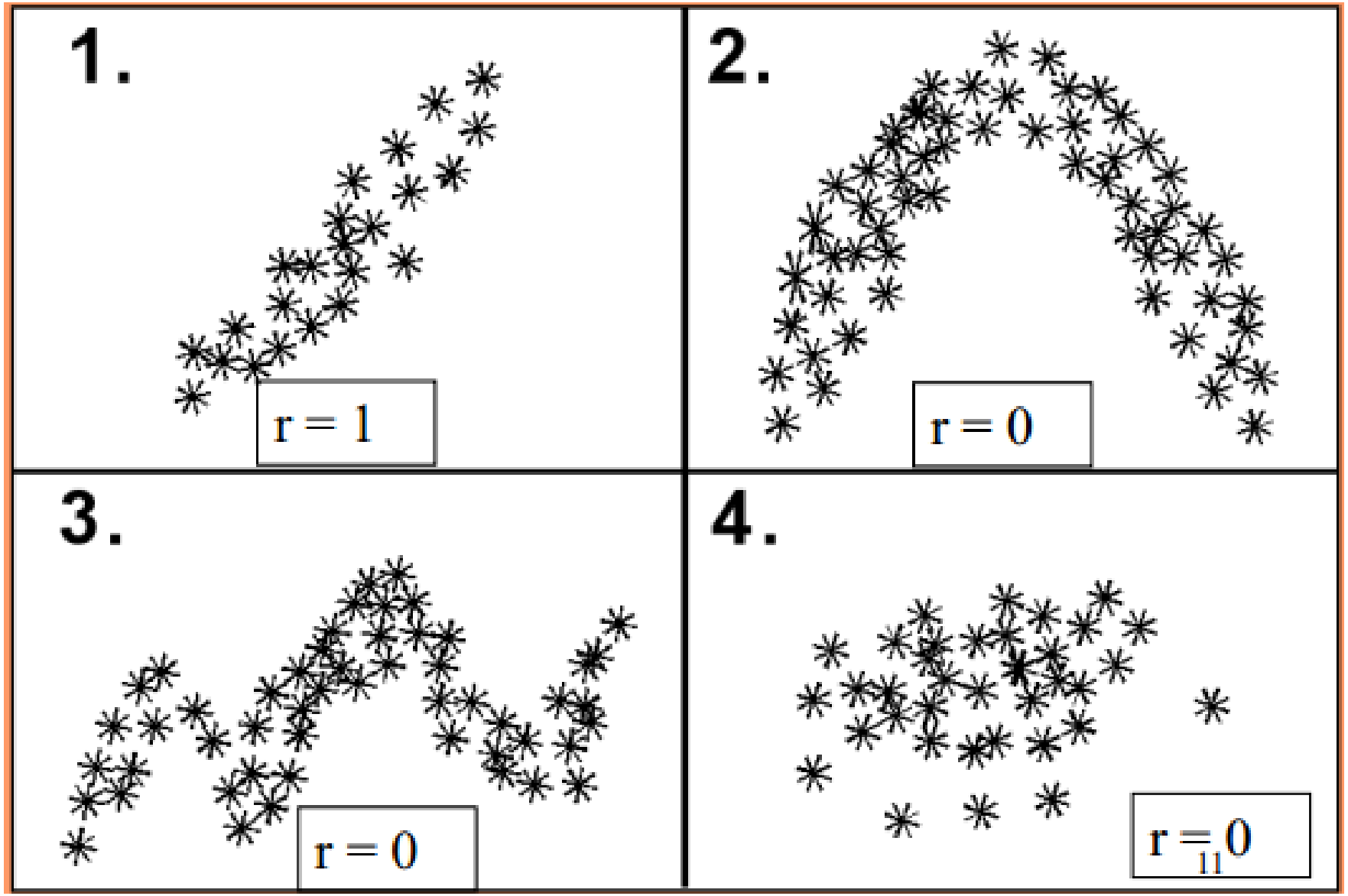
Pendahuluan

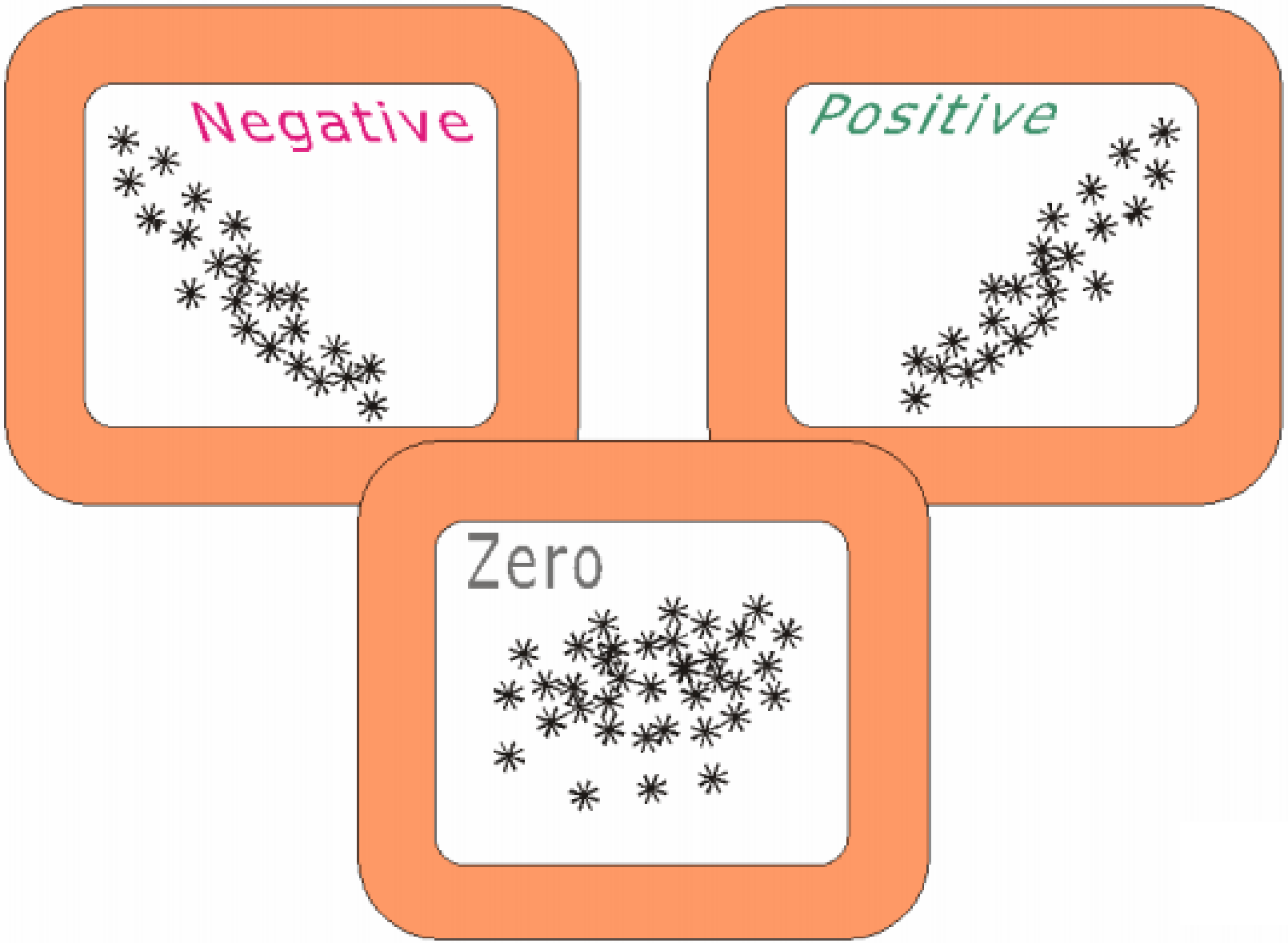
- Penelitian kuantitatif maupun kualitatif hampir selalu menggunakan variabel-variabel untuk merepresentasikan data hasil observasi.
- Antar variabel-variabel tersebut kemungkinan memiliki hubungan yang dinyatakan secara statistika sebagai nilai korelasi.
- Bentuk hubungan ini juga dapat dinyatakan sebagai suatu formulasi dengan pendekatan regresi linear.

Korelasi

Korelasi (r)



Korelasi (r)



Koefisien Korelasi (r)

- Tidak menggambarkan hubungan sebab akibat
- Nilainya berkisar antara -1 dan 1
- Tanda (+) atau (-) → arah hubungan
  - (+) searah;
  - (-) beralawanan arah
- Koefisien Korelasi Pearson → hubungan linier
- Koefisien Korelasi Spearman (rank correlation) → trend relationship

Koefisien Korelasi Pearson (r)

$$r_{xy} = \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}} \qquad S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y})$$

$$S_{xx} = \sum (x_i - \bar{x})^2 \text{ dan } S_{yy} = \sum (y_i - \bar{y})^2$$

Notasi lain:

$$S_{xx} = \Sigma x^2 - \frac{(\Sigma x)^2}{n}, \quad S_{yy} = \Sigma y^2 - \frac{(\Sigma y)^2}{n}$$

$$S_{xy} = \Sigma xy - \frac{(\Sigma x)(\Sigma y)}{n}.$$

Contoh

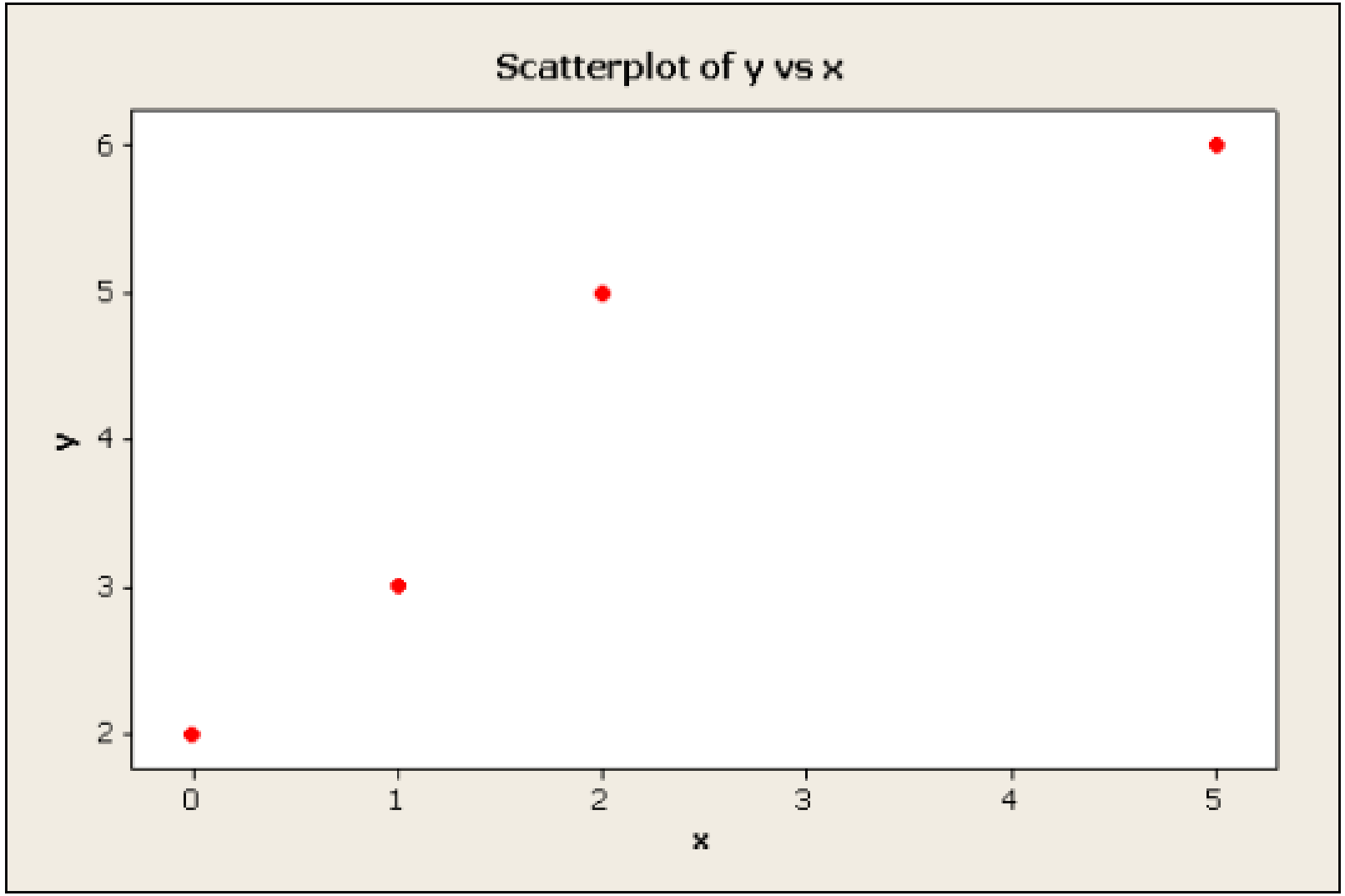
Diketahui data pengeluaran iklan (x milyar) dengan total profit penjualan suatu produk komputer perbulan (y milyar) selama 4 bulan sebagai berikut:

x	:	2	1	5	0
y	:	5	3	6	2

a. Buat scatter plot untuk data tersebut.

b. Hitung koefisien korelasinya.

(a) Scatter Plot : x dengan y

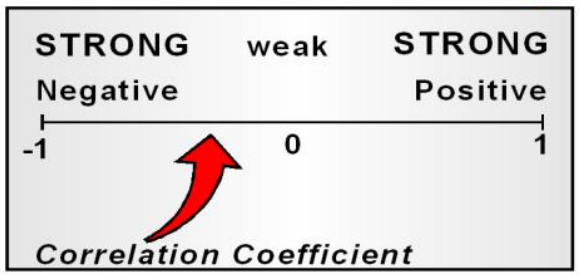


(b) Koefien Korelasi (r)

$$r_{xy} = \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}}$$

	x	y	x^2	y^2	xy
	2	5	4	25	10
	1	3	1	9	3
	5	6	25	36	30
	0	2	0	4	0
Total	8	16	30	74	43
	Σx	Σy	Σx^2	Σy^2	Σxy

$$r = \frac{43 - \frac{8 \times 16}{4}}{\sqrt{30 - \frac{8^2}{4}}\sqrt{74 - \frac{(16)^2}{4}}} = .930$$



Latihan

Mendenhall : Example 12.7, hlm. 534

The heights and weights of $n = 10$ offensive backfield football players are randomly selected from a county's football all-stars. Calculate the correlation coefficient for the heights (in inches) and weights (in pounds) given in Table 12.4.

Heights and Weights of $n = 10$ Backfield All-Stars

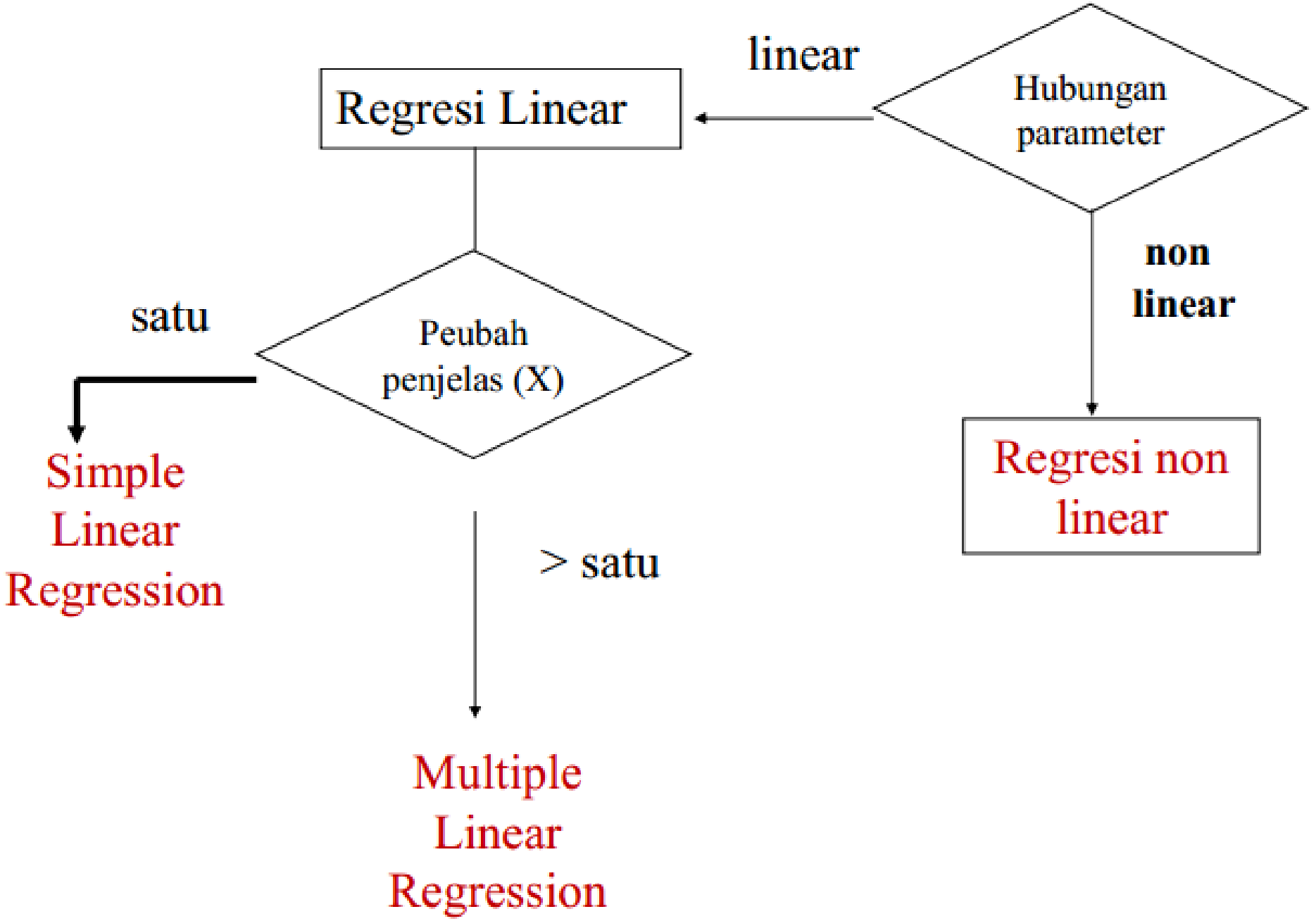
Player	Height, x	Weight, y
1	73	185
2	71	175
3	75	200
4	72	210
5	72	190
6	75	195
7	67	150
8	69	170
9	71	180
10	69	175

Regresi

Definisi

- Linier (*linear*) : linier dalam parameter
- Sederhana (*simple*) : hanya satu peubah penjelas (x)
- Berganda (*multiple*) : lebih dari satu peubah penjelas (x)

- Regresi linear adalah suatu pendekatan yang memodelkan hubungan antar variabel hasil observasi dengan membangun persamaan linear yang merupakan representasi terbaik dari data hasil observasi tersebut.
- Variabel pertama disebut predictor atau ***independent variable*** dan variabel kedua disebut response atau ***dependent variable***


Hubungan parameter
linear
non linear
Regresi Linear
Regresi non linear
Peubah penjelas (X)
satu
> satu
Simple Linear Regression
Multiple Linear Regression

ANALISIS REGRESI

- **Hubungan Antar Peubah:**
  - Fungsional (deterministik) → Y=f(X) ; misalnya: Y=10X
  - Statistik (stokastik) → amatan tidak jatuh pas pada kurva
  - Misal: IQ vs Prestasi, Berat vs Tinggi, Dosis Pupuk vs Produksi, Profit vs Biaya Iklan

- **Model regresi linear sederhana:**

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \;\; ; i = 1,2,\ldots,n$$

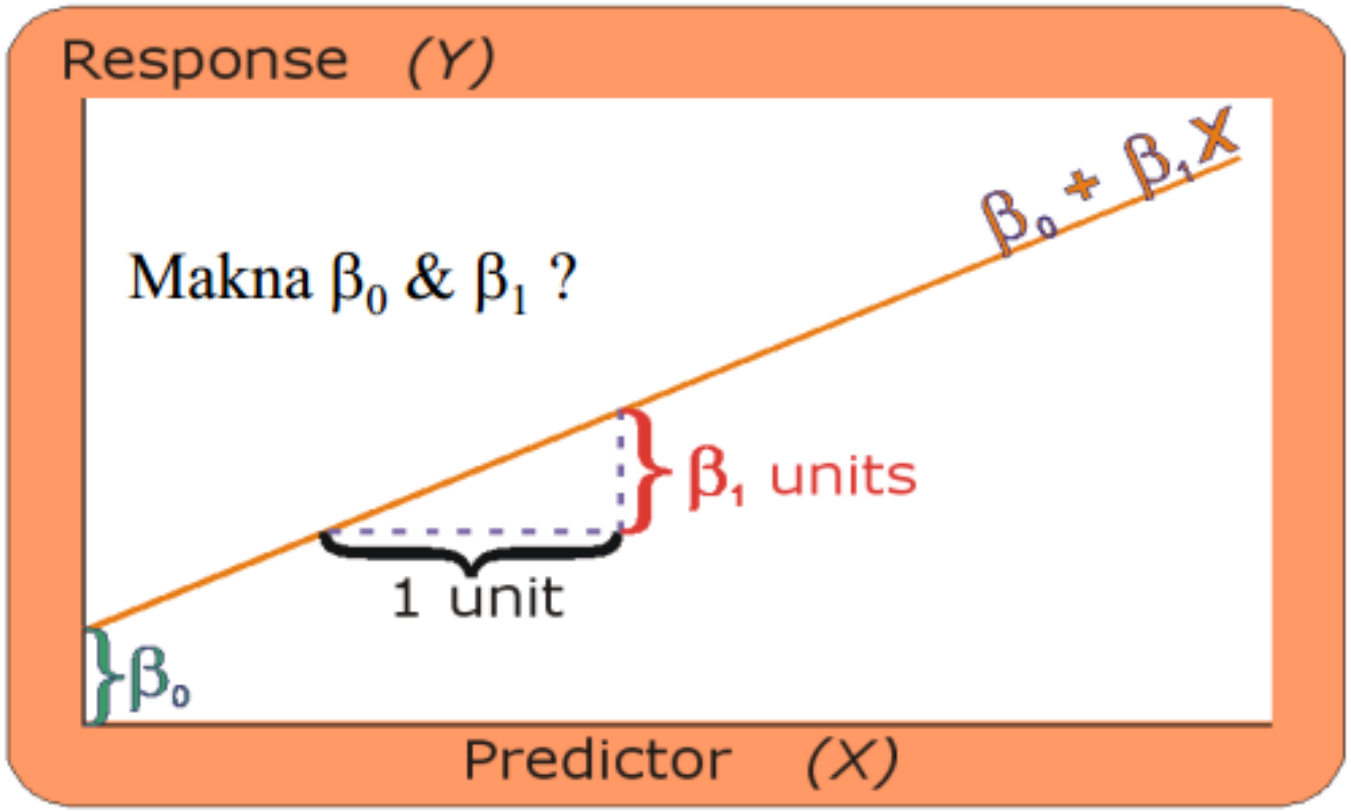
Asumsi dalam Regresi Linear

- Hubungan X dan Y adalah linear.
- Y berdistribusi normal pada setiap nilai X.
- Variance Y pada setiap nilai X adalah seragam.
- Hasil observasi masing-masing independent.

Formulasi Regresi Linear

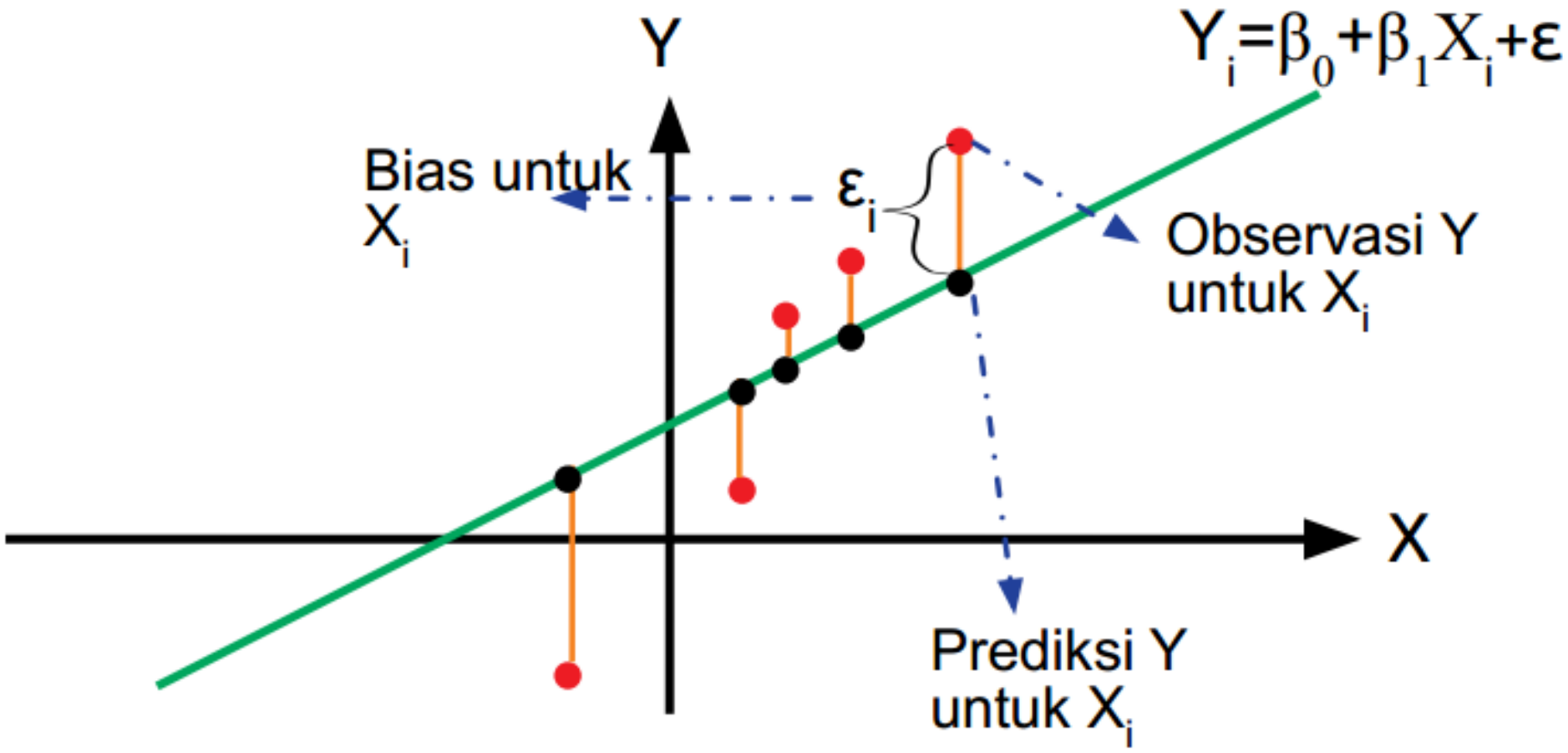
- Ekspektasi Y pada nilai X tertentu diberikan oleh persamaan $E(Y_i/X_i) = \beta_0 + \beta_1 X_i$
- Nilai prediksi individual $Y = \beta_0 + \beta_1 X_i + \varepsilon$
- β_0 dan β_1 adalah parameter regresi.
- Nilai ε adalah bias yang diasumsikan independent dan berdistribusi $N(0, \sigma^2)$.

Simple Linear Regression Model



Interpretasi : β_0 adalah nilai Y ketika X = 0, sedangkan β_1 adalah perubahan nilai Y untuk setiap perubahan X sebesar satu satuan unit.

Estimasi garis regresi dengan meminimalkan total kuadrat bias.



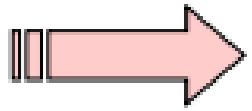
Analisis Regresi

- Pendugaan terhadap koefisien regresi:

  → b_0 penduga bagi β_0 dan b_1 penduga bagi β_1

$$b_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum xy - \frac{(\sum x)(\sum y)}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$



Metode Kuadrat Terkecil (Least Square)

Analisis Regresi

Bagaimana Pengujian terhadap model regresi ??

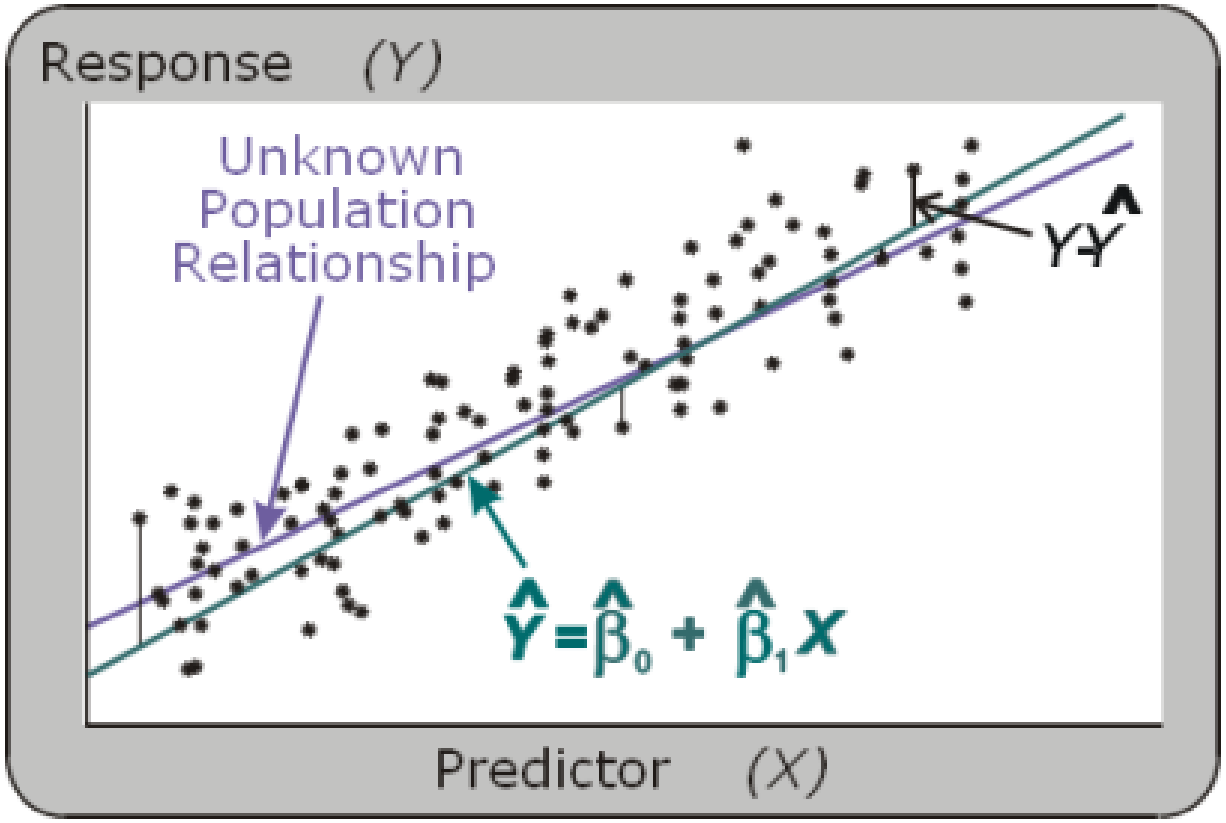
- parsial (per koefisien) → uji-t
- bersama → uji-F (Anova)

Bagaimana menilai kesesuaian model ??

- R^2 (Koefisien Determinasi: persentase keragaman Y yang mampu dijelaskan oleh X)

Metoda Kuadrat Terkecil

- Pendugaan parameter pada regresi didapat dengan meminimumkan jumlah kuadrat galat (*error*).

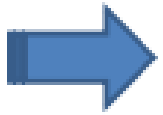


Contoh Data

Apakah semakin besar biaya iklan yang dikeluarkan akan semakin besar pula profit yang diperoleh?

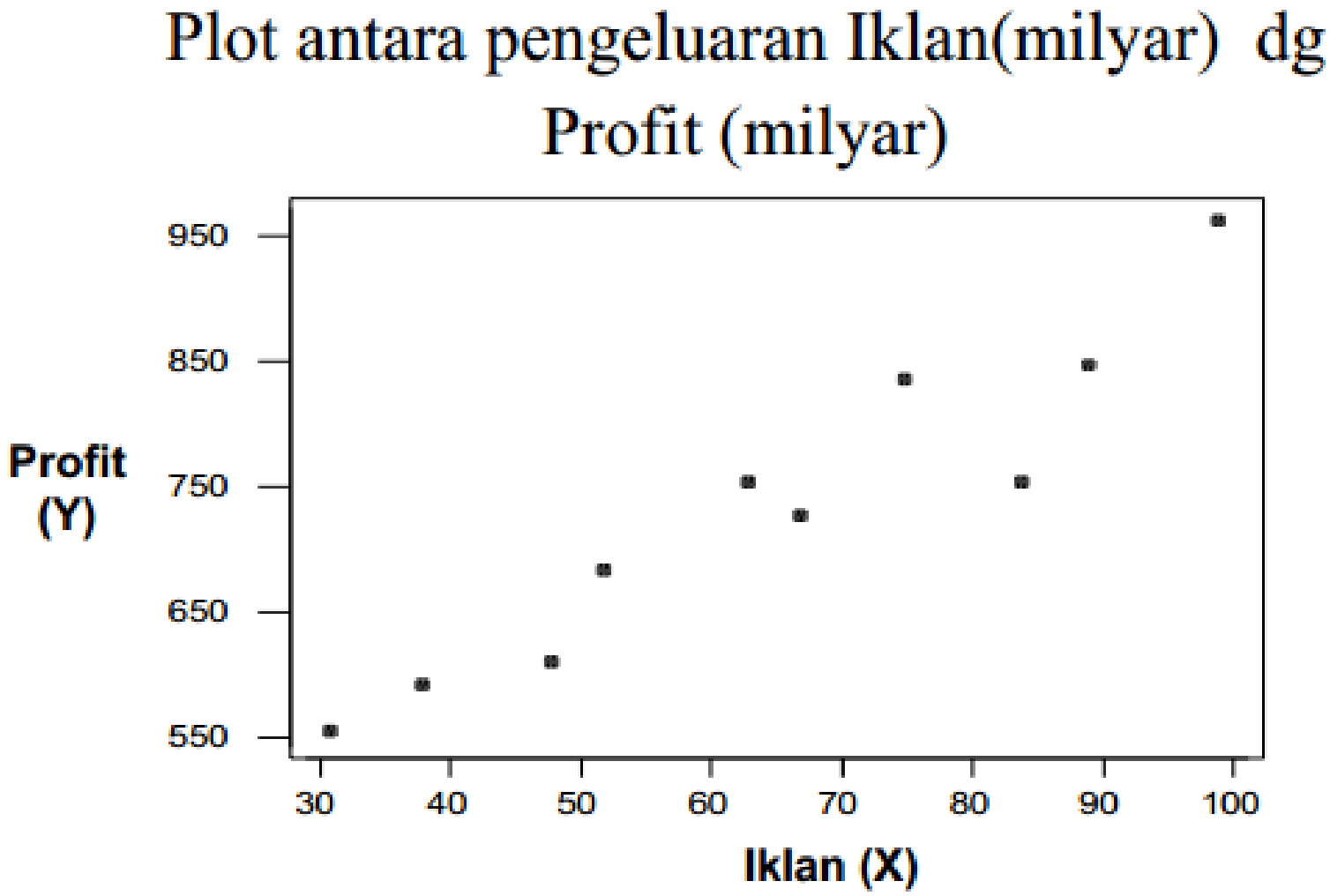
Diamati contoh acak 10 perusahaan yang memproduksi *Laptop*, kemudian dicatat pengeluaran iklan (dalam milyar) dan profit (dalam milyar) selama tahun 2015.

a. Buat scatter plotnya dan jelaskan.
b. Tentukan persamaan model regresinya.
c. Tentukan penduga bagi parameter model regresi tersebut.



Iklan	Profit
31	553
38	590
48	608
52	682
63	752
67	725
75	834
84	752
89	845
99	960

Penyelesaian :



$$\text{Model: } Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \;\; ; i = 1,2,\ldots,n$$

Penyelesaian :

	x	y	x^2	y^2	xy
	31	553	961	305,809	17,143
	38	590	1,444	348,100	22,420
	48	608	2,304	369,664	29,184
	52	682	2,704	465,124	35,464
	63	752	3,969	565,504	47,376
	67	725	4,489	525,625	48,575
	75	834	5,625	695,556	62,550
	84	752	7,056	565,504	63,168
	89	845	7,921	714,025	75,205
	99	960	9,801	921,600	95,040
Jumlah	**646**	**7,301**	**46,274**	**5,476,511**	**496,125**

$$b_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum xy - \frac{(\sum x)(\sum y)}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

Sxy	24,480.4
Sxx	4,542.4
Syy	146,050.9
b_1	**5.39**
b_0	**381.95**

Uji Hipotesis (1) $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$

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

S_{b1} disebut galat baku (*standard of error*) bagi b_1 → $SE(b_1)$

$$S_{b_1} = \sqrt{\frac{s^2}{\sum (x_i - \bar{x})^2}}$$

$$s^2 = \frac{SSE}{n-2}$$

Tolak Ho jika:
$|t\text{-hit}| > t(\alpha/2; db=n-2)$

$$SSE = \Sigma(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = S_{yy} - \frac{S_{xy}^2}{S_{xx}}$$

Uji Hipotesis (2) $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 > 0$

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

$$S_{b_1} = \sqrt{\frac{s^2}{\sum (x_i - \bar{x})^2}}$$

$$s^2 = \frac{SSE}{n-2}$$

Tolak Ho jika:
t-hit > t(α; db=n-2)

$$SSE = \Sigma(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = S_{yy} - \frac{S_{xy}^2}{S_{xx}}$$

Uji Hipotesis (3) $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 < 0$

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

$$S_{b_1} = \sqrt{\frac{s^2}{\sum (x_i - \bar{x})^2}}$$

$$s^2 = \frac{SSE}{n-2}$$

Tolak Ho jika:
t-hit < - t(α; db=n-2)

$$SSE = \Sigma(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = S_{yy} - \frac{S_{xy}^2}{S_{xx}}$$

Latihan (1)

- Apakah iklan berpengaruh pada profit perusahaan? Uji hipotesis Anda pada taraf nyata $\alpha = 0.05$

Jawaban Ringkas $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$

Tolak Ho jika:
$|t\text{-hit}| > t(\alpha/2; db=n-2)$

(4). $t - hitung = \dfrac{b_1 - \beta_1}{S_{b_1}} = \dfrac{b_1 - 0}{S_{b_1}}$ = 5.39/0.623 = **8.64**

(3). $S_{b_1} = \sqrt{\dfrac{s^2}{\sum (x_i - \bar{x})^2}}$ = $\sqrt{(S^2/S_{xx})}$ = $\sqrt{(1{,}764.81/4{,}542.4)}$ = 0.623

(2). $s^2 = \dfrac{SSE}{n-2}$ = 1,764.81

(1). $SSE = \Sigma(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = S_{yy} - \dfrac{S_{xy}^2}{S_{xx}}$ = 14,118.45

Sxy	24,480.4
Sxx	4,542.4
Syy	146,050.9
b_1	**5.39**
b_0	**381.95**

$t(\alpha/2; db=n-2) = t(0.025; 8)$ = **2.306**

Karena (t-hit = 8.64) > 2.306 maka **TOLAK H_0**, artinya iklan berpengaruh pada profit perusahaan untuk taraf uji α = 0.05

t-test table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	**0.50**	**0.25**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**	**0.0005**
two-tails	**1.00**	**0.50**	**0.40**	**0.30**	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.002**	**0.001**
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

Latihan (2)

- Apakah semakin besar iklan akan mengakibatkan semakin besar profit? Uji pada taraf nyata $\alpha = 0.05$

Jawaban Ringkas $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 > 0$

Tolak Ho jika:
t-hit > t(α; db=n-2)

(4). $t - hitung = \frac{b_1 - \beta_1}{S_{b_1}} = \frac{b_1 - 0}{S_{b_1}}$ = 5.39/0.623 = **8.64**

(3). $S_{b_1} = \sqrt{\frac{s^2}{\sum (x_i - \bar{x})^2}}$ = $\sqrt{(S^2/S_{xx})}$ = $\sqrt{(1{,}764.81/4{,}542.4)}$ = 0.623

(2). $s^2 = \frac{SSE}{n-2}$ = 1,764.81

(1). $SSE = \Sigma(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = S_{yy} - \frac{S_{xy}^2}{S_{xx}}$ = 14,118.45

Sxy	24,480.4
Sxx	4,542.4
Syy	146,050.9
b_1	**5.39**
b_0	**381.95**

t(α; db=n-2) = t(0.05; 8) = **1.860**

Karena (t-hit = 8.64) > 1.860 maka **TOLAK H_0**, artinya semakin besar iklan akan mengakibatkan semakin besar profit untuk taraf uji α = 0.05

t-test table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	**0.50**	**0.25**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**	**0.0005**
two-tails	**1.00**	**0.50**	**0.40**	**0.30**	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.002**	**0.001**
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

Koefisien Determinasi (R^2)

Sxy	24,480.4
Sxx	4,542.4
Syy	146,050.9
b_1	5.39
b_0	381.95

R^2 = (JK Regresi)/(JK Total)

$$= \frac{\frac{(S_{xy})^2}{S_{xx}}}{S_{yy}} = \frac{(S_{xy})^2}{S_{xx}S_{yy}} = 0.903 = \mathbf{90.3\%}$$

Artinya, **90.3** persen keragaman pada *Y* (profit) dapat diterangkan oleh keragaman pada *X* (iklan)

Keterbatasan Korelasi dan Regresi Linear

- Korelasi dan Regresi Linear hanya menggambarkan hubungan yang **linear**
- Korelasi dan metode kuadrat terkecil pada regresi linear tidak resisten terhadap **pencilan**
- Prediksi **di luar selang nilai X** tidak diperkenankan karena kurang akurat
- Hubungan antara dua variabel bisa dipengaruhi oleh **variabel lain** di luar model

Catatan

- Apa itu analisis regresi?
- Apa bedanya dengan korelasi?

Analisis Regresi → Analisis statistika yang memanfaatkan hubungan antara dua atau lebih peubah kuantitatif sehingga salah satu peubah dapat diramalkan dari peubah lainnya.

Korelasi → mengukur keeratan HUBUNGAN LINEAR dari dua variabel

Latihan

Mendenhall : Example 12.7, hlm. 534

The heights and weights of $n = 10$ offensive backfield football players are randomly selected from a county's football all-stars. Calculate the correlation coefficient for the heights (in inches) and weights (in pounds) given in Table 12.4.

Heights and Weights of $n = 10$ Backfield All-Stars

Player	Height, x	Weight, y
1	73	185
2	71	175
3	75	200
4	72	210
5	72	190
6	75	195
7	67	150
8	69	170
9	71	180
10	69	175

Apakah semakin tinggi seorang pemain football, maka akan semakin berat bobotnya? Uji pada taraf nyata $\alpha = 0.05$