KORELASI DAN REGRESI LINEAR

HERTANTO WAHYU SUBAGIO

Korelasi

* menunjukkan arah hubungan
* Uji :

  r product moment Pearson

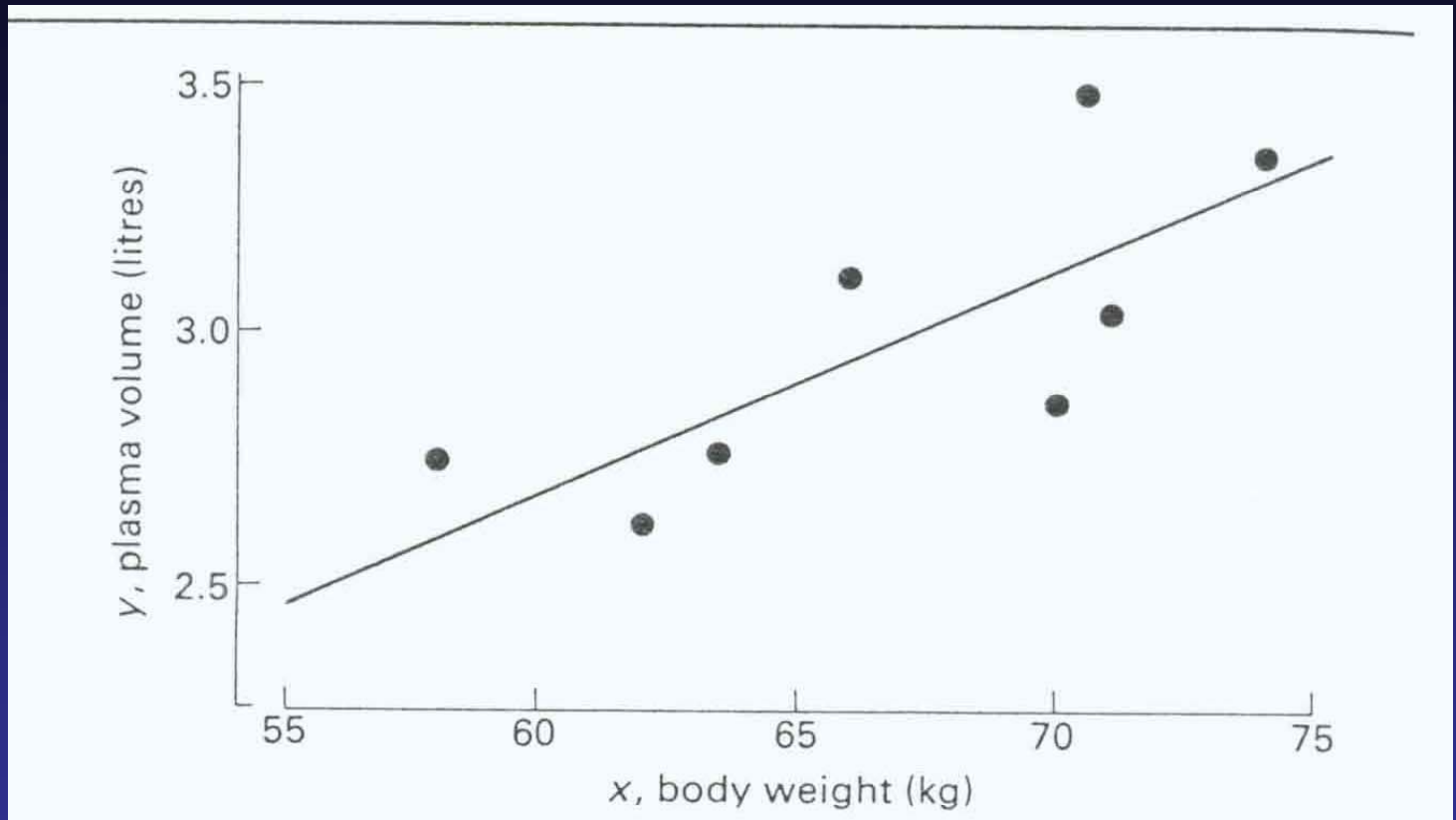
  Spearman

  Kendall
* Nilai : -1 s/d +1

Correlation between Plasma volume and body weight in 8 healthy men

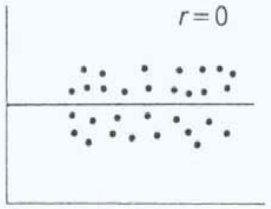
Subject	Body weight (kg)	Plasma vol. (l)
1.	58.0	2.75
2.	70.0	2.86
3.	74.0	3.37
4.	63.5	2.76
5.	62.0	2.62
6.	70.5	3.49
7.	71.0	3.05
8.	66.0	3.12

Scatter Plot

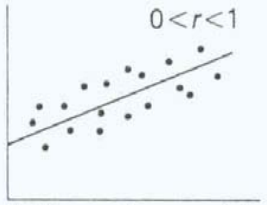


gure 9.1 Scatter diagram of plasma volume and body weight showing the linear regression
e.

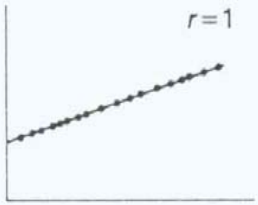
Some possibilities in scatter plot



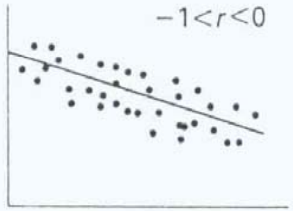
(a) No correlation



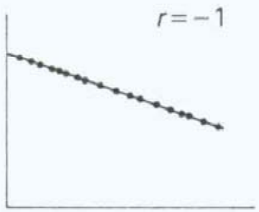
(b) Imperfect positive correlation



(c) Perfect positive correlation



(d) Imperfect negative correlation



(e) Perfect negative correlation

Figure 9.2 Scatter diagrams illustrating different values of the correlation coefficient. Also shown are the regression lines.

Formula

$$r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{[\Sigma(x-\bar{x})^2 \Sigma(y-\bar{y})^2]}}$$

SPSS output

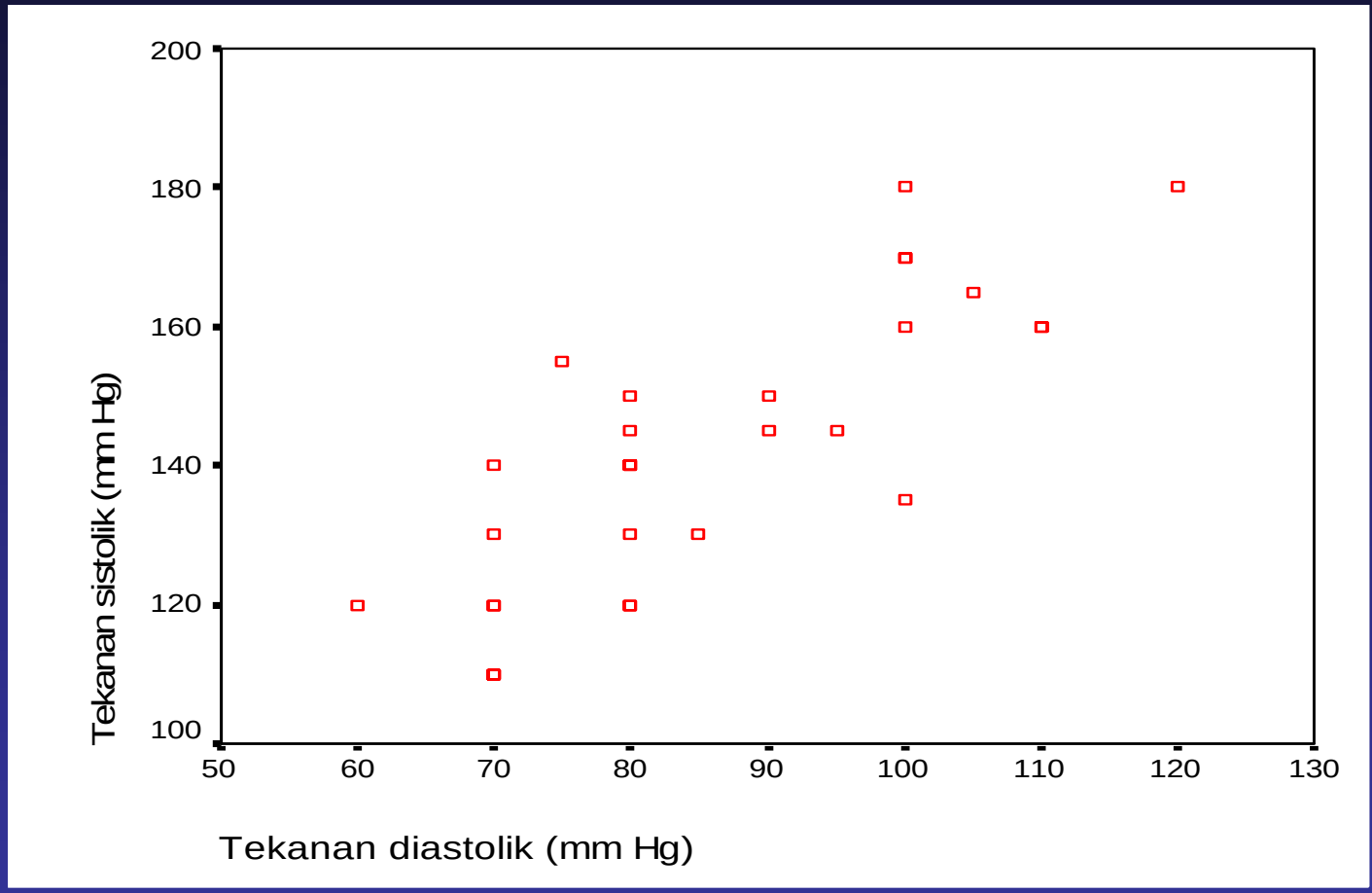
Correlations

		Tekanan sistolik (mm Hg)	Tekanan diastolik (mm Hg)
Tekanan sistolik (mm Hg)	Pearson Correlation	1,000	,814**
	Sig. (2-tailed)	,	,000
	N	30	30
Tekanan diastolik (mm Hg)	Pearson Correlation	,814**	1,000
	Sig. (2-tailed)	,000	,
	N	30	30

**. Correlation is significant at the 0.01 level (2-tailed).

X mempengaruhi Y ?

Diagram Pencar



Regresi

* Mampu memprediksi DV dari perubahan IV
* Uji :
  - regresi sederhana : Y=a+bX.
  - regresi majemuk : Y = a + b1X1 + b2X2 …….biXi
* Tidak otomatis menunjukkan sebab akibat

Linear regression

- Gives the equation of the straight line that best describes it and enables the prediction of one variable from the other

The equation is : $y = a + bx$

a = intercept

b = slope = regression coefficient

y = dependent var.

x = independent var.

The values for a and b are calculated so as to minimize the sum of the squared vertical distances of the points from the line. This is called a least square fit.

Linear Regression Line

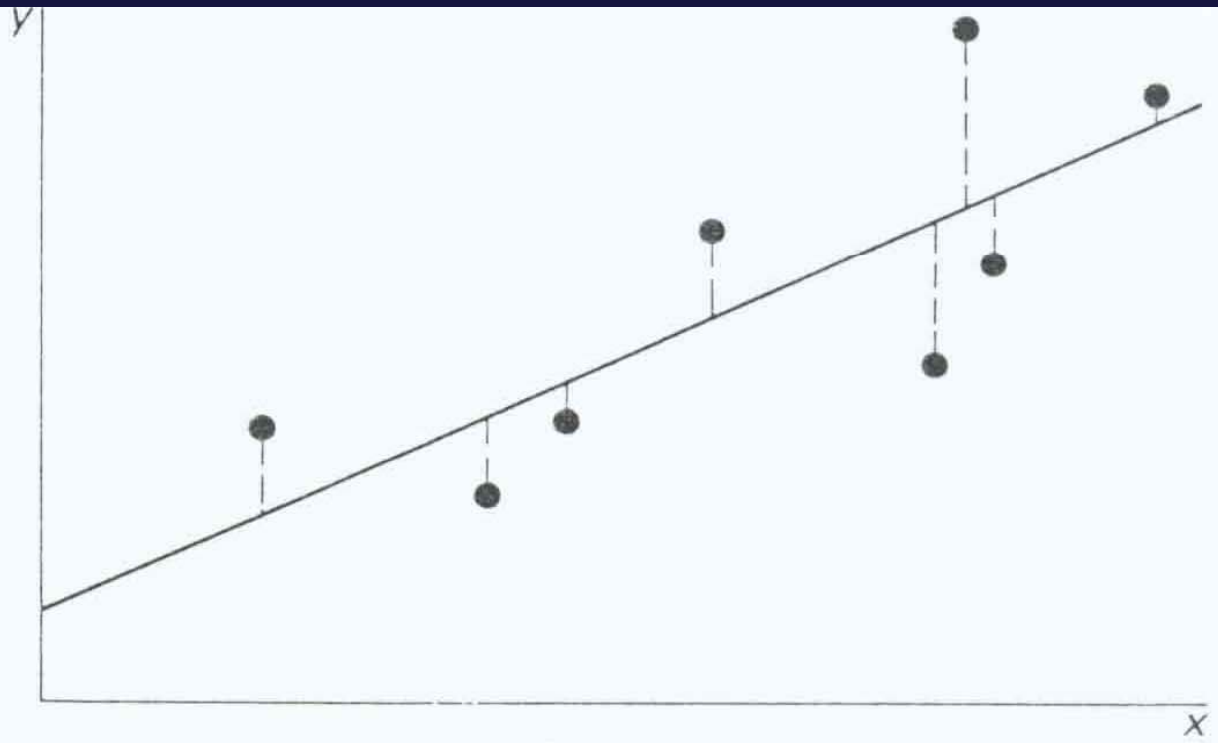


Figure 9.4 Linear regression line, $y = a + bx$, fitted by least squares. a and b are calculated to minimize the sum of squares of the vertical deviations (shown by the dotted lines) of the points about the line; each deviation equals the difference between the observed value of y and the corresponding point on the line, $a + bx$.

The intercept and slope of regression equation y=a+bx

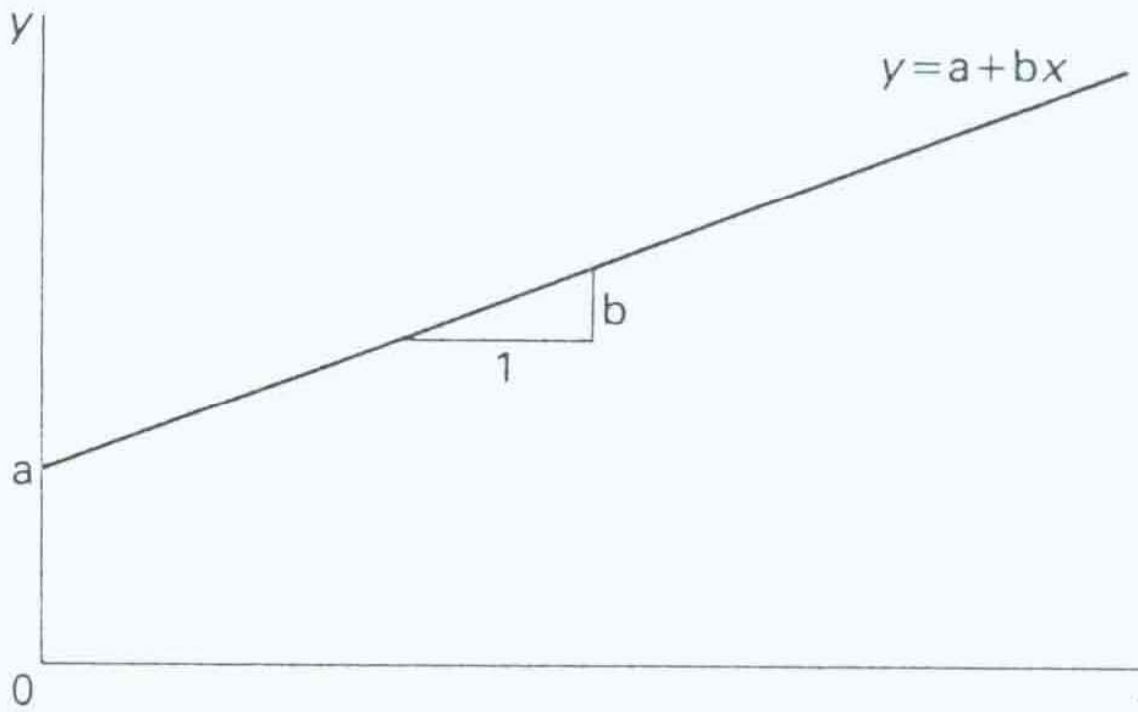


Figure 9.3 The intercept and slope of the regression equation, $y = a + bx$. The intercept, a, is the point where the line crosses the y axis and gives the value of y for $x = 0$. The slope, b, is the increase in y corresponding to a unit increase in x.

Plasma vol=0.086+0.0.044 X bodyweight

SPSS Output :

- Descriptive statistics
- Correlations
- Variables entered/removed
- Model summary
- Anova
- Coefficients

Descriptive Statistics

	Mean	Std. Deviation	N
Tekanan sistolik (mm Hg)	141,67	20,57	30
Tekanan diastolik (mm Hg)	85,67	15,13	30

Correlations

		Tekanan sistolik (mm Hg)	Tekanan diastolik (mm Hg)
Pearson Correlation	Tekanan sistolik (mm Hg)	1,000	,814
	Tekanan diastolik (mm Hg)	,814	1,000
Sig. (1-tailed)	Tekanan sistolik (mm Hg)	,	,000
	Tekanan diastolik (mm Hg)	,000	,
N	Tekanan sistolik (mm Hg)	30	30
	Tekanan diastolik (mm Hg)	30	30

r = 0.814

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	Tekanan diastolik (mm Hg)[a]	,	Enter

a. All requested variables entered.

b. Dependent Variable: Tekanan sistolik (mm Hg)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,814[a]	,662	,650	12,17

a. Predictors: (Constant), Tekanan diastolik (mm Hg)

R^2 = Koefisien determinasi

= Sumbangan var bebas terhadap kejadian var dependent

66.2% kejadian var dependent ditentukan oleh var bebas

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	8121,557	1	8121,557	54,861	,000[a]
	Residual	4145,109	28	148,040		
	Total	12266,667	29			

a. Predictors: (Constant), Tekanan diastolik (mm Hg)

b. Dependent Variable: Tekanan sistolik (mm Hg)

F hitung = 54.89, p = 0.000

Model regresi dapat dipakai untuk memprediksi var dependent

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	46,900	12,986		3,612	,001
	Tekanan diastolik (mm Hg)	1,106	,149	,814	7,407	,000

a. Dependent Variable: Tekanan sistolik (mm Hg)

Y = 46.9 + 1.106 X

Sistole = 46.9 + 1.106 diastole

X mempengaruhi Y : ?