Analisis Regresi

Dr. Agr. Sc. Ernoiz Antriyandarti, SP, MP, M.Ec

Historical origin of the term “Regression”

- Istilah REGRESI diperkenalkan oleh Francis Galton
- Kecenderungan orang tua yang tinggi untuk memiliki anak yang tinggi dan untuk orang tua yang pendek memiliki anak yang pendek, tetapi tinggi rata-rata anak yang lahir dari orang tua dari ketinggian tertentu cenderung untuk bergerak (atau mengalami kemunduran) menuju ketinggian rata-rata dalam populasi secara keseluruhan (F. Galton, “Family Likeness in Stature”)

Historical origin of the term “Regression”

- Hukum Galton dikukuhkan oleh Karl Pearson: Tinggi rata-rata putra dari sekelompok ayah tinggi < tinggi ayah mereka. Dan tinggi rata-rata putra dari sekelompok ayah pendek > tinggi ayah mereka. Dengan demikian "mundur" putra-putra yang tinggi dan pendek sama-sama menuju tinggi rata-rata semua pria. (K. Pearson dan A. Lee, "Tentang Hukum Waris")
- Dengan kata-kata Galton, disebut “*Regression to mediocrity*”

Modern Interpretation of Regression Analysis

- *Cara modern dalam menafsirkan Regresi: Analisis Regresi berkaitan dengan studi tentang ketergantungan satu variabel (Variabel Dependen), pada satu atau lebih variabel lain (Variabel Penjelasan), dengan maksud untuk memperkirakan dan / atau memprediksi nilai rata-rata (populasi) atau rata-rata dari yang pertama dalam hal nilai yang diketahui atau tetap (dalam pengambilan sampel berulang) dari yang terakhir.*

Dependent Variable Y; Explanatory Variable Xs

1. Y = Son's Height; X = Father's Height
2. Y = Height of boys; X = Age of boys
3. Y = Personal Consumption Expenditure
X = Personal Disposable Income
4. Y = Demand; X = Price
5. Y = Rate of Change of Wages
X = Unemployment Rate
6. Y = Money/Income; X = Inflation Rate
7. Y = % Change in Demand; X = % Change in the advertising budget
8. Y = Crop yield; Xs = temperature, rainfall, sunshine, fertilizer

Regresi Linier

- Regresi linier sering digunakan untuk melihat nilai prediksi atau perkiraan yang akan datang

- Apabila X dan Y mempunyai hubungan, maka nilai X yang sudah diketahui dapat digunakan memperkirakan Y

- Perkiraan mengenai terjadinya sesuatu kejadian (nilai variabel untuk waktu yang akan datang, seperti prediksi produksi 3 tahun yang akan datang, prediksi harga bulan depan, ramalan jumlah penduduk 10 tahun mendatang, ramalan hasil penjualan tahun depan).

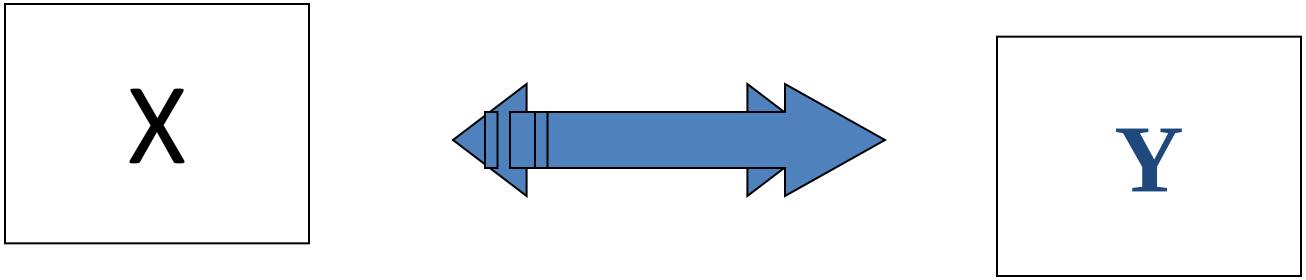
- Ramalan mengetahui suatu kejadian baik secara kualitatif (akan turun hujan, akan terjadi perang, akan lulus ujian)

- Kuantitatif (produksi padi akan mencapai 16 juta ton, indek harga 9 bahan pokok naik 10%, penerimaan devisa turun 5%)

- Melakukan peramalan adalah dengan mengunakan garis regresi

→Variable Y yang nilainya akan diramalkan disebut variable tidak bebas (*dependent variable*)

→ sedangkan variable X yang nilainya digunakan untuk meramalkan nilai Y disebut variable bebas (*independent variable*) atau variable peramal (*predictor*) dan sering kali disebut variable yang menerangkan (*exsplanatory*).



Prediktor

variabel indipenden

Variabel respon

Variabel dependen

Adakah korelasi/ hubungannya nya ?

Analisis Regresi

Dapatkah variabel X memprediksi Y ?

Analisis regresi digunakan untuk mengetahui bagaimana variabel dependen atau kriterium dapat diprediksikan melalui variabel independen atau prediktor secara individu atau parsial maupun secara bersama-sama atau simultan.

Ilustrasi hubungan positif

X

Pupuk
Berat
Badan

Y

Produksi
Tekanan
darah

Ilustrasi hubungan negatif

X

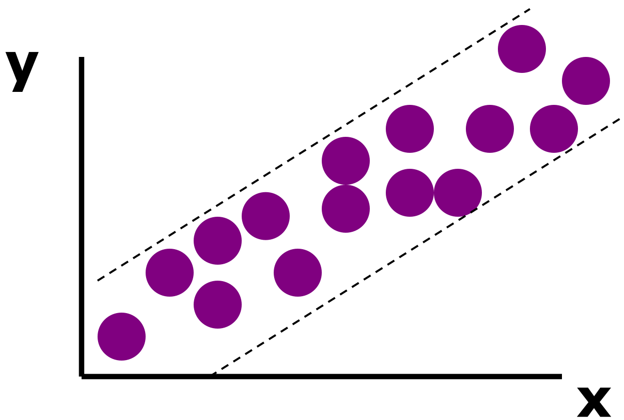
Jumlah aseptor
Harga suatu barang

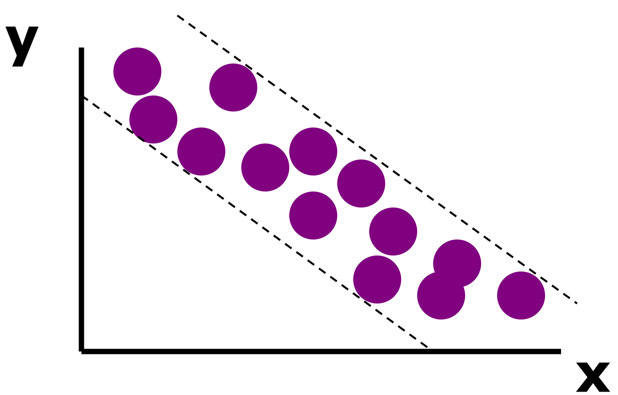
Y

Jumlah kelahiran
Permintaan barang

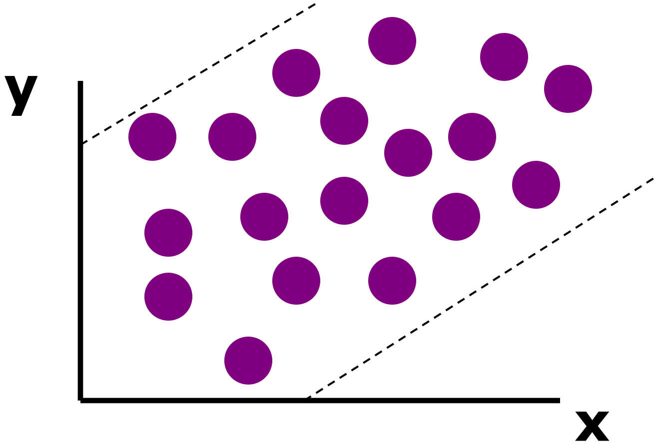
Scatter Plot Examples

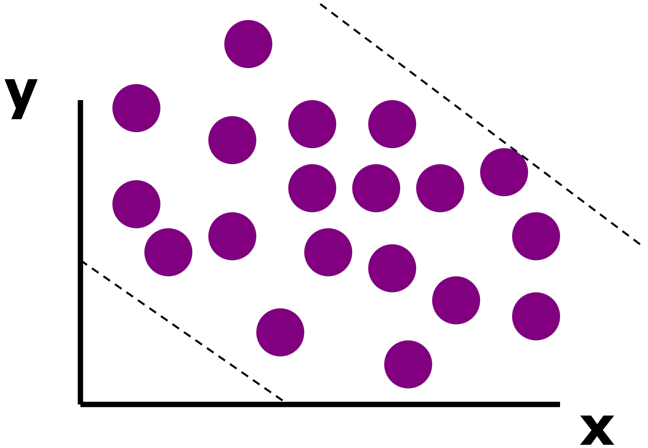
Strong relationships





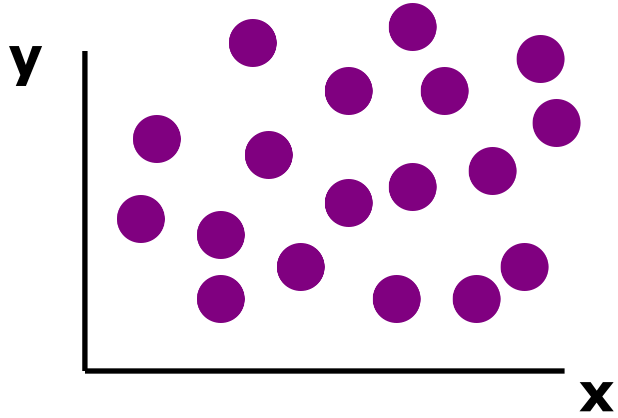
Weak relationships

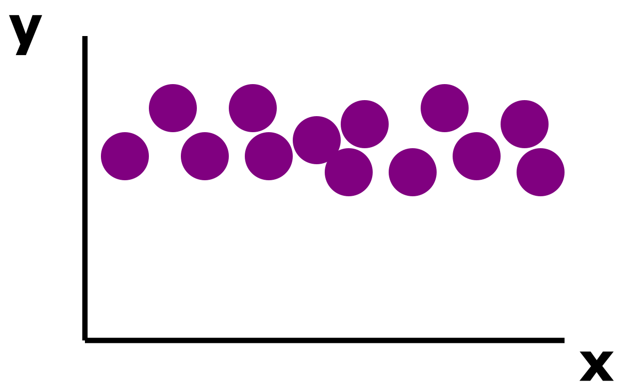




Scatter Plot Examples

No relationship





I. **Regresi linier jika hubungan antara variabel bebas terhadap variabel tak bebas berbentuk linier**

- Regresi linier sederhana → $\hat{Y} = a + bX$
- Regresi linier berganda → $\hat{Y} = a + b_1 X_1 + b_2 X_2 + b_3 X_3$

II. **Regresi tak linier jika hubungan antara variabel bebas terhadap variabel tak berbentuk linier**

$$\hat{Y} = a + bX + cX^2$$

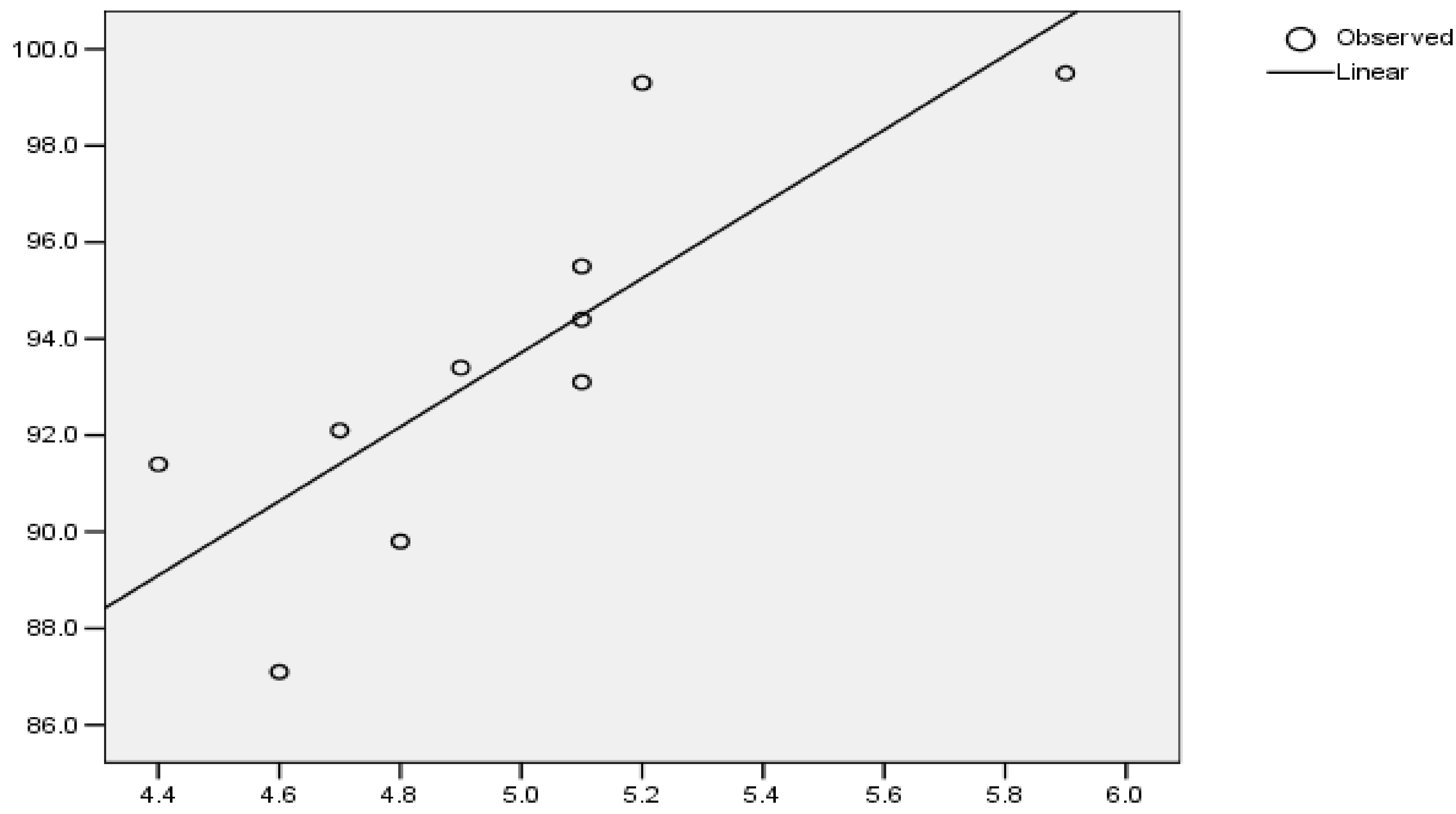
$$\hat{Y} = a + bX^2$$

$$\hat{Y} = a + bX + cX^2 + dX^3$$

$$\hat{Y} = a + bX^2 + cX^3$$

$$\hat{Y} = a + bX^3$$

Regresi Linier Sederhana



- X_i variabel independen ke-*i*
- Y_i variabel dependen ke-*i* maka bentuk model regresi sederhana adalah :

$$Y_i = \alpha + \beta X_i + \varepsilon_i, \quad i = 1,2,\ldots,n$$

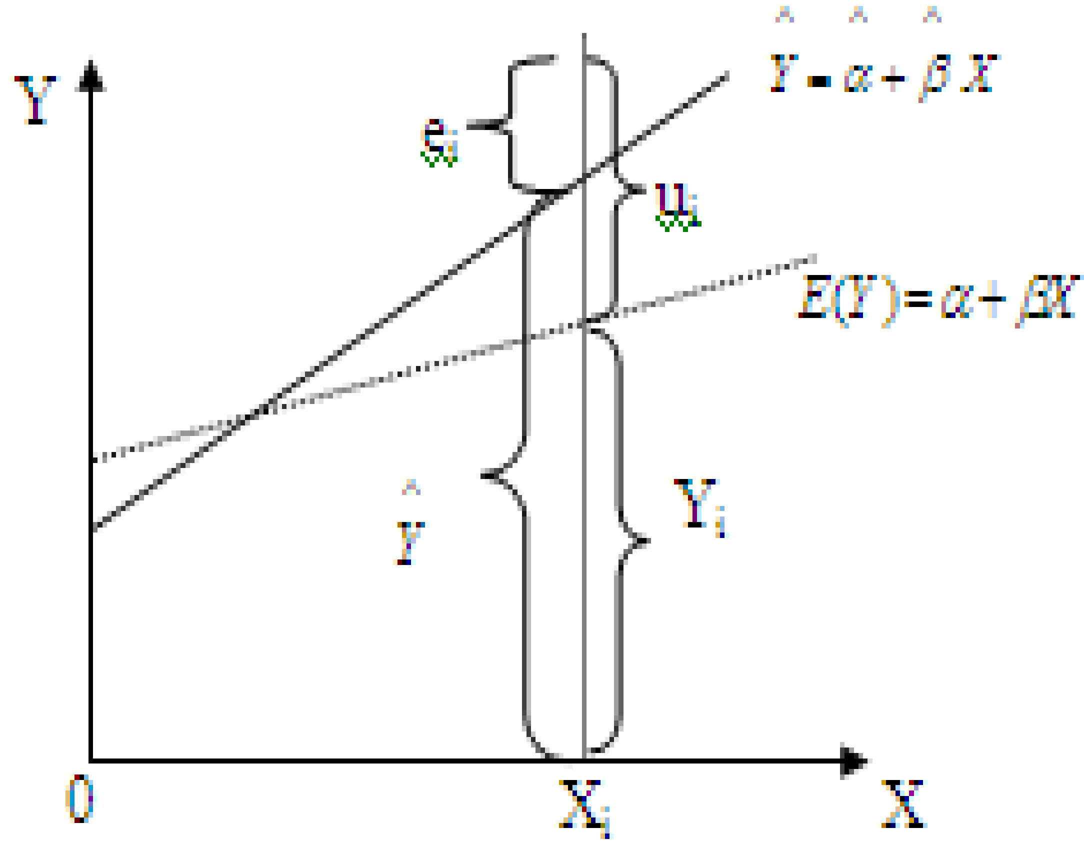
$$\hat{Y}_i = \hat{\alpha} + \hat{\beta} X$$

$$\hat{Y}_i = a + bX$$

dengan

$\hat{\alpha}, \hat{\beta}$ atau a, b ε_i parameter yang tidak diketahui sesatan random dgn asumsi $E[\varepsilon_i] = 0$

$$Var(\varepsilon_i) = \sigma^2$$



DARI GARIS REGRESI SAMPEL DIPEROLEH :

$$e_i = Y_i - (\hat{\alpha} + \hat{\beta} X_i)$$

DAN

$$D = \sum e_i^{\,2} = \sum_{i=1}^{n} (Y_i - (a + bX_i))^2$$

Turunkan D
terhadap
a dan b !!!!

Bentuk persamaan regresi linear sederhana

$$\hat{Y} = a + bX$$

Untuk meramalkan persamaan regresi maka nilai a dan b dirumuskan

$$b = \frac{\sum XY - n\,\overline{X}\,\overline{Y}}{\sum X^2 - n\,\overline{X}^2}$$

$$a = \overline{Y} - b\,\overline{X}$$

- **Pendugaan terhadap koefisien regresi:**

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

$$a = \frac{\left(\sum Y\right)\left(\sum X^2\right) - \left(\sum X\right)\left(\sum XY\right)}{n\sum X^2 - \left(\sum X\right)^2}$$

Metode Kuadrat Terkecil

Bagaimana Pengujian terhadap model regresi ??

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

Bagaimana menilai kesesuaian model ??

R^2 (Koef. Determinasi: % keragaman Y yang mampu dijelaskan oleh X)

ATAU

y	x	xy	x^2	y^2
. .	. .	. .	. .	. .
Σy	Σx	Σxy	Σx^2	Σy^2

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

$$a = \bar{y} - b\bar{x}$$

$$\bar{y} = \frac{\sum y}{n} \qquad \bar{x} = \frac{\sum x}{n}$$

Latihan

Mat (X)	Fis (Y)	XY	X2	Y2
60	80	4800	3600	6400
45	69	3105	2025	4761
50	71	3550	2500	5041
60	85	5100	3600	7225
50	80	4000	2500	6400
65	82	5330	4225	6724
60	89	5340	3600	7921
65	93	6045	4225	8649
50	76	3800	2500	5776
65	86	5590	4225	7396
45	71	3195	2025	5041
50	69	3450	2500	4761
665	951	53305	37525	76095

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

$$= \frac{53305 - \frac{(665)(951)}{12}}{37525 - \frac{(665)^2}{12}} = 0.8972$$

$$a = \bar{y} - b_1\bar{x} = 29.53$$

$$\text{jadi diperoleh persamaan regresi :}$$

$$\hat{Y}_i = 29.5294 + 0.8972 X_i$$