



POLITEKNIK STATISTIKA STIS

For Better Official Statistics

Pertemuan 1

KONSEP DASAR REGRESI

Efri Diah Utami, M.Stat

Mata Kuliah : Analisis Regresi

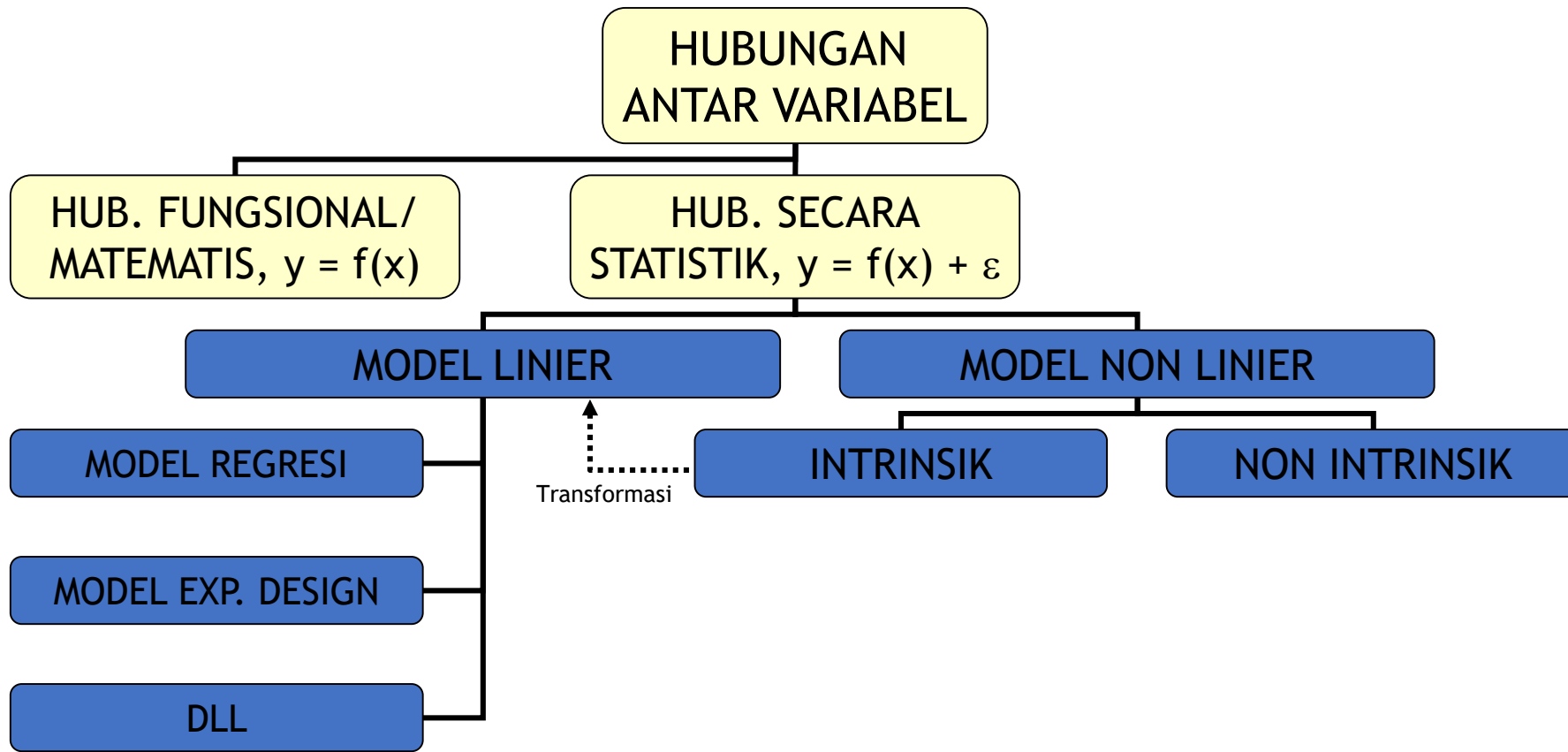
Program Studi D-IV Komputasi Statistik





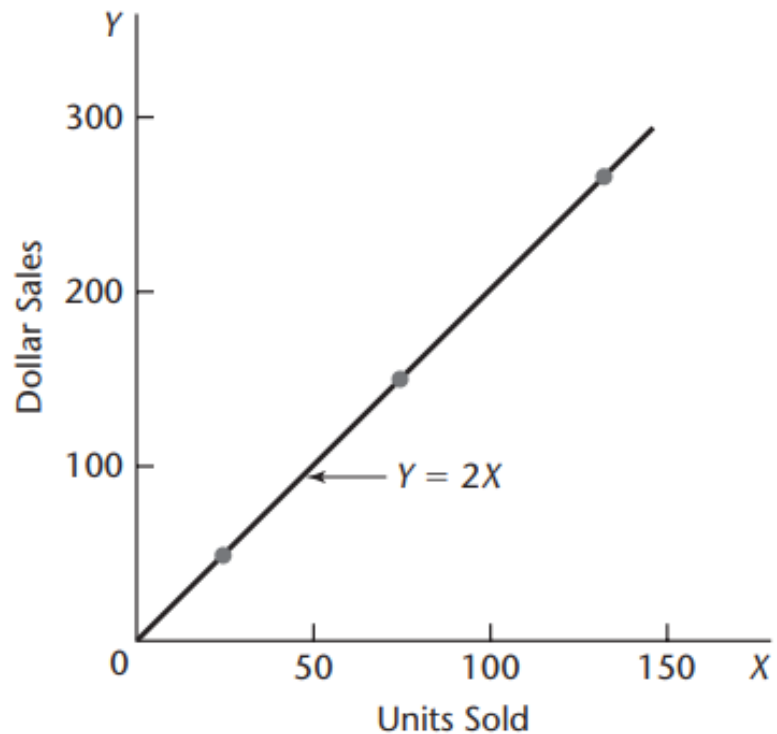
HUBUNGAN ANTAR VARIABEL





Contoh Hubungan Fungsional / Matematis :

Hubungan antara penjualan (dolar) dan produk yang terjual



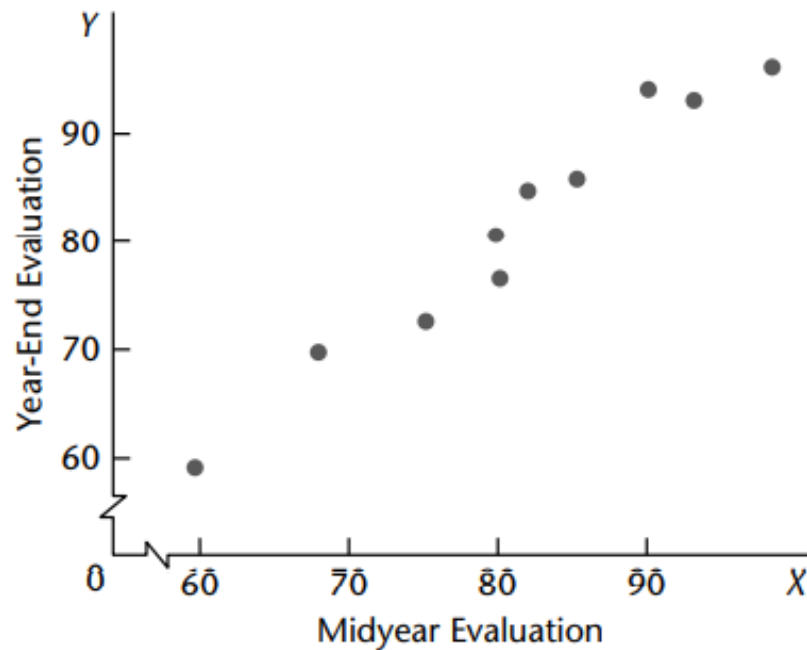
Period	Number of Units Sold	Dollar Sales
1	75	\$150
2	25	50
3	130	260

Contoh Hubungan Statistik :

Hubungan antara evaluasi performa tengah tahun dan akhir tahun

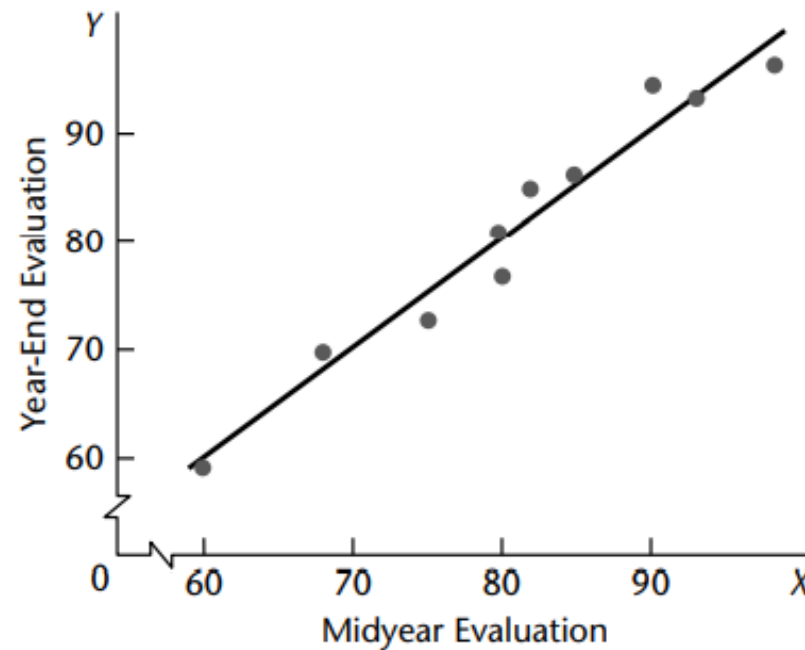
(a)

Scatter Plot



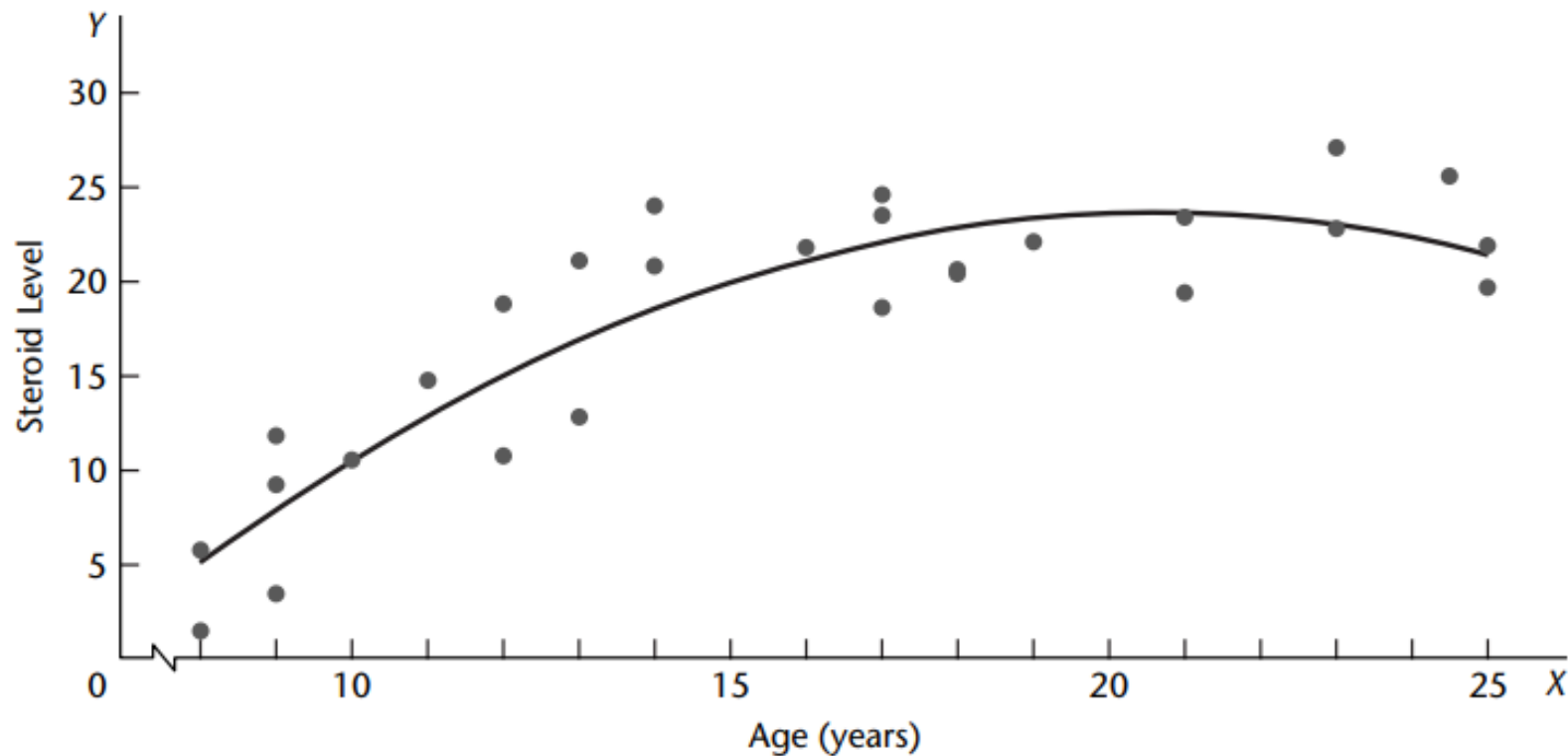
(b)

Scatter Plot and Line of Statistical Relationship



Contoh Hubungan Statistik :

Hubungan antara umur dan level steroid pada wanita sehat



- ❖ Keduanya mempelajari hubungan antar variabel

KORELASI

- Mempelajari keeratan hubungan antar 2 variabel (kuantitatif dan kualitatif) yang bisa dilihat dari besarnya angka, bukan tandanya
- Dapat mengetahui arah hubungan yang terjadi (berbanding lurus jika tandanya positif, dan berbanding terbalik jika tandanya negatif)
- Nilainya berkisar -1 sampai dengan 1
- Tidak bisa menyatakan hubungan sebab akibat

Korelasi yang tinggi tidak selalu berarti bahwa suatu variabel menyebabkan/memengaruhi variabel yang lain

KORELASI

Contoh:

- (1) Banyak kematian karena kekeringan di musim panas.
Banyak minuman ringan yang dikonsumsi di musim panas.

High positive correlation

Apakah minuman ringan menyebabkan kematian?

- (2) Banyaknya gigi yang rusak pada anak usia SD.
Jumlah kosa kata anak SD.

High positive correlation

Apakah kosa kata anak harus dibatasi agar giginya tidak rusak?



REGRESI

- Mempelajari bentuk hubungan antar variabel melalui suatu persamaan (RLS, RLB, Regresi non Linier). Hubungan bisa berupa hubungan sebab akibat.
- Dapat mengukur seberapa besar suatu variabel memengaruhi variabel lain
- Dapat digunakan untuk melakukan peramalan (prediksi) nilai suatu variabel berdasarkan variabel lain





REGRESI



- ***Dependent Variable***/Variabel Tak Bebas (Y): Variabel yang nilainya ditentukan oleh variabel lain. Diasumsikan bersifat *random/stochastic*
- ***Independent Variable***/Variabel Bebas (X): Variabel yang nilainya ditentukan secara bebas (variabel yang diduga mempengaruhi variabel tak bebas). Diasumsikan bersifat *fixed/non stochastic*.
- **Syarat :**
 - Y: Berjenis data kuantitatif
 - X: Berjenis data kuantitatif atau kualitatif/kategorik

JENIS DATA UNTUK Y

1. Data Observasi

diperoleh tanpa melakukan kontrol thd var. X

→ tdk kuat menyatakan *cause-effect relationships*

Misal :

- Sebuah perusahaan ingin mempelajari tentang hubungan antara umur karyawan (X) dan jumlah hari sakit karyawan tersebut selama setahun (Y)
- Umur tidak bisa di kontrol
- Jumlah hari sakit seorang karyawan tidak langsung disebabkan oleh umur, bisa ada penyebab lain.
- Harus dipertimbangkan variabel penjelas lain yang lebih menyebabkan secara langsung

JENIS DATA UNTUK Y

2. Data Eksperimen

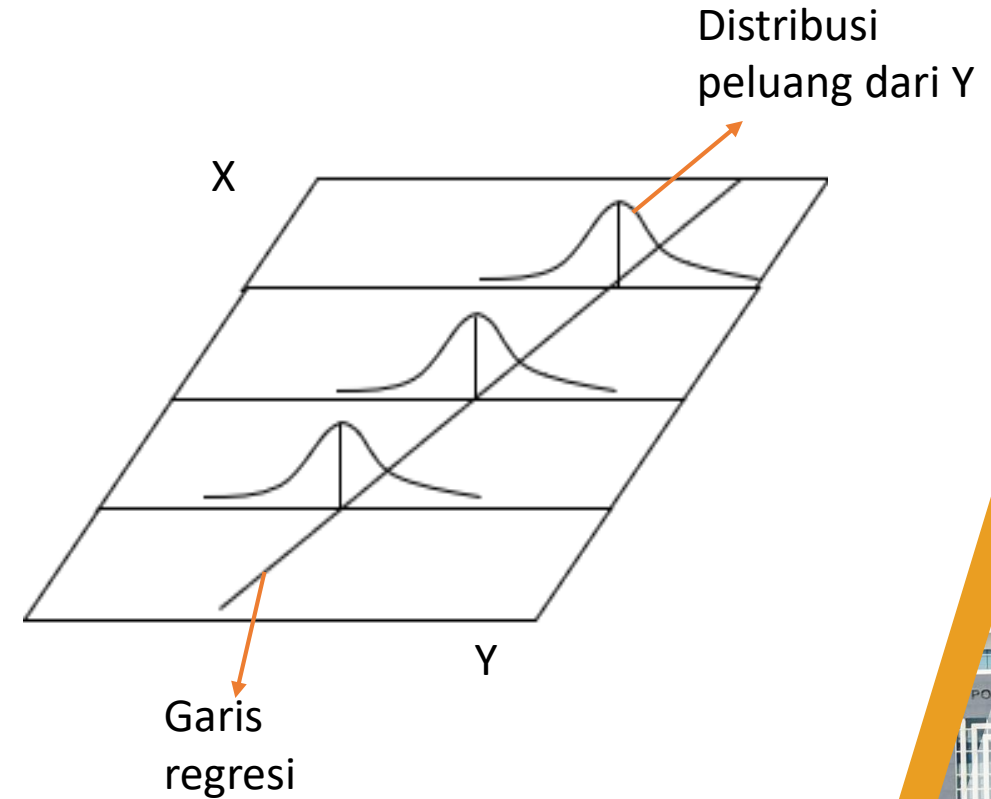
diperoleh dengan melakukan kontrol thd var. X

→ dapat menyatakan *cause-effect relationships*

Misal :

- Sebuah perusahaan asuransi ingin mempelajari tentang produktivitas analisnya (Y) dalam memproses claim dan lamanya pelatihan (X). Sampel acak 10 orang analis. Setiap 3 orang analis dilatih dengan durasi berbeda. Selama 10 minggu produktifitasnya diamati.
- Data eksperimen yang dihasilkan memberikan informasi lebih kuat tentang hubungan sebab-akibat daripada data observasi.
- Alasannya adalah bahwa pengacakan cenderung menyeimbangkan efek dari variabel lain yang mungkin mempengaruhi variabel respons.

- Pada suatu nilai X tertentu akan terdapat banyak kemungkinan nilai-nilai Y (Y akan terdistribusi mengikuti suatu fungsi peluang tertentu → Distribusi Normal dengan Nilai rata-rata $E(Y)$ dan Nilai varians σ^2 tertentu
- Nilai rata-rata $E(Y)$ diasumsikan berubah secara sistematis mengikuti perubahan nilai X , yg digambarkan dalam bentuk garis linier
- Nilai varians σ^2 pada setiap nilai X akan sama

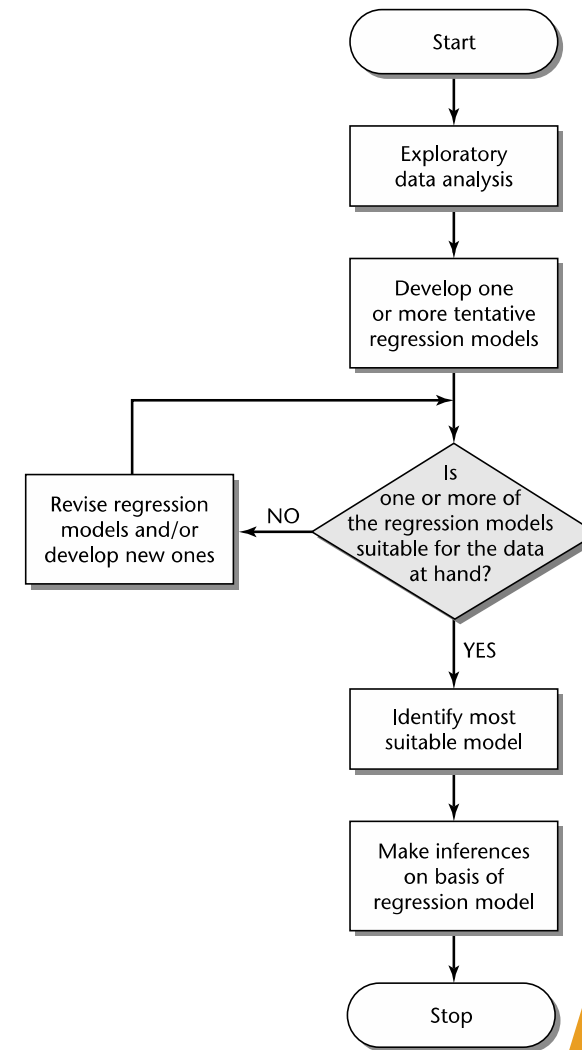


1. Pengaruh usia mobil terhadap harga jualnya (seberapa besar usia mobil mampu memprediksi harga jualnya)
2. Pengaruh usia seseorang terhadap kemampuan menghafal kata/istilah (seberapa besar usia mampu memprediksi tingkat kemampuan menghafal)
3. Pengaruh tinggi badan terhadap berat badan seseorang (seberapa besar tinggi badan seseorang mampu memprediksi berat badannya)
4. Pengaruh pendapatan rumah tangga terhadap pengeluaran untuk konsumsi (seberapa besar pendapatan rumah tangga mampu memprediksi pengeluaran untuk konsumsi)
5. Pengaruh motivasi belajar mahasiswa terhadap hasil belajar (seberapa besar motivasi belajar mahasiswa mampu memprediksi hasil belajarnya)
6. dan lain-lain



1. Identifikasi dan pembentukan model
 - * Eksplorasi data: sebaran data, missing value, pencilan (outlier), dsb.
 - * Identifikasi pola hubungan antar variabel
2. Pendugaan parameter model
 - * Estimasi titik
 - * Estimasi interval
3. Pengujian keberartian parameter model
 - * Uji Simultan (Simultan Test)
 - * Uji Parsial (Partial Test)
4. Penilaian ketepatan model (*goodness of fit*) dan pemeriksaan asumsi

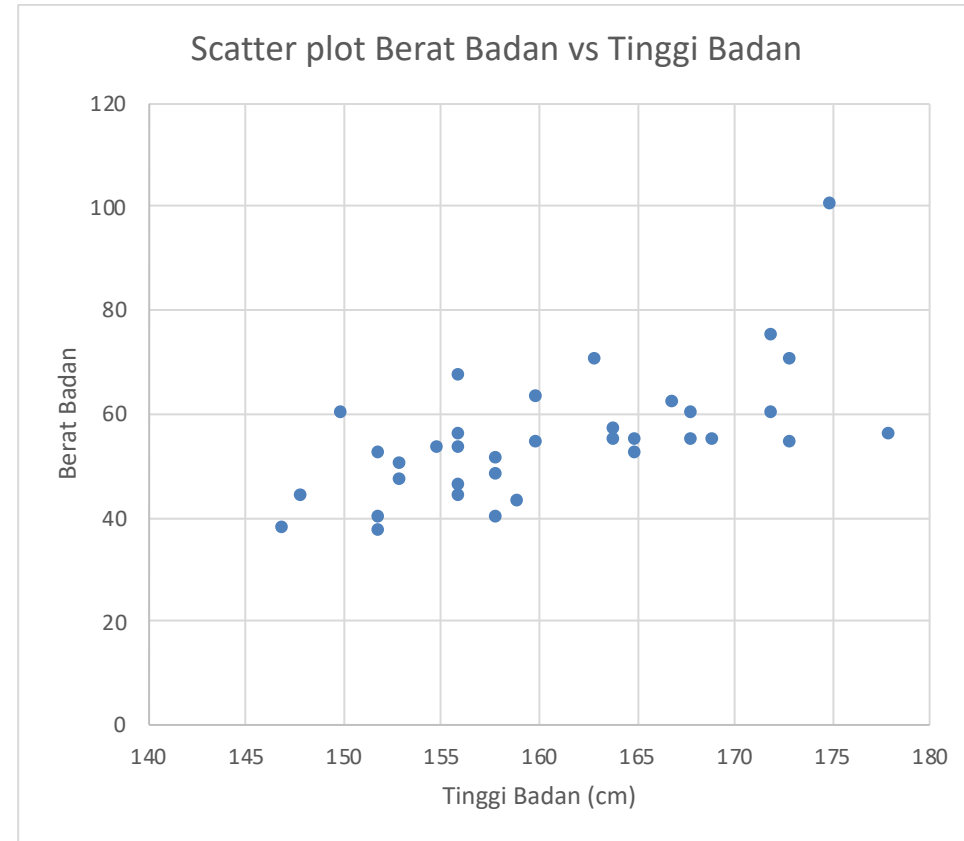
FIGURE 1.8
Typical
Strategy for
Regression
Analysis.



Scatter plot (diagram pencar)

- Berguna utk mengidentifikasi model hubungan antara variabel X dan Y.
- Bila pencaran titik-titik pada plot ini menunjukkan adanya suatu kecenderungan (trend) yang linier, maka model regresi linier layak digunakan.

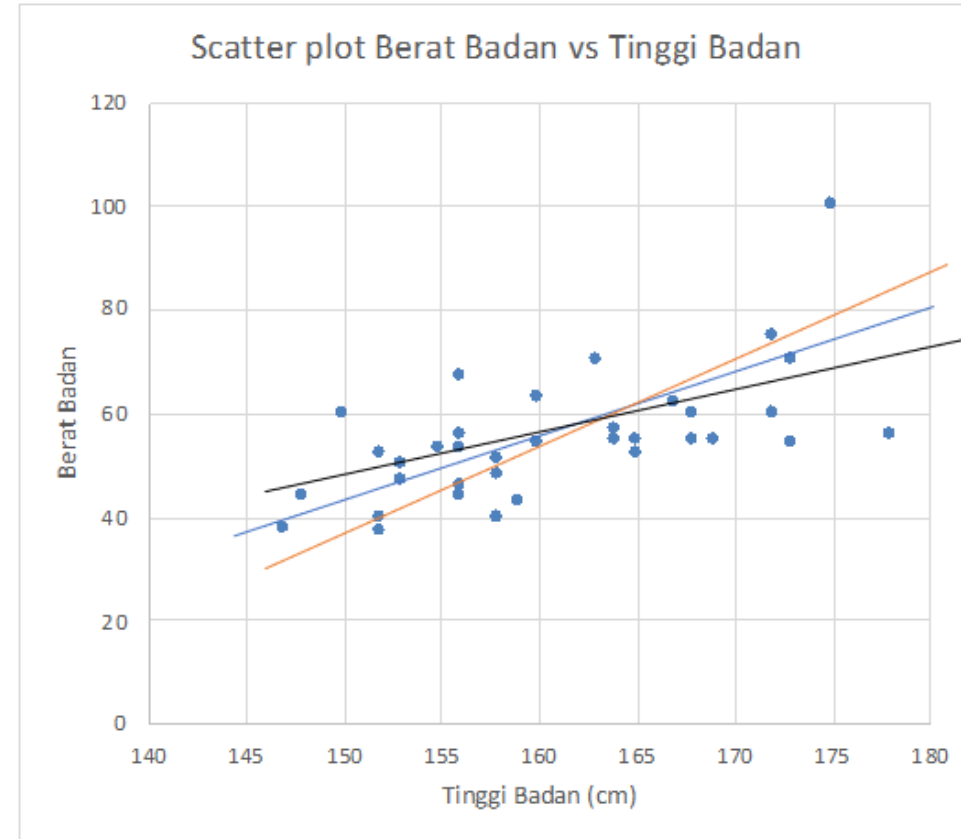
Contoh Plot Tinggi Badan (cm) vs Berat Badan (kg)



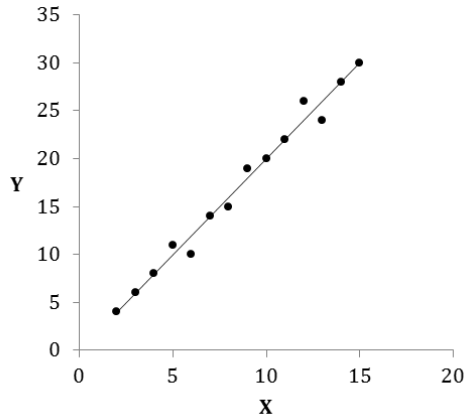
Relationship can be represented by line of best fit

Keterangan :

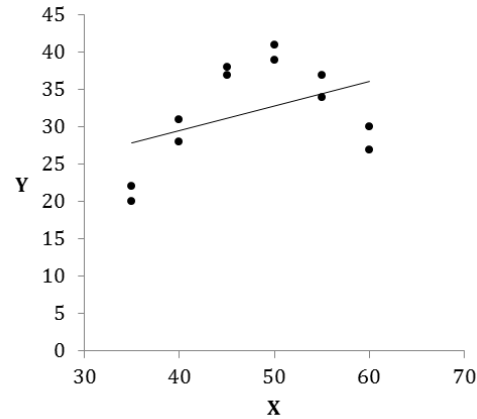
- Ternyata titik-titik (plotting data) tersebut terlihat mengelompok di sekitar garis lurus
- Pada scatter plot tersebut, sebenarnya bisa ditarik beberapa garis yang dekat terhadap titik-titik tersebut
- Tujuan kita di sini adalah
 1. Mencari garis yang paling tepat
 2. Melakukan Peramalan
 3. Ingin mengetahui hubungan yang terjadi (seberapa besar pengaruh tinggi badan terhadap berat badan)



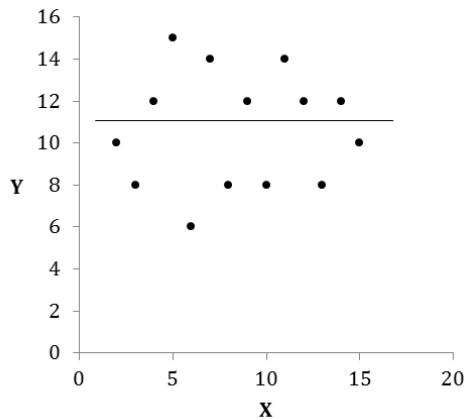
Relationship can be represented by line of best fit



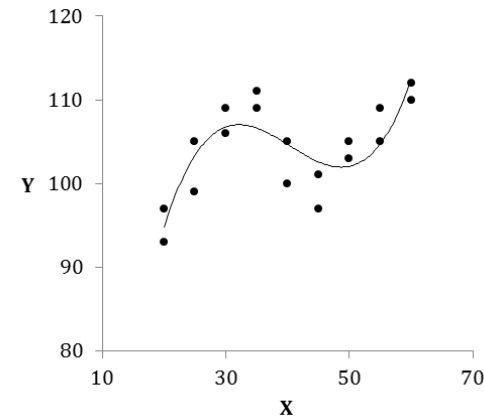
$$\hat{Y}_i = b_0 + b_1 X_i$$



$$\hat{Y}_i = b_0 + b_1 X_i + b_{11} X_i^2$$



$$\hat{Y}_i = b_0 = \bar{Y}$$



$$\hat{Y}_i = b_0 + b_1 X_i + b_{11} X_i^2 + b_{111} X_i^3$$

Beberapa Contoh Model Regresi

- Model Regresi Linier Sederhana (Simple Linear Regression Model):

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

jika masih ada error maka bukan sebuah estimasi

- Model Regresi Linier Berganda (Multiple Linear Regression Model) dg p-1 variabel bebas:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-1} X_{i(p-1)} + \varepsilon_i$$

- Model Regresi Polinomial Ordo-2 (Kuadratik) dg satu variabel bebas (Second-Order Polynomial Regression Model)

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \varepsilon_i$$

- dsb

Terima kasih