

**PENGOLAHAN DATA SECARA STATISTIKAL**

- Isu-isu yang terkait :
  1. preprocessing data : menyiapkan data
  2. appropriate technique(s) : teknik-teknik yang sesuai
  3. output analysis and interpretation : analisis output dan tafsiran
- Objektif Dasar :
 

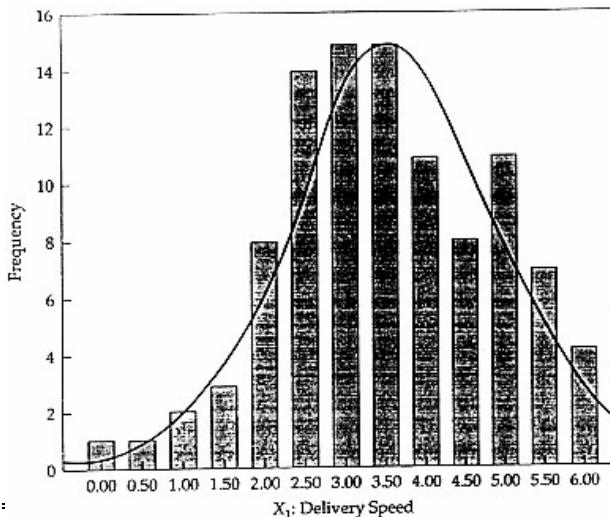
Menurut Sekaran (1992) terdapat 3 objektif dasar dalam pengolahan data :

  1. *getting the feel for the data* : merasakan data
  2. *testing the goodness of data* : menguji kebaikan data
  3. *testing the hypothesis or in general answering the research questions* : menguji hipotesis atau secara umum menjawab pertanyaan2 riset.

**PREPROCESSING**

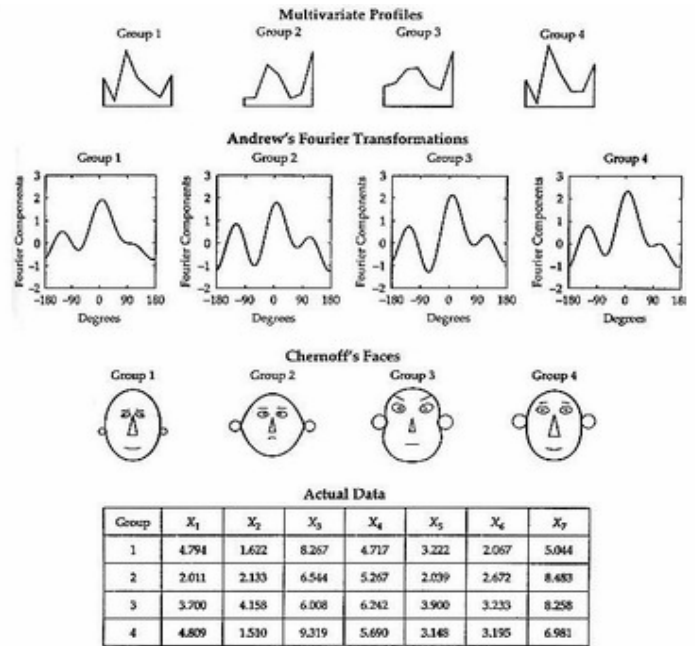
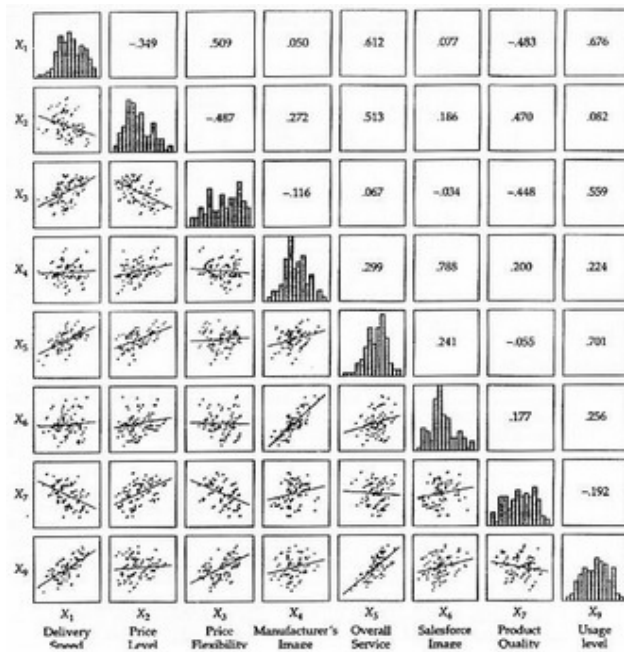
**Getting the feel for the data**

- ‘merasakan data’ → diagnostic awal perilaku data
- Dapat dilakukan dengan mengamati statistik deskriptif (mean, range, dll)
- Hal yang patut diwaspadai :
  1. adanya statistic yang ekstrim → salah input
  2. variabilitas respons yang sangat rendah → pertanyaan tidak tepat *atau* responden tidak mengerti apa yang ditanyakan
  3. respons yang cenderung sama → pertanyaan bias
  4. nilai korelasi antar variable yang tidak sesuai dengan logika/teori
- mengapa data perlu disaring ?
  1. Fenomena GIGO
  2. Validitas penggunaan analisis
  3. Isu generasilisasi hasil penelitian
  4. Investasi dalam analisis
- Salah satu cara untuk merasakan data adalah dengan pengamatan grafis :
  1. histogram
  2. scatterplot
  3. boxplot
  4. wajah-wajah Chernoff, dll
- Yang patut diingat : “merasakan data” ini hanya ditujukan sebagai diagnostic awal dan tidak untuk menggantikan teknik-teknik diagnostic statistik.



Frequency	Stem and Leaf
1.00	0 * 0
1.00	0 . 6
3.00	1 * 013
7.00	1 . 6688999
12.00	2 * 0013334444444
10.00	2 . 5566788899
18.00	3 * 0000011112334444444
10.00	3 . 5666777889
10.00	4 * 001122233
10.00	4 . 556778999
11.00	5 * 00112223344
5.00	5 . 55689
2.00	6 * 01

Stem width: 1.0  
 Each leaf: 1 case(s)  
 Valid cases: 100.0      Missing cases: .0      Percent missing: .0



**Missing Value**

- *Missing data process* : kejadian sistematis eksternal pada responden atau perilaku pada sebagian responden yang menyebabkan *missing value*.
- Apa yang harus dilakukan terhadap kuesioner/informasi dengan jawaban/data yang tidak lengkap ?
  1. abaikan kuesioner tersebut (*delete cases/object*)
  2. ganti nilai yang hilang dengan suatu nilai (opsi : nilai tengah dari skala, nilai rata-rata dari respons yang lain, dengan nilai yang didapat melalui teknik statistik tertentu)
  3. hapus variable yang mempunyai *missing value* paling banyak (*delete variable*)
- Sebab Missing Value :
  1. Eksternal : salah penginput (dari peneliti)
  2. Internal : dari si responden, karena malas untuk menjawab
- Contoh Kasus Missing Value :

Case ID	V <sub>1</sub>	V <sub>2</sub>	V <sub>3</sub>	V <sub>4</sub>	V <sub>5</sub>	Missing Data by Case	
						Number	Percent
1	1.3	9.9	6.7	3.0	2.6	0	0
2	4.1	5.7			2.9	2	40
3		9.9		3.0		3	60
4	.9	8.6		2.1	1.8	1	20
5	.4	8.3		1.2	1.7	1	20
6	1.5	6.7	4.8		2.5	1	20
7	.2	8.8	4.5	3.0	2.4	0	0
8	2.1	8.0	3.0	3.8	1.4	0	0
9	1.8	7.6		3.2	2.5	1	20
10	4.5	8.0		3.3	2.2	1	20
11	2.5	9.2		3.3	3.9	1	20
12	4.5	6.4	5.3	3.0	2.5	0	9
13					2.7	4	80
14	2.8	6.1	6.4		3.8	1	20
15	3.7			3.0		3	60
16	1.6	6.4	5.0		2.1	1	20
17	.5	9.2		3.3	2.8	1	20
18	2.8	5.2	5.0		2.7	1	20
19	2.2	6.7		2.6	2.9	1	20
20	1.8	9.0	5.0	2.2	3.0	0	0

MISSING DATA BY VARIABLE						TOTAL MISSING VALUES	
Number	2	2	11	6	2	Number: 23	
Percent	10	10	55	30	10	Percent: 23	

- Sifat Missing Value :
  1. Dapat diabaikan (*ignorable*)
    - ✓ karena konsekuensi sampling (observasi yg tidak disertakan saat sampling)
    - ✓ data yang disensor
  2. Tidak dapat diabaikan (*unignorable*)
    - ✓ faktor prosedural
    - ✓ respon tidak dapat dipakai (*inapplicable*)
    - ✓ penolakan respon pada pertanyaan-pertanyaan tertentu
- Pertanyaan yang berkenaan dengan *missing value* :
  1. Apakah *missing data* tersebar secara acak di seluruh observasi atau ada pola tertentu yang teridentifikasi?
  2. seberapa merata *missing data*?
- Memeriksa pola *missing data* :
  1. Missing not random (nonrandom)  
Misal terdapat 2 variabel, X & Y. Jika data yang hilang pada Y berkorelasi dengan nilai X, maka missing bersifat nonrandom
  2. Missing at Random (MAR)  
Missing pada Y tergantung pada X, tapi tidak pada Y
  3. Missing Completely at Random (MCAR)  
Tidak terdapat pola sama sekali

**Missing not Random**

X	Y
1	-
2	-
7	3
2	-
3	-
6	8
7	4
8	5

Pada X yang kecil Y cenderung hilang  
Y tergantung X, tetapi hilangnya Y  
ada polanya

**Missing at Random**

Gender (X)	Income (Y)
0	10
0	12
0	-
0	-
0	-
1	6
1	5
1	4

Hilangnya Y di tengah-tengah  
Y tergantung X, tetapi hilangnya Y  
bersifat random

- Diagnosa Keacakan
  1. Uji statistik perbedaan antar dua kelompok (berdasarkan variabel lain yang menjadi perhatian) → uji t
  2. Korelasi dikotomi → dibandingkan 2 variabel yang memiliki *missing value* kemudian dilihat apakah ada korelasi antara keduanya.
  3. Uji keacakan secara keseluruhan (contoh : Little's MCAR test)
- Pendekatan untuk menangani *Missing Value* : (hanya berlaku jika prosesnya MCAR)
  1. Menggunakan data yang lengkap saja  
Kelebihan : mudah, tinggal buang saja data yang tidak lengkap  
Kekurangan : data menjadi tidak cukup

2. Menghapus case(s) dan/atau Variable(s)  
Sebuah variabel dapat dihapus apabila datanya redundant (pengulangan) sehingga apabila salah satu dihapus ia sudah diwakilkan oleh data yang lain.
3. Metode Imputasi (suntikan data pada *missing value*)
  - Case substitution → menggunakan data dari case lain yang mirip perilakunya
  - Mean substitution → menggunakan rata-rata variabel
  - Cold deck imputation → menggunakan data sembarang (acak)
  - Regression imputation → menggunakan data dari hasil metode regresi
  - Multiple imputation → kombinasi dari metode di atas
4. Prosedur berdasarkan Model
  - Estimasi kemungkinan (likelihood) maksimum
  - Langsung melibatkan *missing data* pada analisis
  - Model E-M (*Estimation Missing*) Replacement  
Hanya berlaku untuk MCAR

**OUTLIER (Data Ekstrim)**

- Outlier dapat dikelompokkan ke dalam 4 kelas :
  1. kesalahan prosedur (*procedural error*)
  2. kejadian diluar kebiasaan dengan penjelasan
  3. kejadian diluar kebiasaan tanpa penjelasan
  4. tidak diluar kebiasaan tapi dengan kombinasi yang unik
- Cara mendeteksi outlier :
  1. teknik univariat  
normalisasi dan menggunakan batas  $2,5\sigma$  (ukuran sampel kecil) atau  $3\sigma$  atau  $4\sigma$
  2. teknik bivariat → plot dua variabel
  3. teknik multivariat  
menggunakan jarak Mahalanobis  $D^2$  (jarak di ruang multidimensional antara sebuah observasi dengan pusat dari semua pengamatan)

$$D^2 = D_1^2 = \sum_{j=1}^m (x_{1j} - \bar{x})^2$$

$$\bar{x} = \begin{pmatrix} \sum_{i=1}^n \frac{x_{i1}}{n} \\ \vdots \\ \sum_{i=1}^n \frac{x_{im}}{n} \end{pmatrix} \begin{matrix} \leftarrow \bar{x}_1 \\ \vdots \\ \leftarrow \bar{x}_m \end{matrix}$$

- Pendeskripsian dan Pembuatan Profil Outlier :
  1. sama seperti pada *missing value*, outlier jangan buru-buru dibuang karena akan mengurangi tingkat generalisasi penelitian yang dilakukan.
  2. beberapa analisis multivariat dapat digunakan untuk mendapatkan penjelasan secara statistik mengapa hal ini terjadi, misalnya analisis regresi linier sederhana dan analisis diskriminan.
- Kapan Outlier dibuang ?
  1. pembuangan outlier dapat meningkatkan kualitas analisis tapi di sisi lain dapat mengurangi *generalizability* penelitian
  2. beberapa teknik dapat mengakomodasikan outlier dalam perhitungan
  3. pembuangan outlier sebaiknya dilakukan jika data benar-benar bersifat *aberrant* (menyimpang)

**PENGUJIAN ASUMSI INPUT DATA**

Pengujian asumsi dapat dilakukan secara pada variabel atau variat, pemenuhan asumsi pada masing-masing variabel tidak menjamin asumsi variat-nya terpenuhi

**1. Normalitas**

Pengujian normalitas dapat dilakukan secara grafis dengan menggunakan :

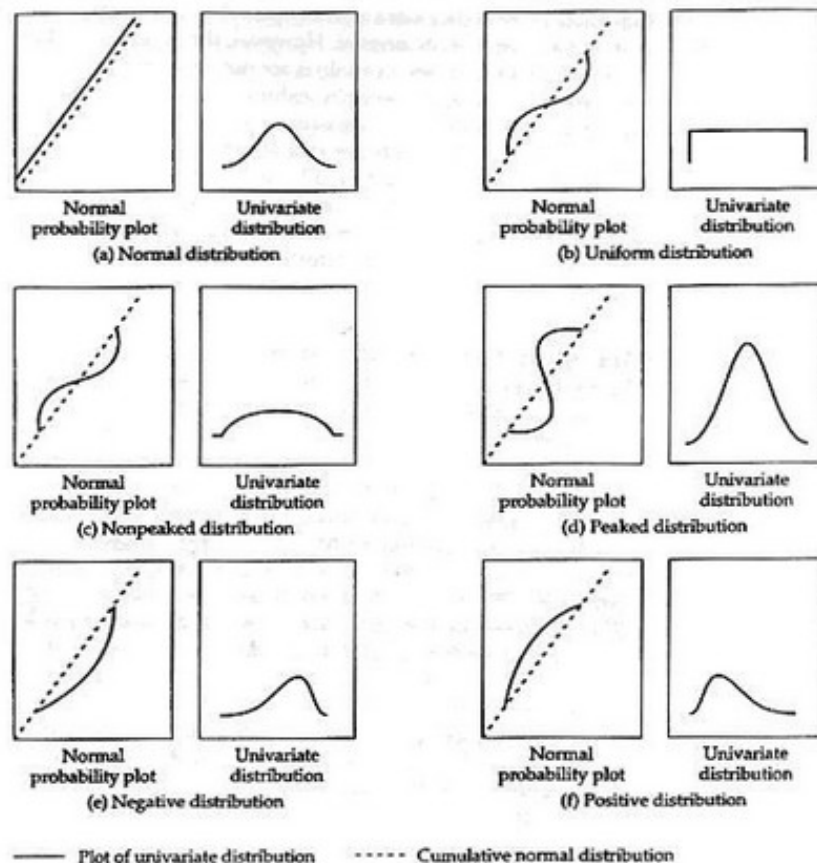
- normal probability plot
- pengujian statistikal (Kolmogorov-Smirnov, Saphiro-Wilks, dll)
- pengujian yang lebih sederhana dapat menggunakan pengujian skewness (kemencengan) dan kurtosis (keruncingan) berikut :

$$Z_{skewness} = \frac{skewness}{\sqrt{\frac{6}{N}}} = \frac{1}{\sqrt{\frac{6}{N}}} \frac{momen\ ke - 3}{S^3} = \sqrt{\frac{N}{6}} \frac{\sum (x - \bar{x})^3}{S^3 N^3}$$

$$Z_{kurtosis} = \frac{kurtosis}{\sqrt{\frac{24}{N}}} = \frac{1}{\sqrt{\frac{24}{N}}} \frac{momen\ ke - 4}{S^4} = \sqrt{\frac{N}{24}} \frac{\sum (x - \bar{x})^4}{S^4 N^4}$$

$$momen\ ke - k = \frac{\sum (x - \bar{x})^k}{N^k}$$

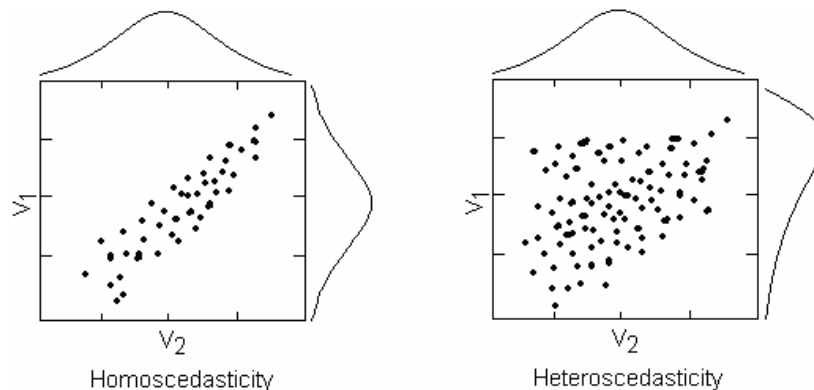
Z semakin mendekati nol maka semakin normal



- Remedial terhadap pelanggaran asumsi normalitas dapat dilakukan dengan melakukan transformasi variabel
- Namun seringkali perbaikan atas pelanggaran asumsi yang lain dapat memperbaiki normalitas data
- Beberapa variabel normal belum tentu membentuk Variat (kombinasi dari beberapa variabel) yang normal juga.
- Normal multivariat : kombinasi linier dari variabel yang ada berdistribusi normal.

## 2. Homoscedasticity

- Asumsi yang didasarkan pada hubungan dependensi antar variabel
- Asumsi ini menyatakan variabel dependen memiliki variansi yang sama sepanjang nilai variabel dependen
- Pengujian dapat dilakukan secara graphical atau dengan menggunakan uji statistik (Levene Test, Box M, dll).
- Remedial untuk pelanggaran asumsi ini dapat dilakukan dengan jalan transformasi variabel. Namun seringkali pada saat remedial terhadap nonnormalitas dilakukan, pelanggaran asumsi homoscedasticity turut terselesaikan



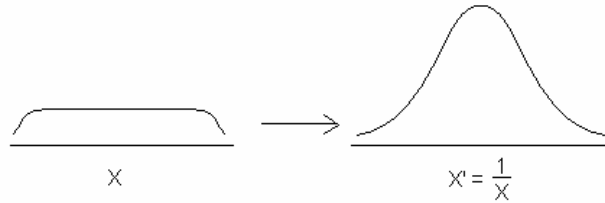
## 3. Linearitas

- Cara pendeteksian pelanggaran asumsi linearitas dapat dilakukan melalui pengamatan graphical atas scatter plot atau menggunakan teknik analisis regresi sederhana dan selanjutnya menggunakan analisis residual
- Perbaikan atas pelanggaran asumsi ini dapat dilakukan dengan melalui transformasi data

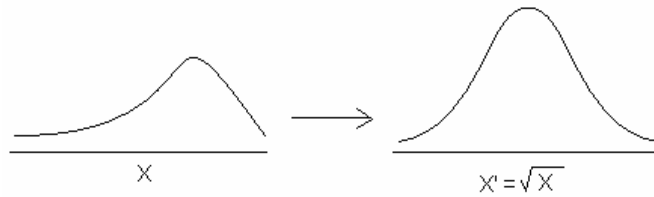
**TRANSFORMASI DATA**

**1. Pencapaian Normalitas**

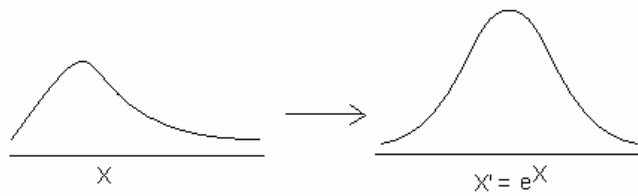
- pada *flat distribution*, transformasi yang dilakukan adalah dengan menggunakan invers ( $1/X$ )



- pada *negatively skewed distribution*, dapat digunakan transformasi akar kuadrat

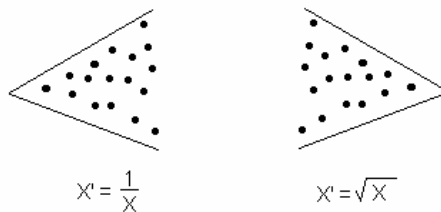


- pada *positively skewed distribution*, dapat digunakan transformasi logaritmik

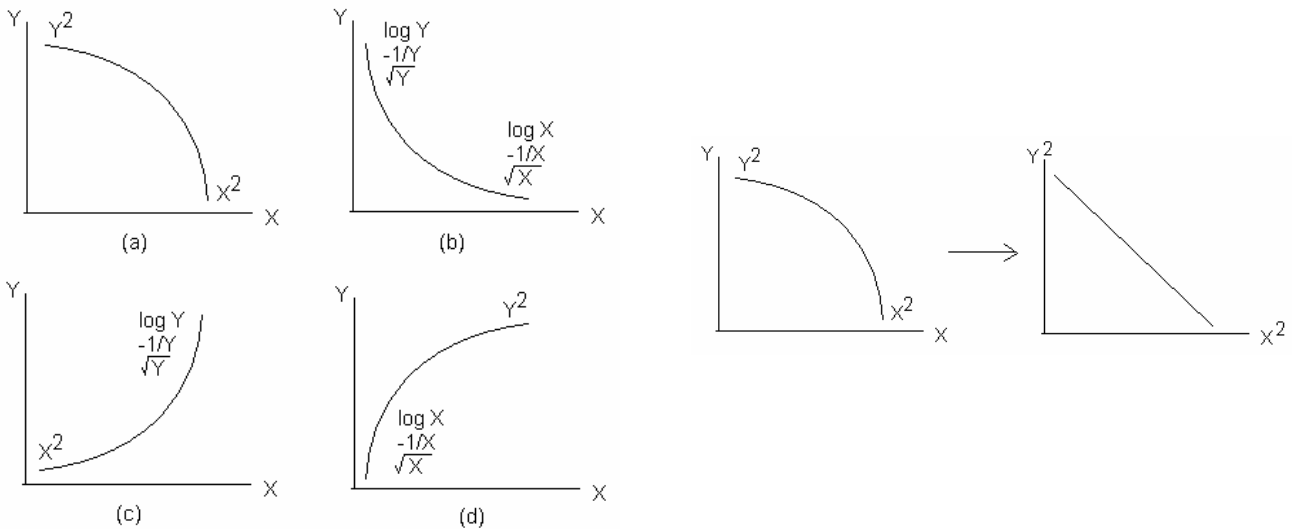


**2. Pencapaian Homoscedasticity**

- Jika kerucut membuka ke kanan gunakan invers
- Jika kerucut membuka ke kiri gunakan akar kuadrat



**3. Pencapaian Linearitas**



**General Guideline**

- Untuk mendapat efek transformasi, rasio mean dan deviasi standar dari sebuah variabel harus kurang dari 4.0

$$\frac{\bar{X}}{S} < 4$$

- Jika transformasi dapat dilakukan pada salah satu variabel, pilih variabel dengan rasio terkecil
- Transformasi dilakukan pada variabel independent kecuali pada kasus heteroscedasticity
- Remedial atas heteroscedasticity dapat dilakukan pada variabel dependent
- Transformasi dapat mengubah interpretasi atas variabel

**Uji Kebaikan Data (Testing the goodness of data)**

- Di sini dilakukan pengujian atas validitas data dan reliabilitas alat ukur
- Validitas dapat diuji lewat analisis faktor (*confirmatory factor analysis*) dan teknik-teknik pengujian validitas data yang lain
- Reliabilitas dapat diukur dengan menggunakan *alpha-Cronbach*

**Menjawab Pertanyaan-pertanyaan Riset**

- Langkah ini dilakukan hanya setelah data 'bersih' dan ukuran kebaikan data sudah terpenuhi
- Pemilihan *tools* statistik yang tepat merupakan syarat yang harus terpenuhi
- Pemilihan *tools* statistik yang tepat terkait erat dengan tujuan studi dan data yang dimiliki

**APPROPRIATE TECHNIQUE(S)**

**Pertimbangan Dalam Memilih**

**1. Jenis data**

Jenis data di sini terkait dengan skala alat ukur yang digunakan :

- o Variabel nonmetrik : nominal/kategoris, ordinal
- o Variabel Metrik : interval, rasio

Masing-masing jenis variabel/skala, sesuai untuk diolah dengan teknik statistika khusus

Contoh : uji-uji parametric tidak dapat digunakan untuk variabel berjenis nonmetrik

**2. Desain penelitian**

Terdapat 3 faktor dasar pada sebuah rancangan penelitian :

1. independensi sample
2. jumlah kelompok objek/data
3. jumlah variabel penelitian yang terlibat

Rancangan yang berbeda akan menuntut penggunaan teknik analisis statistika yang berbeda pula, contoh :

1. untuk menguji pengaruh suatu metode diet digunakan uji-t
2. untuk menguji apakah rata-rata produktivitas dari penggunaan tiga jenis pupuk adalah berbeda digunakan ANOVA
3. untuk melihat pengaruh jenis media, besar budget, lama ekspose, dan bintang yang digunakan dalam sebuah iklan terhadap penjualan produk harus digunakan analisis multivariat.

**3. Asumsi yang mendasari sebuah uji statistik**

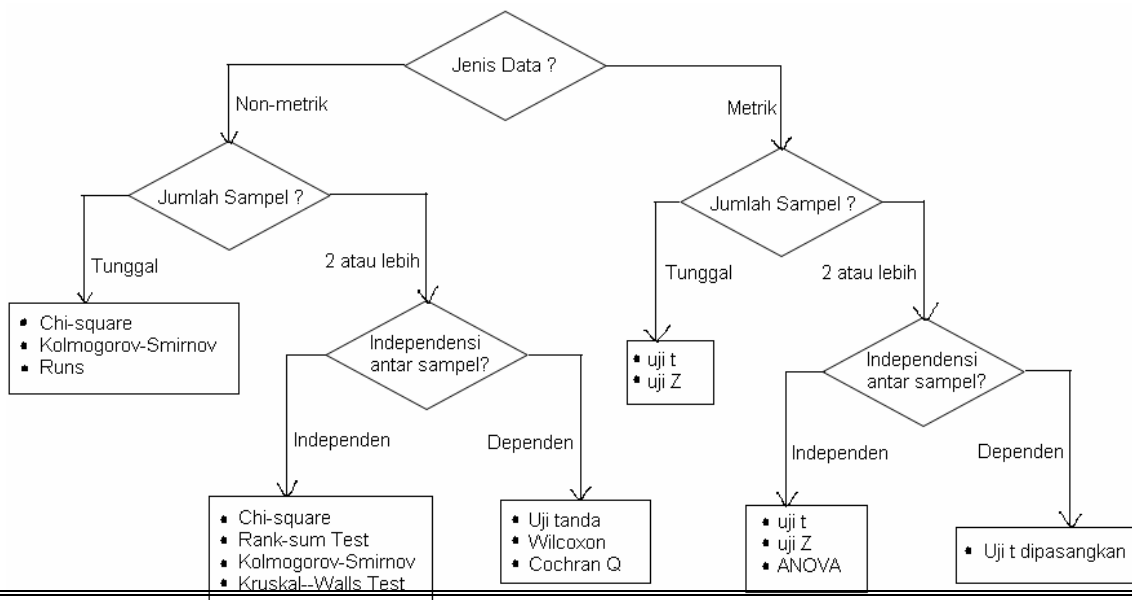
Setiap teknik analisis statistik dibangun dengan dasar asumsi-asumsi tertentu

Contoh :

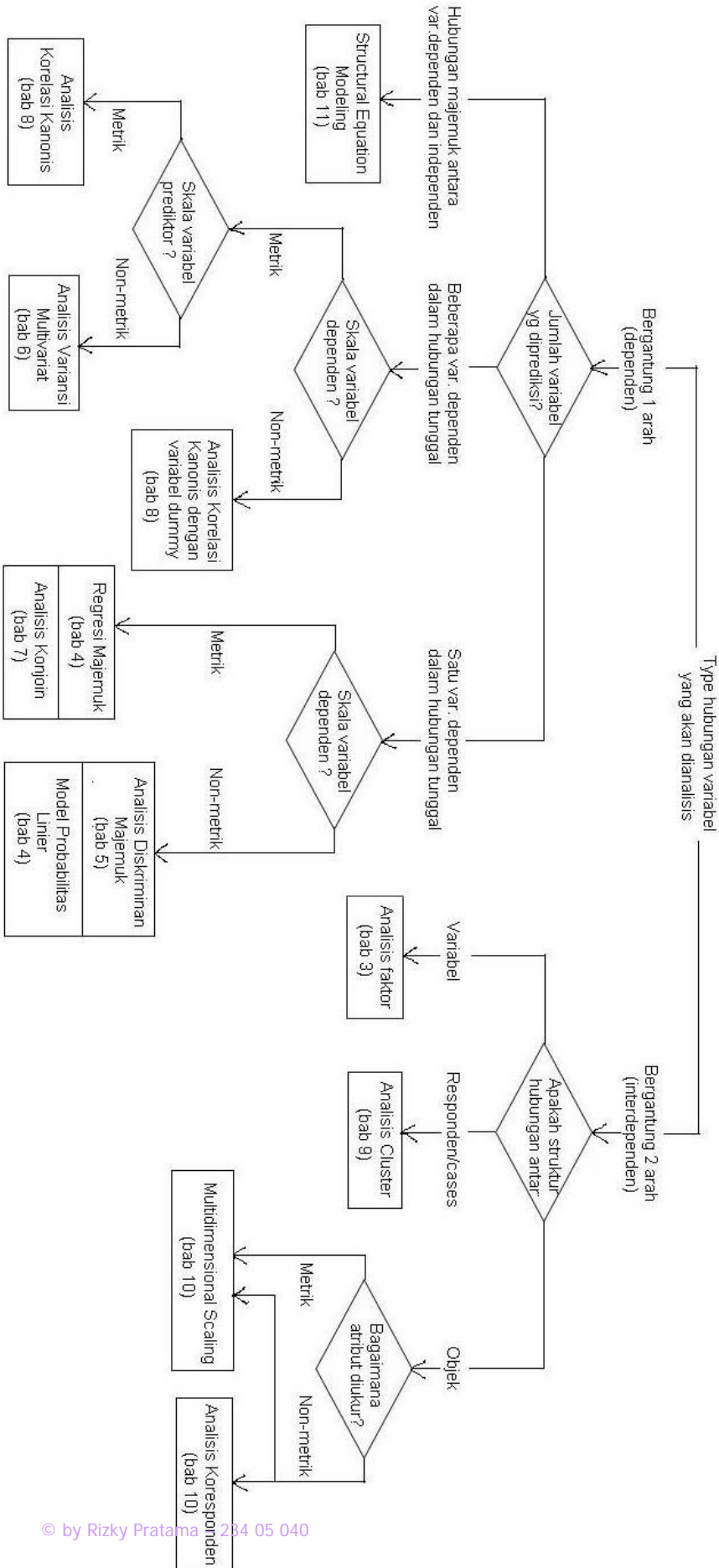
1. uji-t mengasumsikan sampel diambil dari populasi normal
2. analisis regresi mengasumsikan variansi nilai y adalah sama untuk semua nilai x yang diberikan

Pelanggaran terhadap asumsi tersebut dapat menyebabkan hasil analisis tidak valid

**PEDOMAN PEMILIHAN TEKNIK ANALISIS STATISTIKA YANG TEPAT (KASUS DATA UNIVARIAT)**



## PEDOMAN PEMILIHAN TEKNIK ANALISIS STATISTIKA (KASUS DATA MULTIVARIAT)



**Teknik Multivariat Dependence  
Hubungan Matematis Antar Variabel**

- Regresi Linier Majemuk

$$Y_1 = X_1 + X_2 + \dots + X_n$$

(metrik)                      (metrik)

- Analisis Konjoin

$$Y_1 = X_1 + X_2 + \dots + X_n$$

(metrik, non-metrik)                      (non-metrik)

- Analisis Diskriminan

$$Y_1 = X_1 + X_2 + \dots + X_n$$

(non-metrik)                      (metrik)

- Structural Equation Modeling

$$Y_1 = X_{11} + X_{12} + \dots + X_{1n}$$

$$Y_2 = X_{21} + X_{22} + \dots + X_{2n}$$

$$Y_m = X_{m1} + X_{m2} + \dots + X_{mn}$$

(metrik)                      (metrik, non-metrik)

- Korelasi Kanonis

$$Y_1 + Y_2 + \dots + Y_n = X_1 + X_2 + \dots + X_n$$

(metrik, non-metrik)                      (metrik, non-metrik)

**Definisi Analisis Multivariat**

Adalah penerapan metode-metode yang berhubungan dengan pengukuran-pengukuran (misalnya variabel) dalam jumlah sangat banyak yang dilakukan pada tiap obyek dalam satu atau lebih sampel secara *simultan*.

Jadi, analisis multivariat membahas hubungan simultan antar beberapa (lebih dari dua) variabel.

**Jenis-jenis Skala Pengukuran**

Skala	Pemakaian Empiris	Statistik yang dapat dipakai
Nominal	Penentuan kesetaraan	Jumlah cacah (cases), Modus, Korelasi kontijensi (Chi-Square, Fisher's exact test)
Ordinal	Penentuan lebih atau kurang	Median, Persentil, rank-order correlation, sign test, run test
Interval	Penentuan kesamaan atau perbedaan interval	Rataan aritmetis, deviasi standar, deviasi rata-rata, korelasi, uji t, uji F
Rasio	Penentuan kesamaan atau perbedaan rasio	Koefisien variasi, Rataan Geometris, Rataan Harmonis

**Larik Data (Data Arrays)**

Sering kali data yang tersedia adalah hasil pengukuran n obyek atas  $p > 1$  variabel. Di sini kita gunakan notasi  $X_{ij}$ , yang menunjukkan hasil pengukuran atau nilai variabel ke-j atas case ke-i.

	Var 1	Var 2	...	Var j	...	Var p
Case 1	$x_{11}$	$x_{12}$	...	$x_{1j}$	...	$x_{1p}$
Case 2	$x_{21}$	$x_{22}$	...	$x_{2j}$	...	$x_{2p}$
...	...	...	...	...	...	...
Case i	$x_{i1}$	$x_{i2}$	...	$x_{ij}$	...	$x_{ip}$
...	...	...	...	...	...	...
Case n	$x_{n1}$	$x_{n2}$	...	$x_{nj}$	...	$x_{np}$

Presentasi data juga dapat dilakukan dalam bentuk matriks sebagai berikut :

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1j} & \vdots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2j} & \vdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{i1} & x_{i2} & \dots & x_{ij} & \dots & x_{ip} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nj} & \dots & x_{np} \end{pmatrix}$$

**KONSEP VEKTOR DAN STATISTIKA DESKRIPTIF**

**Data Awal**

Data berikut merupakan hasil pengukuran konsumen atas variabel-variabel *attitude*, opini, dan niat untuk membeli terhadap suatu produk.

Attitude score (X1)	Opinion score (X2)	Intention score (Y)
2	3	3
2	2	1
4	6	5
3	5	6
6	8	8
8	8	7
5	4	5
2	7	8
7	6	4
6	5	3
2	4	4
5	7	4
4	6	7
8	9	8
3	5	4

**Deviasi Standar**

- Deviasi standar memiliki hubungan dengan panjang vektor. Vektor yang dimaksud di sini adalah vektor nilai yang telah dikoreksi mean (*mean corrected score*)
- Simbolkan *mean corrected score* dari obyek tertentu sebagai dengan  $x_i$  atau  $x_i = X_i - \bar{X}$
- Maka deviasi standar dapat dihitung menggunakan rumus berikut :

$$\sigma_x = \frac{\|\mathbf{x}'\|}{\sqrt{n}} = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n}} \tag{1}$$

**Korelasi dan Kovariansi**

- Statistik deskriptif korelasi dan korelasi berhubungan dengan perkalian dalam (*inner product*) dari dua vektor
- Diberikan  $\mathbf{x}$  dan  $\mathbf{y}$  masing-masing adalah vektor yang elemen-elemennya berupa *mean corrected score*.
- *Inner product* dari  $\mathbf{x}$  dan  $\mathbf{y}$  didefinisikan sebagai :

$$\begin{aligned} \mathbf{x}'\mathbf{y} &= x_1y_1 + x_2y_2 + \dots + x_ny_n \\ &= \sum_{i=1}^n x_iy_i \end{aligned} \tag{2}$$

- Kovariansi antara  $X_i$  dan  $Y_i$  diberikan oleh :

$$C_{XY} = \frac{1}{n} \sum_{i=1}^n x_iy_i \quad \text{atau} \quad C_{XY} = \frac{\mathbf{x}'\mathbf{y}}{n} = \frac{1}{n} \mathbf{x}'\mathbf{y} \tag{3}$$

Jadi, kovariansi antar dua variabel adalah sama dengan *inner product* dari *mean corrected score* kedua variabel dibagi dengan ukuran sampelnya.

- Jika kita standardisasikan  $\mathbf{x}$  dan  $\mathbf{y}$  dengan jalan membagi tiap elemennya masing-masing dengan  $\sigma_x$  dan  $\sigma_y$  sehingga didapatkan  $\mathbf{x}^*$  dan  $\mathbf{y}^*$  maka dengan rumus yang analog dengan rumus kovariansi di atas, kita akan dapatkan nilai korelasi antara  $\mathbf{x}$  dan  $\mathbf{y}$

$$r_{XY} = \frac{\mathbf{x}^*\mathbf{y}^*}{n} = \frac{1}{n} \mathbf{x}^*\mathbf{y}^* \tag{4}$$

- Dengan memasukkan persamaan (1) ke persamaan (4) kita akan dapatkan :

$$r_{XY} = \frac{\mathbf{x}^*\mathbf{y}^*}{n} = \frac{1}{n} \frac{\mathbf{x}'\mathbf{y}}{\sigma_x \sigma_y} = \frac{\mathbf{x}'\mathbf{y}}{(\sqrt{n}\sigma_x)(\sqrt{n}\sigma_y)} = \frac{\mathbf{x}'\mathbf{y}}{\|\mathbf{x}'\|\|\mathbf{y}'\|}$$

Perhatikan bahwa bentuk terakhir ini tidak lain adalah cosinus sudut yang dibentuk oleh vektor  $\mathbf{x}$  dan  $\mathbf{y}$ . Jadi

$$r_{XY} = \cos \theta_{XY} \tag{5}$$

**Tendensi Sentral (Centroid)**

- $\mathbf{X}$  menggambarkan sebuah matriks  $n \times p$ , dimana  $n$  mewakili jumlah observasi dan  $p$  mewakili jumlah variabel
- Vektor baris dari mean  $\mathbf{X}$  disimbolkan dengan  $\bar{\mathbf{x}}$  disebut dengan centroid
- Centroid ini dapat dihitung dengan menggunakan rumus berikut :

$$\bar{\mathbf{x}}' = \frac{1}{n} \mathbf{1}'\mathbf{X} \tag{6}$$

**Mean Corrected Data**

- *Mean corrected data* adalah data hasil pengurangan mean atas masing-masing data awal
- Rumus perhitungan *mean corrected data* dalam notasi matriks adalah sebagai berikut :

$$\mathbf{X}_d = \mathbf{X} - \mathbf{1}\bar{\mathbf{x}}' \quad (7)$$

**Variansi**

Dengan menggunakan informasi matriks *mean corrected data* kita dapat dengan mudah menghitung variansi dengan menggunakan rumus berikut :

$$s^2 = \frac{1}{n-1} \mathbf{X}_d' \mathbf{X}_d \quad (8)$$

**Standardized Data**

- Diberikan **D** adalah matriks diagonal dengan nilai diagonalnya adalah variansi dari masing-masing variabel
- Data yang distandardisasikan dapat diperoleh melalui rumus :

$$\mathbf{X}_s = \frac{\mathbf{X}_d}{\sqrt{\mathbf{D}}} \quad (9)$$

**Sum dan Sum of Square**

- Untuk menghitung *sum* dari suatu variabel, misalkan  $X_1$ , kita dapat menggunakan rumus :

$$\sum X_1 = \mathbf{1}' \mathbf{X}_1 \quad (10)$$

- Untuk menghitung *sum of square* dari suatu variabel, misalkan  $X_1$ , kita dapat menggunakan rumus :

$$\sum X_1^2 = \mathbf{X}_1' \mathbf{X}_1 \quad (11)$$

**Cross Product**

- *Cross product* dari dua variabel, misalkan  $X_1$  dan  $Y_1$  dapat dihitung dengan menggunakan rumus :

$$\sum X_1 Y_1 = \mathbf{X}_1' \mathbf{Y}_1 \quad (12)$$

- Secara umum kita dapat merangkum *sum of square* dan *cross product* dari beberapa variabel dalam matriks **B** berikut :

$$\mathbf{B} = \mathbf{X}' \mathbf{X} \quad (13)$$

- Dalam matriks **B** ini, diagonal menunjukkan *sum of square* dan *entry* lainnya menunjukkan *cross product* antara variabel pada kolom dengan variabel pada baris

**Mean Corrected Sum of Square**

- Matriks *mean corrected sum of square* dan *cross product* (disimbolkan dengan **S**) dapat dihitung dengan menggunakan rumus :

$$\mathbf{S} = \mathbf{X}' \mathbf{X} = \frac{1}{n} (\mathbf{X}' \mathbf{1})(\mathbf{1}' \mathbf{X}) \quad (14)$$

- Jika menggunakan matriks *mean corrected data*, kita dapat memakai rumus :

$$\mathbf{S} = \mathbf{X}_d' \mathbf{X}_d \quad (15)$$

**Matriks Variansi – Kovariansi**

- Matriks variansi-kovariansi dapat diperoleh dengan menggunakan rumus :

$$\mathbf{C} = \frac{1}{n} \mathbf{S} \quad (16)$$

- Sedangkan matriks korelasi dapat diperoleh dengan menggunakan rumus :

$$\mathbf{C} = \frac{1}{n-1} \left( \frac{1}{\sqrt{\mathbf{D}}} \mathbf{S} \frac{1}{\sqrt{\mathbf{D}}} \right) \quad (17)$$

## Gambaran Influential Observation

### INTERPRETASI

#### Regresi vs Koefisien Beta

- Koefisien regresi sensitif pada satuan pengukuran (misal : cm vs km)
- Koefisien beta dihitung berdasarkan nilai terstandarisasi → memberikan keuntungan saat kita berhadapan dengan beberapa satuan pengukuran yang berbeda pada variabel independen

#### Efek Multikolinearitas

- Proses memisahkan pengaruh tiap variabel independen menjadi sulit
- Membatasi besar koefisien determinasi
- Singularitas → dapat menyebabkan koefisien regresi salah terestimasi

#### Cara Pendeteksian Kolinearitas

- Matriks korelasi Pearson
- Nilai toleransi → jumlah variabilitas dari variabel independen yang dipilih tidak dijelaskan oleh variabel independen yang lain.  
Threshold : 0,1
- Nilai Variance Inflation Factor (VIF) atau faktor kenaikan variansi → derajat kemampuan tiap variabel independen berubah menjadi variabel dependen dan kemudian diregresi terhadap variabel independen sisanya.  
Threshold : 10

#### Remedial bagi kolinearitas

- Hilangkan salah satu atau lebih variabel independen berkorelasi tinggi
- Gunakan model dengan variabel independen berkorelasi tinggi untuk prediksi saja
- Gunakan korelasi sederhana antara variabel independen dan dependen untuk memahami hubungan diantara keduanya
- Gunakan metode yang lebih canggih

### VALIDASI HASIL

- Validasi dapat dilakukan dengan menggunakan sampel baru (*additional sampling*) atau dengan menggunakan data yang ada (metode split atau *bootstrapping*).
- Khusus dengan menggunakan data yang ada dapat digunakan statistik PRESS
- Untuk membandingkan beberapa model gunakan *adjusted R<sup>2</sup>*.

#### Prediksi Menggunakan Model

Beberapa hal yang harus diperhatikan :

- sampel bervariasi
- kondisi dan hubungan yang mungkin berubah seiring dengan berubahnya waktu
- tidak menggunakan prediksi di luar range variabel independen

**MENGIDENTIFIKASI INFLUENTIAL**

**Outliers (menggunakan residual)**

- outlier tidak harus menjadi *influential observation* dan sebaliknya
- kita dapat menemukan outlier dengan menggunakan residual :
  1. (normal) residual :  $Y_i - \hat{Y}_i$
  2. Standardized Residual:  $\frac{e}{\sigma_e}$
  3. Residual yang dihapus :  $Y_i - \hat{Y}_i^{(i)}$   
Dimana  $\hat{Y}_i^{(i)}$  artinya nilai Y yang terprediksi saat case ke-i dikeluarkan
  4. Studentized residual : sama seperti standardized residual tetapi penyebutnya adalah standar deviasi residual saat case ke-i dikeluarkan  $\rightarrow \frac{e}{\sigma_e^{(i)}}$
  5. Studentized deleted residual : sama seperti studentized residual tetapi pembilangnya adalah residual yang dihilangkan  $\rightarrow \frac{e^{(deleted)}}{\sigma_e^{(i)}}$

**Outliers (menggunakan plot regresi parsial)**

Plot regresi parsial dibuat dengan cara memplot nilai-nilai setiap variabel independen vs dependen

**Leverage (*Hat Matrix* / Matriks Topi)**

$$\hat{Y} = HY$$

- Di sini matriks H disebut dengan *hat matrix* dan diagonalnya disebut dengan *hat values*
- *Hat value* membawa dua informasi :
  - ukuran jarak dari observasi tersebut terhadap pusat (*mean center*) dari seluruh observasi atas seluruh variabel independen yang ada (mirip dengan jarak Mahalanobis)
  - *hat value* yang besar menunjukkan bobot yang tidak proporsional dalam menentukan nilai variabel dependen yang diprediksi
- Nilai *hat value* dianggap besar jika

$$IH > \frac{2p}{n} \quad \text{untuk sampel besar}$$

$$IH > \frac{3p}{n} \quad \text{untuk sampel kecil}$$

dengan :  $I = \text{matriks identitas} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$

p = jumlah variabel independen

n = jumlah observasi

**REGRESI LINIER SEDERHANA**

**Pengertian**

- Sebuah model yang berusaha menerangkan hubungan antara dua buah variabel, dimana satu variabel menjadi **prediktor** atas satu variabel lainnya.
- Sebuah model yang berusaha menerangkan hubungan dependensi dari dua variabel, di mana salah satu variabel bersifat **dependen** atas variabel lainnya.
- Hubungan ini tidak bersifat deterministik

**Model Dasar**

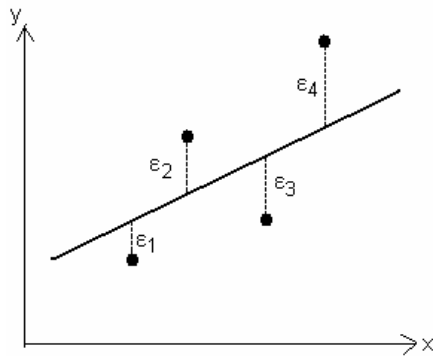
$$Y = \alpha + \beta x + \varepsilon$$

$\alpha$  adalah parameter intercept

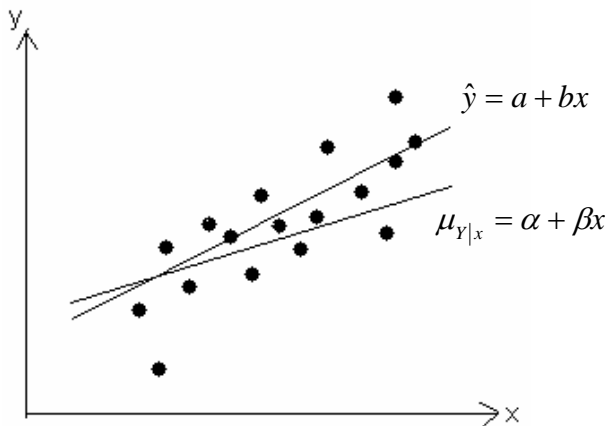
$\beta$  adalah parameter slope

$\varepsilon$  adalah residual yang merupakan variabel random dengan  $E(\varepsilon) = 0$  dan berdistribusi normal

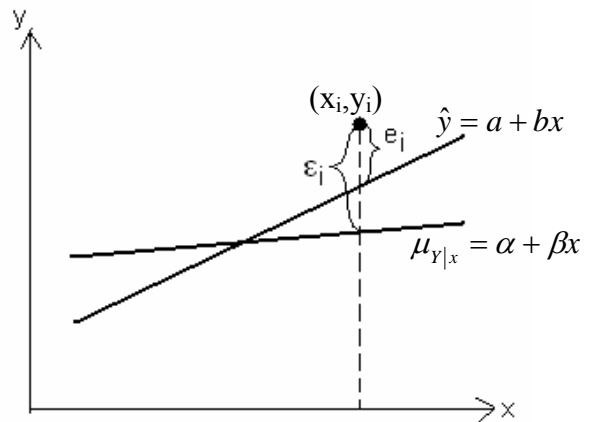
**Contoh Garis Regresi**



**Garis Regresi vs Garis Regresi di-fit-kan**



**Random Error vs Residual**



**Metode Least Square**

- Digunakan untuk mencari parameter  $a$  dan  $b$  yang merupakan estimator bagi  $\alpha$  dan  $\beta$
- Residual :  $e_i = y_i - \hat{y}_i$
- Parameter  $a$  dan  $b$  dicari sedemikian hingga yang akan meminimalkan nilai  $SSE = \sum (e_i)^2$

$$SSE = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - a - bx_i)^2$$

Selanjutnya, dengan menurunkan terhadap  $a$  dan  $b$ , akan diperoleh :

$$\frac{\partial(SSE)}{\partial a} = -2(y_i - a - bx_i)$$

$$\frac{\partial(SSE)}{\partial b} = -2(y_i - a - bx_i)x_i$$

Dengan menetapkan nilai turunan di atas sama dengan nol dan menyederhanakan persamaan yang diperoleh akan di dapatkan :

$$b = \frac{n \sum_{i=1}^n x_i y_i - \left( \sum_{i=1}^n x_i \right) \left( \sum_{i=1}^n y_i \right)}{n \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$a = \frac{\sum_{i=1}^n y_i - b \sum_{i=1}^n x_i}{n} = \bar{y} - b\bar{x}$$

**Inferensi pada Model Regresi**

- Perhatikan bahwa  $a$  dan  $b$  adalah variabel random (mengapa?)
- Jika diasumsikan  $\epsilon$  berdistribusi normal dengan  $\mu_\epsilon=0$  dan  $\sigma_\epsilon^2=\sigma^2$ , maka jika  $x$  diasumsikan bukan variabel random maka akan berlaku :  
 $\sigma_{Y|x_i}^2 = \sigma^2$ , untuk  $i = 1,2, \dots$

**Variansi Parameter  $b$  dan  $a$**

$$\sigma_B^2 = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \qquad \sigma_A^2 = \frac{\sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})} \sigma^2$$

**Estimasi atas Nilai  $\sigma^2$**

$$s^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - 2}$$

Besaran ini sering pula disebut sebagai **Mean Squared Error (MSE)**

**Inferensi atas Parameter  $\beta$**

- Selang kepercayaan untuk  $\beta$

$$b - \frac{t_{\alpha/2} s}{\sqrt{S_{xx}}} < \beta < b + \frac{t_{\alpha/2} s}{\sqrt{S_{xx}}}$$

dimana t berdistribusi t-student dengan d.o.f = n-2

- Uji hipotesis menggunakan statistik :

$$T = \frac{B - \beta}{s / \sqrt{S_{xx}}}$$

di mana t berdistribusi t-student dengan d.o.f = n-2

**Inferensi atas Parameter  $\alpha$**

- Selang kepercayaan untuk  $\alpha$  :

$$a - \frac{t_{\alpha/2} s \sqrt{\sum_{i=1}^n x_i^2}}{\sqrt{nS_{xx}}} < \alpha < a + \frac{t_{\alpha/2} s \sqrt{\sum_{i=1}^n x_i^2}}{\sqrt{nS_{xx}}}$$

di mana t berdistribusi t-student dengan d.o.f = n-2

- Uji hipotesis menggunakan statistik :

$$T = \frac{A - \alpha}{s \sqrt{\frac{\sum_{i=1}^n x_i^2}{nS_{xx}}}}$$

**Contoh Penggunaan**

Misalkan, diperoleh data nilai investasi asing dan tingkat pajak yang dikenakan seperti ditunjukkan oleh tabel 1

Daerah	Tingkat pajak (%)	Nilai Investasi (ribu US\$)
1	15	4178
2	14	4100
3	20	3978
4	25	3850
5	30	3340
6	27	3500
7	5	4900
8	7.5	4625
9	12.5	4260
10	11	4400

**OUTPUT SPSS**

**Coefficients(a)**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	5032,046	81,094		62,052	,000	4845,042	5219,049
	PAJAK	-55,027	4,376	-,976	-12,573	,000	-65,119	-44,935

a Dependent Variable: INVEST

**Seberapa Baik Model yang dibuat?**

- Jika kita tidak memiliki informasi mengenai tingkat pajak, prediksi hanya didasarkan pada nilai rata-rata investasi  $\bar{Y}$ .
- Error yang didapat di sini :  $Y_i - \bar{Y}$
- Dengan menggunakan model regresi, error akan menjadi :

$$(Y_i - \bar{Y}) - (Y_i - \hat{Y}) = (\hat{Y}_i - \bar{Y})$$

**Penguraian Variansi**

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

$$SST = SSR + SSE$$

**Ukuran Kebaikan Model : Koefisien Determinasi**

$$r^2 = \frac{SSR}{SST}$$

Menunjukkan berapa persen variansi pada nilai Y yang dapat dijelaskan oleh variabel prediktor (X).

**Output SPSS Untuk Kebaikan Model :**

**ANOVA<sup>(b)</sup>**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1951807,459	1	1951807,459	158,090	,000(a)
	Residual	98769,441	8	12346,180		
	Total	2050576,900	9			

a Predictors: (Constant), PAJAK

b Dependent Variable: INVEST

**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,976(a)	,952	,946	111,11337

a Predictors: (Constant), PAJAK

**ANALISIS KORELASI**

Analisis korelasi adalah tools statistika yang digunakan untuk mengukur kekuatan hubungan antara dua variabel random melalui sebuah *koefien korelasi*

**Penurunan Koefisien Korelasi**

- Sekarang, asumsikan bahwa X adalah variabel random berdistribusi normal dengan rataaan sama dengan  $\mu_x$  dan variansi  $\sigma_x^2$
- Maka, *joint density function (jdf)* untuk X dan Y adalah :

$$f(x, y) = n(y|x; \alpha + \beta x, \sigma)n(x; \mu_x, \sigma_x)$$

$$= \frac{1}{2\pi\sigma_x\sigma} \exp\left\{-\frac{1}{2}\left[\left(\frac{y - (\alpha + \beta x)}{\sigma}\right)^2 + \left(\frac{x - \mu_x}{\sigma_x}\right)^2\right]\right\}$$

- Ingat bahwa :

$$Y = \alpha + \beta X$$

- Karena rataaan error random  $\varepsilon$  adalah nol, maka

$$\mu_y = \alpha + \beta\mu_x \qquad \sigma_y^2 = \sigma^2 + \beta^2\sigma_x^2$$

- Selanjutnya, dengan mensubstitusikan  $\alpha$  dan  $\sigma^2$  pada persamaan untuk  $f(x,y)$  akan didapatkan distribusi normal bivariat :

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \times \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x - \mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x - \mu_x}{\sigma_x}\right)\left(\frac{y - \mu_y}{\sigma_y}\right) + \left(\frac{y - \mu_y}{\sigma_y}\right)^2\right]\right\}$$

**Koefisien Korelasi Populasi**

- Pada ekspresi terakhir, terdapat parameter  $\rho$  yang nilainya :

$$\rho^2 = \beta^2 \frac{\sigma_x^2}{\sigma_y^2}$$

- Parameter ini menunjukkan tingkat asosiasi linear antara dua variabel random (X dan Y)

**Estimasi untuk  $\rho$**

Koefisien korelasi populasi  $\rho$  dapat diestimasi dengan menggunakan rumus berikut :

$$r = b \sqrt{\frac{S_{xx}}{S_{yy}}}$$

**REGRESI LINIER MAJEMUK**

Regresi linier majemuk merupakan generalisasi dari regresi linier sederhana dan memiliki model umum :

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$$

Model regresi di atas dapat didekati dengan menggunakan data sampel sehingga didapatkan model yang di-fit-kan :

$$\hat{Y} = a + b_1 X_1 + b_2 X_2 + \dots + b_k X_k$$

**Residual pada Model yang di-fit-kan**

$$SSE = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - b_0 - b_1 X_{1i} - b_2 X_{2i} - b_k X_{ki})^2$$

Dengan mendiferensiasi SSE terhadap  $b_0, b_1, \dots, b_k$  akan didapatkan :

$$\begin{matrix} nb_0 & + & b_1 \sum_{i=1}^n x_{1i} & + & b_2 \sum_{i=1}^n x_{2i} & + \dots + & b_k \sum_{i=1}^n x_{ki} & = & \sum_{i=1}^n y_i \\ b_0 \sum_{i=1}^n x_{1i} & + & b_1 \sum_{i=1}^n x_{1i}^2 & + & b_2 \sum_{i=1}^n x_{1i} x_{2i} & + \dots + & b_k \sum_{i=1}^n x_{1i} x_{ki} & = & \sum_{i=1}^n x_{1i} y_i \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ b_0 \sum_{i=1}^n x_{ki} & + & b_1 \sum_{i=1}^n x_{ki} x_{1i} & + & b_2 \sum_{i=1}^n x_{ki} x_{2i} & + \dots + & b_k \sum_{i=1}^n x_{ki}^2 & = & \sum_{i=1}^n x_{ki} y_i \end{matrix}$$

**Penyelesaian dengan Menggunakan Matriks**

$$SSE = (\mathbf{y} - \mathbf{Xb})'(\mathbf{y} - \mathbf{Xb})$$

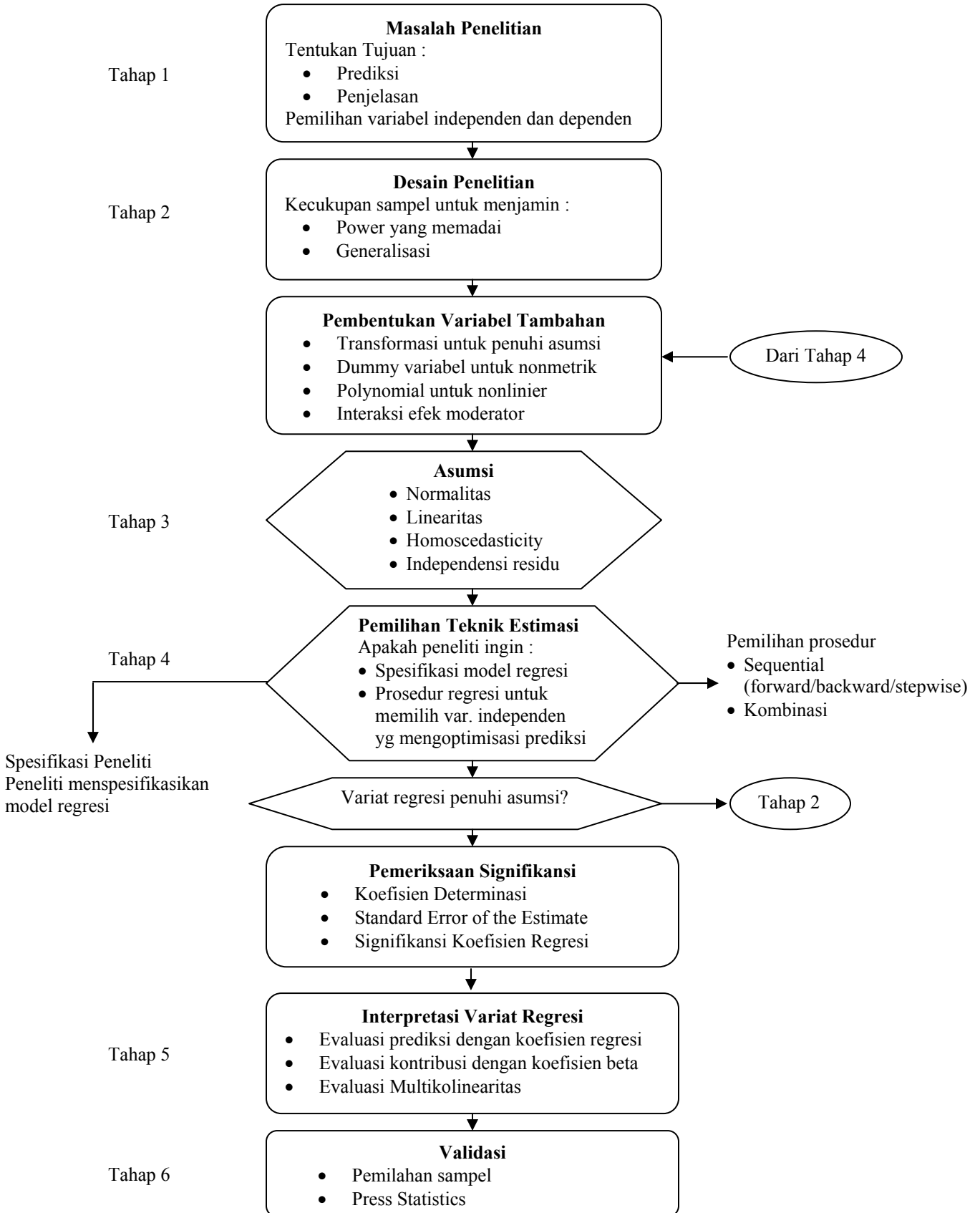
Untuk meminimalkan SSE :  $\frac{\partial}{\partial \mathbf{b}} (SSE) = 0$

Seolusi yang akan didapatkan adalah :  $(\mathbf{X}'\mathbf{X})\mathbf{b} = \mathbf{X}'\mathbf{y}$

**Partisi Variasi**

Source	Sum of Squares	d.f
Regresi	$\mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$	k+1
Error	$\mathbf{Y}'[\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y}$	n-(k+1)
Total	$\mathbf{Y}'\mathbf{y}$	n

**Langkah-langkah dalam Analisis Regresi Linier Majemuk**



**Tujuan Pemakaian Regresi Linier Majemuk**

Regresi linier berganda umum digunakan untuk dua tujuan berikut :

1. Prediksi
  - Untuk memaksimalkan kemampuan prediksi secara keseluruhan dari variabel independen
  - Untuk membandingkan dua atau lebih kelompok variabel independen
  - Prediksi sering termaksimasi pada saat mengeluarkan tafsiran → deret waktu
2. Penjelasan
  - Model harus berdasarkan pada hubungan teoritis dari variabel dependen dan independen
  - Tiga sudut pandang dalam mengartikan hasil :
    - pentingnya suatu variabel independen
    - tipe-tipe hubungan yang ditemukan
    - antarhubungan diantara variabel independen

Hubungan fungsional menghitung nilai eksak, sedangkan hubungan statistikal memperkirakan nilai rata-rata.

**DESAIN**

**Pemilihan Variabel**

- Pemilihan sebaiknya berdasarkan secara prinsip pada kerangka kerja teoritikal atau konseptual
- Jika tujuan riset adalah penyelidikan (*exploratory*), kita sebaiknya waspada terhadap:
  - Kesalahan pengukuran (*measurement error*)
  - Kesalahan perincian (*specification error*)
    - ✓ Melibatkan variabel yang tidak berhubungan
    - ✓ Tidak melibatkan variabel yang sebenarnya berhubungan

**Ukuran Sampel**

Ukuran sampel berpengaruh terhadap *statistical power* dan generalisasi model. Nilai  $R^2$  minimum yang dinyatakan signifikan secara statistik untuk power 0.8

Uk. Sampel	Level signifikansi = 0.01				Level signifikansi = 0.05			
	Jumlah Variabel Independen				Jumlah Variabel Independen			
	2	5	10	20	2	5	10	20
20	45	56	71	NA	39	48	64	NA
50	23	29	36	49	19	23	29	42
100	13	16	20	26	10	12	15	21
250	5	7	8	11	4	5	6	8
500	3	3	4	6	3	4	5	9
1000	1	2	2	3	1	1	2	2

NA = *Not Applicable* (tidak dapat dilakukan)

Untuk mendapatkan tingkat generalisasi yang mencukupi, aturan umum yang berlaku adalah jumlah sampel paling sedikit lima kali dari jumlah variabel independen (Hair, 1998)

**Fixed vs Random Effect Models**

Perbedaan kedua model didasari oleh asumsi model yang berbeda :

- pada *fixed effect* diasumsikan nilai pengukuran variabel independen tidak random

– pada *random effect* diasumsikan nilai pengukuran variabel independen adalah random  
**ASUMSI-ASUMSI**

- Hubungan linier
- Variansi error konstan (homoscedasticity)
- Independensi error
- Error berdistribusi normal

### Penggunaan Variabel Dummy

- Setiap variabel nonmetrik dengan  $k$  kategori dapat diwakili oleh  $k-1$  variabel dummy
- Untuk  $k$  lebih dari 2, terdapat 2 cara :
  - indikator coding
  - effect coding

Contoh penggunaan Variabel Dummy

- Untuk  $k = 2$ 
  - $X_1 = 1$ , jika perempuan
  - $X_1 = 0$ , jika laki-laki
- Untuk  $k = 3$ , menggunakan *indicator coding*
  - $X_1 = 1, X_2 = 0$  jika mahasiswa S3
  - $X_1 = 0, X_2 = 1$  jika mahasiswa S2
  - $X_1 = 0, X_2 = 0$  jika mahasiswa S1

Disini mahasiswa S1 berperan sebagai *control group*. Nilai koefisien merupakan selisih nilai variabel dependen dari grup bersangkutan terhadap grup kontrol

$X_1 = 1, X_2 = 0$  jika mahasiswa S3  
 $X_1 = 0, X_2 = 1$  jika mahasiswa S2  
 $X_1 = -1, X_2 = -1$  jika mahasiswa S1

Disini nilai koefisien merupakan selisih nilai variabel dependen rai grup bersangkutan terhadap nilai rataan dari semua grup yang ada

### Estimasi Model

- Confirmatory specification
  - Juga dikenal sebagai metode enter
  - Setiap variabel dimasukkan pada model
  - Perhatian : model parsimony (model sederhana)
- Sequential Search
- Combinatorial

**ANALISIS FAKTOR**

**TUJUAN**

Terdapat dua tujuan utama seorang peneliti menggunakan *tools* analisis faktor:

1. mengidentifikasi konstruksi yang mendasari sebuah data
2. mengurangi jumlah variabel penelitian sehingga lebih mudah ditangani

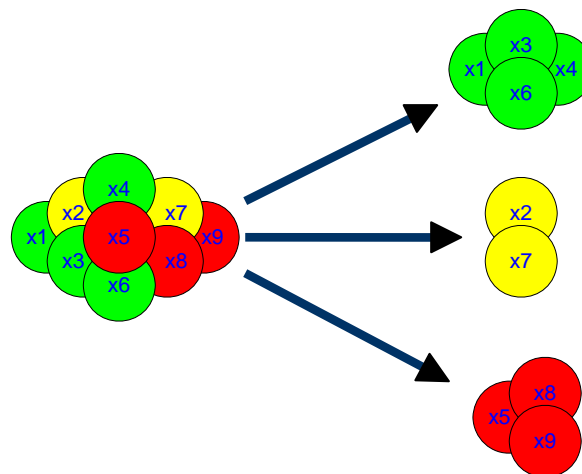
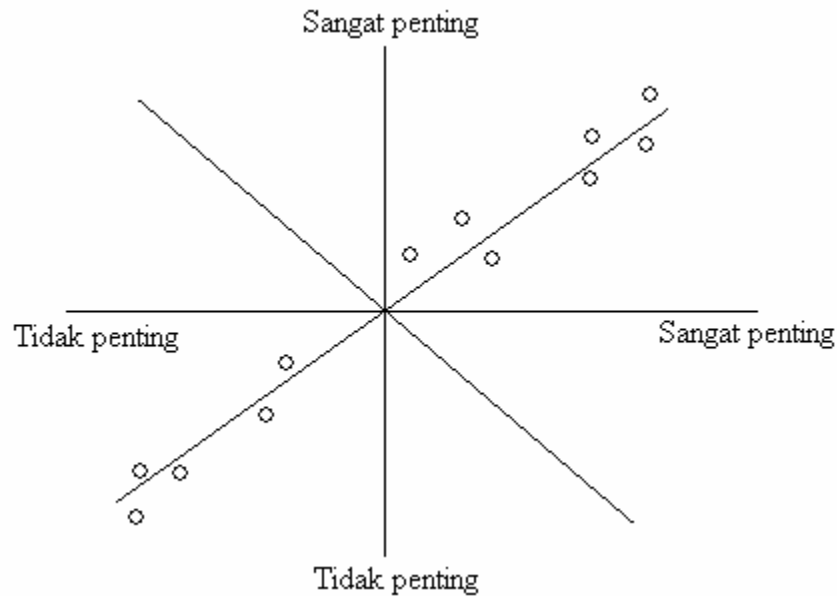
**ILUSTRASI AWAL**

Sebuah pengelola pusat perbelanjaan melakukan survey terhadap calon *tenant* yang prospektif. Para calon diminta memberikan penilaian dalam skala -5 sampai dengan 5 untuk atribut yang mereka pertimbangkan dalam memilih pusat perbelanjaan yang akan mereka sewa (-5 menunjukkan bahwa atribut tersebut tidak penting dan angka 5 menunjukkan bahwa atribut tersebut sangat penting). Untuk menyederhanakan permasalahan, di sini diasumsikan bahwa atribut yang diukur hanya dua buah yaitu :

$X_1$  : Tampilan Fisik Plaza/Mal yang menarik

$X_2$  : Letak yang strategis

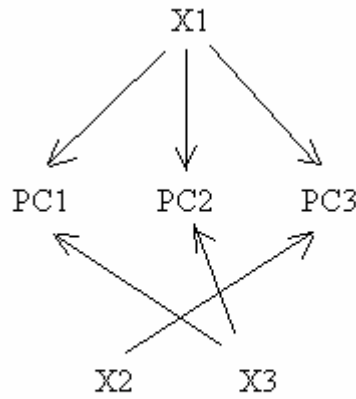
Hasil repons dari para calon *tenant* yang prospektif ini digambarkan dalam grafik pada gambar



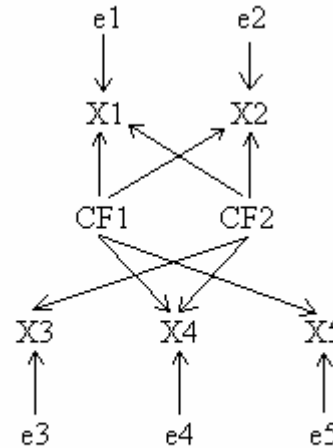
**IDE DASAR ANALISIS FAKTOR**

- Metodologi analisis faktor merupakan proses transformasi variabel-variabel original (awal) menjadi variabel-variabel baru yang saling tidak berkorelasi.
- Variabel baru ini disebut dengan faktor
- Masing-masing faktor merupakan kombinasi linier dari variabel original atau variabel original merupakan kombinasi linier dari faktor-faktor.

**Principal Component Analysis (PCA) vs Common Factor Analysis (CFA)**



PRINCIPAL COMPONENT ANALYSIS



COMMON FACTOR ANALYSIS

**PERSAMAAN PCA**

$$PC_1 = w_{11}X_1 + w_{12}X_2 + \dots + w_{1p}X_p$$

$$PC_2 = w_{21}X_1 + w_{22}X_2 + \dots + w_{2p}X_p$$

$$PC_m = w_{m1}X_1 + w_{m2}X_2 + \dots + w_{mp}X_p$$

**PERSAMAAN CFA**

$$X_1 = v_{11}CF_1 + v_{12}CF_2 + \dots + v_{1m}CF_m + e_1$$

$$X_2 = v_{21}CF_1 + v_{22}CF_2 + \dots + v_{2m}CF_m + e_2$$

$$X_p = v_{p1}CF_1 + v_{p2}CF_2 + \dots + v_{pm}CF_m + e_p$$

- Salah satu ukuran jumlah informasi yang dibawa atau diteruskan oleh masing-masing faktor adalah variansi nya.
- Sehubungan dengan hal ini, faktor-faktor disusun dengan urutan variansi yang menurun.
- Faktor pertama merupakan faktor yang paling informatif (memiliki variansi terjelaskan yang maksimum) dan faktor terakhir adalah faktor yang paling sedikit meneruskan informasi (memiliki variansi terjelaskan yang minimum).
- Jumlah faktor yang dibangkitkan adalah maksimum sebanyak jumlah variabel awal.
- Namun dikaitkan dengan tujuannya, pada umumnya jumlah faktor yang dibangkitkan adalah sejumlah kecil faktor yang dinilai mencukupi oleh peneliti.

**KECUKUPAN DATA**

- Kecukupan jumlah sampel secara keseluruhan dapat dilihat dari angka **Kaiser-Meyer-Olkin (KMO)**

Marvelous (0,90)	Mediocre (0,60)
Meritorius (0,80)	Miserable (0,50)
Middling (0,7)	Unacceptable (<0,50)

- Diagonal matriks korelasi anti-image disebut MSA (Measure of Sampling Adequacy) menunjukkan kecukupan untuk masing-masing variabel. Jika nilainya kurang dari 0,50 maka penyertaan variabel tersebut perlu dipertimbangkan kembali.

**INPUT ANALISIS FAKTOR**

- Variabel awal harus memiliki landasan teoritis/logis
- Sejauh mungkin metrik
- Memiliki magnitude (skala) yang seragam (jika tidak harus distandarisasikan atau gunakan metode input matriks korelasi)
- Contoh :

Responden Ke -	Nilai yang diberikan				
	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>
1	9	6	9	2	2
2	4	6	2	6	7
3	0	0	5	0	0
4	2	2	0	9	9
5	6	9	8	3	3
6	3	8	5	4	7
7	4	5	6	3	6
8	8	6	8	2	2
9	4	4	0	8	8
10	2	8	4	5	7

- X<sub>1</sub> : biaya perizinan usaha di pusat lebih besar dibandingkan biaya perizinan usaha di daerah
- X<sub>2</sub> : pusat lebih sering menimbulkan masalah dibandingkan daerah
- X<sub>3</sub> : para pegawai yang melayani urusan investasi harus cukup ramah dan sopan
- X<sub>4</sub> : kami ingin agar kami dikenal secara personal
- X<sub>5</sub> : jika kami diperlakukan secara tidak profesional di suatu tempat maka kami tidak pernah akan mencoba untuk berinvestasi kembali di tempat tersebut.

**OUTPUT**

	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>
X <sub>1</sub>	1.00000	0.60980	0.46870	-0.01795	-0.09642
X <sub>2</sub>		1.00000	0.23048	0.18969	0.31863
X <sub>3</sub>			1.00000	-0.83183	-0.77394
X <sub>4</sub>				1.00000	0.92732
X <sub>5</sub>					1.00000

Matriks korelasi antar variabel

	Faktor ke-				
	1	2	3	4	5
Eigenvalue	2.7546	1.7749	0.3771	0.0650	0.0285
Difference	0.9797	1.3978	0.3121	0.0365	
Proportion	0.5509	0.3550	0.0754	0.0130	0.0057
Cumulative	0.5509	0.9059	0.9813	0.9943	1.0000

Nilai Eigen dari masing-masing komponen utama

	Faktor 1	Faktor 2	Komunalitas
X <sub>1</sub>	-0.30	0.85	0.81
X <sub>2</sub>	0.05	0.92	0.83
X <sub>3</sub>	-0.94	0.28	0.94
X <sub>4</sub>	0.95	0.23	0.94
X <sub>5</sub>	0.94	0.27	0.96

Bobot Faktor (Tanpa rotasi)

	Faktor 1	Faktor 2	Komunalitas
X <sub>1</sub>	-0.17	0.89	0.81
X <sub>2</sub>	0.018	0.90	0.83
X <sub>3</sub>	-0.89	0.41	0.94
X <sub>4</sub>	0.97	0.09	0.94
X <sub>5</sub>	0.97	0.13	0.96

Bobot Faktor (Rotasi Varimax)

	Faktor 1	Faktor 2
Tanpa Rotasi	2.755	1.775
Rotasi Varimax	2.735	1.794

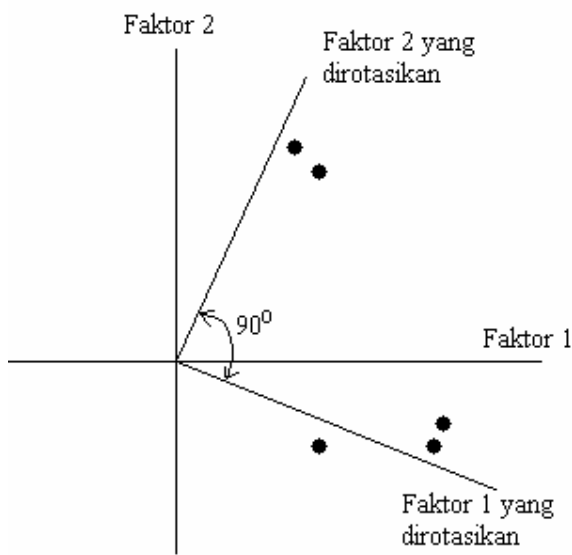
Variansi yang dijelaskan oleh masing-masing faktor =  $\sum_{i=1}^5 F_i$

	Faktor 1	Faktor 2
X <sub>1</sub>	-0.03915	0.49096
X <sub>2</sub>	0.08936	0.51083
X <sub>3</sub>	-0.31526	0.20227
X <sub>4</sub>	0.35929	0.07915
X <sub>5</sub>	0.35906	0.10182

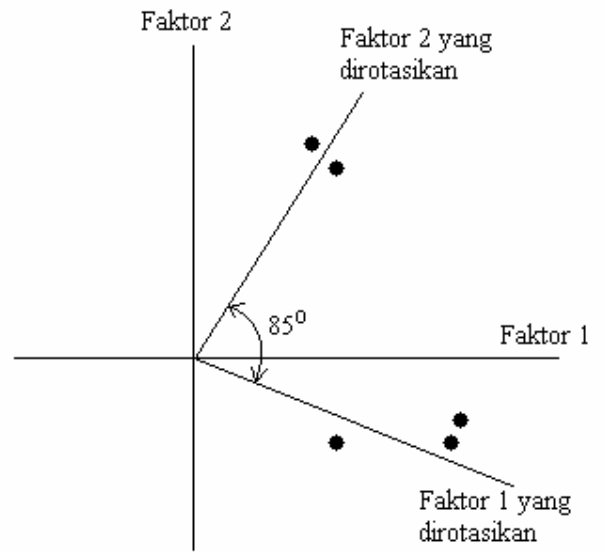
Koefisien Skor Faktor (*Standardized*)

Responden ke-	Faktor 1	Faktor 2
1	-0.91581	1.03767
2	0.93263	-0.00767
3	-1.05059	-1.94294
4	1.64856	-1.03405
5	-0.44159	1.04045
6	0.44671	0.30815
7	-0.00002	-0.03933
8	-0.80081	0.80036
9	1.44238	-0.3993
10	0.69027	0.09916

Skor Faktor



ROTASI ORTHOGONAL

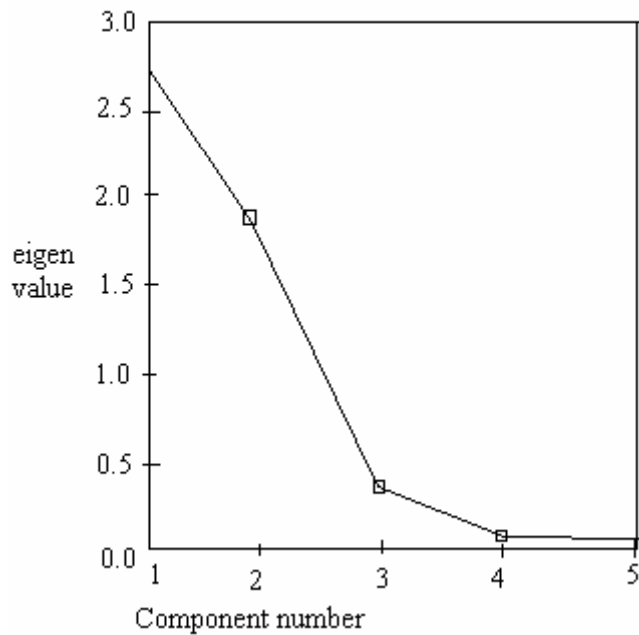


ROTASI OBLIQUE

Jumlah faktor yang diekstraksi

- Kriteria a priori : jumlah faktor ditentukan sendiri di awal oleh analis
- Kriteria nilai eigen : hanya nilai eigen lebih dari angka tertentu yang dipakai
- Kriteria persen variansi terjelaskan
- Kriteria scree plot : berdasarkan nilai eigen

SCREE PLOT



**VALIDASI ANALISIS FAKTOR**

Salah satu cara mem-validasi hasil analisis faktor adalah dengan melihat *replicability* hasil analisis faktor. Caranya dengan membagi dua data tersebut, atau dengan menggunakan data lain yang terpisah. Kemudian bandingkan hasilnya dengan hasil analisis faktor yang telah didapat sebelumnya.

**PENGGUNAAN UNTUK ANALISIS LEBIH LANJUT**

Output dari analisis faktor (skor faktor) dapat digunakan sebagai input untuk analisis-*analisis* statistika multivariat lainnya (misalkan analisis cluster, diskriminan, dan regresi linier berganda). Dengan menggunakan analisis faktor (*principal component analysis*) jumlah variabel awal akan direduksi menjadi sejumlah faktor yang berhasil diekstraksi sehingga akan lebih memudahkan perhitungan.

Contoh :

$$\text{Skor faktor 1} = -0.039 X_1 + 0.089 X_2 - 0.315 X_3 + 0.359 X_4 + 0.359 X_5$$

**ANALISIS DISKRIMINAN**

**TUJUAN**

- Menguji apakah terdapat perbedaan yang signifikan antara dua kelompok (atau lebih), didasarkan pada pusat kelompok (*groups centroids*).
- Menentukan kombinasi linier dari variabel-variabel prediktor untuk memisahkan kelompok-kelompok dengan jalan memaksimalkan variasi antar-group (*between group*) relatif terhadap variasi dalam grup (*within-grup*). Atau dengan kata lain objek-objek dalam kelompok yang berbeda dipisahkan secara maksimal.
- Membangun prosedur untuk menempatkan objek-objek baru yang profilnya (kecuali keanggotaan kelompoknya) diketahui ke dalam salah satu dari kelompok yang ada.
- Menentukan variabel mana yang paling mampu menjelaskan perbedaan antar kelompok.

**MODEL DASAR**

- Analisis diskriminan melibatkan penurunan kombinasi linier dari dua (atau lebih) variabel-variabel independen yang akan membedakan (*discriminate*) secara paling baik kelompok-kelompok yang telah didefinisikan secara *a priori*.

$$Z = const + b_1 X_1 + b_2 X_2 + b_3 X_3 + \dots + b_n X_n$$

di mana

Z = skor diskriminan

b = bobot diskriminan

X = variabel prediktor (independen)

- Dengan merata-ratakan skor diskriminan untuk semua individu dalam sebuah kelompok tertentu, akan didapatkan rata-rata kelompok (*centroid*)
- Centroid ini mengindikasikan lokasi yang paling tipikal dari sebuah individu dalam sebuah kelompok tertentu.
- Selanjutnya adalah menentukan garis batas kapan suatu observasi dimasukkan ke dalam kelompok dengan centroid tertentu
- Garis batas tersebut dikenal juga sebagai *cut-off score*

**PERBANDINGAN ANALISIS DISKRIMINAN DAN REGRESI MAJEMUK**

<b>Analisi Regresi</b>	<b>Analisis Diskriminan</b>
Variabel dependen berdistribusi normal	Variabel dependen tetap (kategorikal)
Variabel independen tetap (metrik)	Variabel independen berdistribusi normal
Tujuan : menjelaskan dan memprediksi nilai rata-rata populasi variabel berdasarkan nilai variabel independen yang telah diketahui dan tetap	Tujuan : mencari kombinasi linier dari variabel independen yang meminimasi kemungkinan atau probabilitas melakukan kesalahan klasifikasi individu atau objek ke dalam kelompoknya masing-masing
Suatu model formal di mana asumsi pasti dibuat untuk mendapatkan estimasi parameter yang memiliki properti statistik yang diinginkan	Suatu strategi untuk mencari alat pengelompokan individu atau objek secara akurat

**ASUMSI DASAR**

- Variabel independen berdistribusi normal multivariat (masalah estimasi fungsi)

- Matriks variansi-kovariansi dari variabel-variabel independen dalam masing-masing kelompok adalah sama (pengaruh: klasifikasi objek)
- Multikolinearitas
- Ukuran sampel  
Studi empirik mensyaratkan bahwa perbandingan minimum antara jumlah variabel independen dengan jumlah sampel adalah 1 : 5 (idealnya 1 : 20)

### ESTIMASI MODEL DISKRIMINAN

- **Simultaneous Estimation (Enter)**  
Semua variabel independen diinput secara bersamaan untuk membentuk model tanpa mempertimbangkan daya pembeda antar variabel
- **Stepwise Estimation**  
Variabel independen diinput satu persatu ke dalam model berdasarkan daya pembedanya. Metode ini sangat bermanfaat manakala analisis melibatkan variabel independen dalam jumlah yang besar.

### PENGUJIAN SIGNIFIKANSI MODEL

- Pengujian signifikansi model dilakukan dengan menguji hipotesis bahwa rata-rata populasi atas nilai fungsi diskriminan pada masing-masing kelompok sama ( $H_0 : \mu_A = \mu_B$ )  
Jika  $Z_A = Z_B$ , signifikan secara statistik. Artinya model tidak mampu menunjukkan perbedaan yang berarti dari rata-rata nilai kelompok A dan B.
- Jika model ini diterima, konsekuensi logisnya adalah model diskriminan yang dibuat tidak mampu membuat suatu perbedaan.
- Untuk menguji hipotesis ini secara statistik digunakan nilai  $\lambda$  Wilk (Wilk's  $\lambda$ ).
- Wilk's  $\lambda$  didefinisikan sebagai rasio variansi dalam kelompok terhadap variansi antar kelompok.
- Nilai Wilk's  $\lambda$  berkisar antara 0 sampai 1, di mana semakin kecil nilainya semakin signifikan model yang telah dibuat.

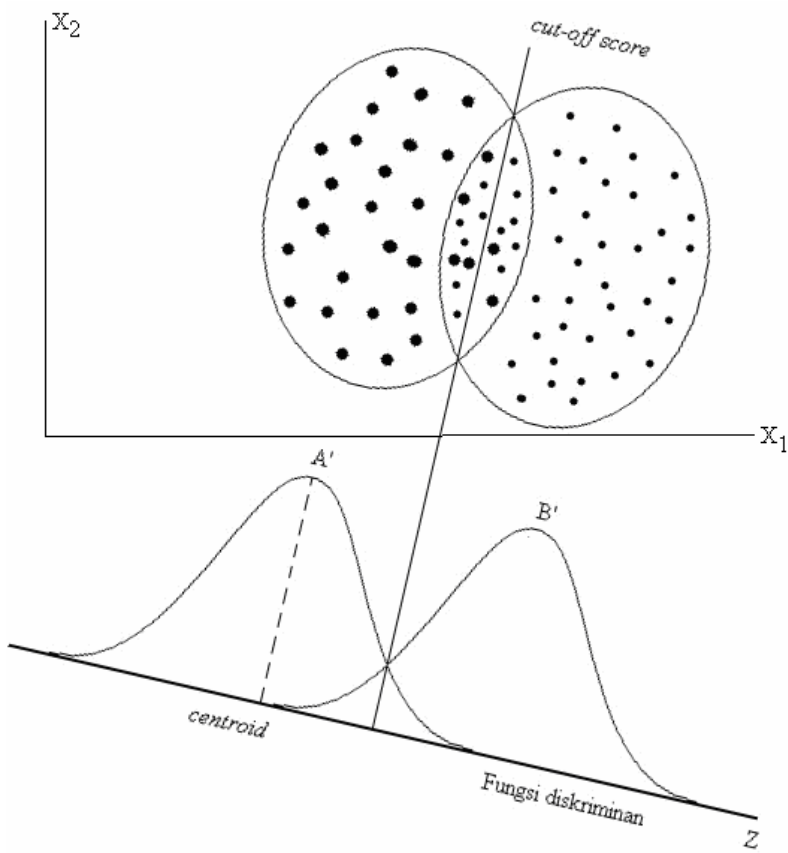
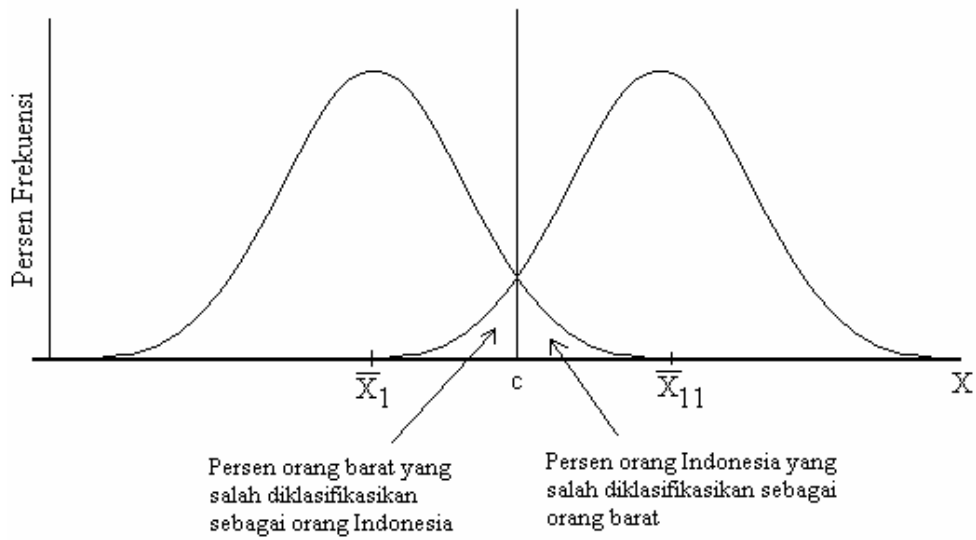
$$\text{Wilk's } \lambda = \frac{\text{var dalam kelompok}}{\text{var antar kelompok}} = \frac{\text{kecil}}{\text{besar}} \rightarrow \text{kecil}$$

### INTERPRETASI MODEL ANALISIS DISKRIMINAN

- *Canonical Discriminant Function Coefficients*
- *Standardized coefficient*
- *Canonical loadings* atau *discriminant loadings*
- *Centroid*

### VALIDASI OUTPUT ANALISIS DISKRIMINAN

- Membandingkan dengan probabilitas asal (verifikasi)
- Metode *jack-knife*
- Metode *hold-out*



Cut-off score :

$$c = \frac{Z_A + Z_B}{2}$$

jika ukuran sampel sama

$$c = \frac{N_A Z_A + N_B Z_B}{N_A + N_B}$$

jika ukuran sampel tidak sama

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	.795	12.881	4	.012

Jumlah kelompok = 2 → karena test of function = 2 – 1 = 1

Jika Wilks Lambda signifikan (<0,05) → dapat dilanjutkan

Jika Wilks Lambda tidak signifikan (>0,05) → tidak dapat dilanjutkan

Canonical Discriminant Function Coefficients

	Function
	1
PDPTN	.129
UKRN	.084
USIA	.875
JMLPROY	.017
(Constant)	-9.399

Standardized Canonical Discriminant Function Coefficients

	Function
	1
PDPTN	.111
UKRN	.862
USIA	.781
JMLPROY	.044

Structure Matrix

	Function
	1
UKRN	.636
USIA	.540
JMLPROY	.331
PDPTN	.134

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions  
 Variables ordered by absolute size of correlation within function

Fungsi diskriminan :

$$Z = -9,399 + 0,129 \text{ PDPTN} + 0,084 \text{ UKRN} + 0,875 \text{ USIA} + 0,017 \text{ JMLPROY}$$

Jika ingin melihat mana yang paling membedakan, lihat dari standardized Function > 0,3 maka signifikan

Beda function di structure matrix → masih ada kolinearitas

Pada structure matrix, nilai 0,636 menunjukkan nilai korelasi ukuran dengan Z

Semakin tinggi korelasi maka semakin mampu untuk memprediksi

Function at Group Centroids

MINAT	Function
	1
1.00	-.380
2.00	.657

Unstandardized canonical discriminant functions evaluated at group means

Classification Result<sup>a</sup>

			Predicted Group Membership		Total
			1.00	2.00	
Original	Count	1.00	31	7	38
		2.00	6	16	22
	%	1.00	81.6	18.4	100.0
		2.00	27.3	72.7	100.0

a. 78.3% of original grouped cases correctly classified

---

---

## ANALISIS CLUSTER

### TUJUAN

- Tujuan analisis cluster adalah untuk mengelompokkan objek penelitian (orang, perusahaan, atau produk) ke dalam kelompok-kelompok dengan menggunakan atribut (variabel) pengelompok tertentu.
- Dalam satu kelompok diharapkan anggotanya memiliki sifat kemiripan yang tinggi dan anggota satu kelompok dengan anggota kelompok yang lain diharapkan memiliki perbedaan yang berarti.

### METODOLOGI

- Pengelompokkan data pada analisis cluster dapat didasarkan pada kesamaan (*similarity*) atau ketidaksamaan (*dissimilarity*) dari masing-masing objek.
- Proses dimulai dengan pengambilan p pengukuran variabel dan n objek pengamatan (matriks  $n \times p$ ). Matriks tersebut ditransformasikan ke dalam bentuk matriks similaritas berukuran  $n \times n$  yang dihitung berdasarkan pasangan-pasangan obyek p variabel.
- Kemudian suatu algoritma pengelompokkan dipilih.
- Terdapat dua pendekatan dalam algoritma pengelompokkan, yaitu pendekatan hierarkis (*hierarchical clustering*) dan pendekatan non-hierarkis (*nonhierarchical clustering*)

### HUBUNGAN DENGAN TEKNIK REDUKSI DATA LAINNYA

Analisis Cluster vs Analisis Komponen Utama

- ✓ AC : lebih terfokus untuk mereduksi data obyek
- ✓ AKU : lebih terfokus untuk mereduksi variabel

Analisis Cluster vs Analisis Diskriminan

- ✓ AC : dimulai dengan grup yang belum berdiferensiasi, kemudian grup itu dijadikan sub grup yang dapat dibedakan
- ✓ AD : dimulai dengan grup yang telah jelas (*well defined*) dan yang dipertanyakan adalah bagaimana grup tersebut dibedakan.

### UKURAN SIMILARITAS

- Ukuran jarak
- Koefisien korelasi
- Koefisien asosiasi

### TEKNIK PENGELOMPOKAN NON-HIERARKIS

- Sequential Threshold
- Parallel Threshold
- Optimizing

### HAL-HAL BERKAITAN DENGAN INPUT ANALISIS CLUSTER

- Satuan yang digunakan
- Adanya outlier
- Variabel-variabel pengelompok
- Dependensi antar variabel

**JUMLAH CLUSTER YANG DIBENTUK**

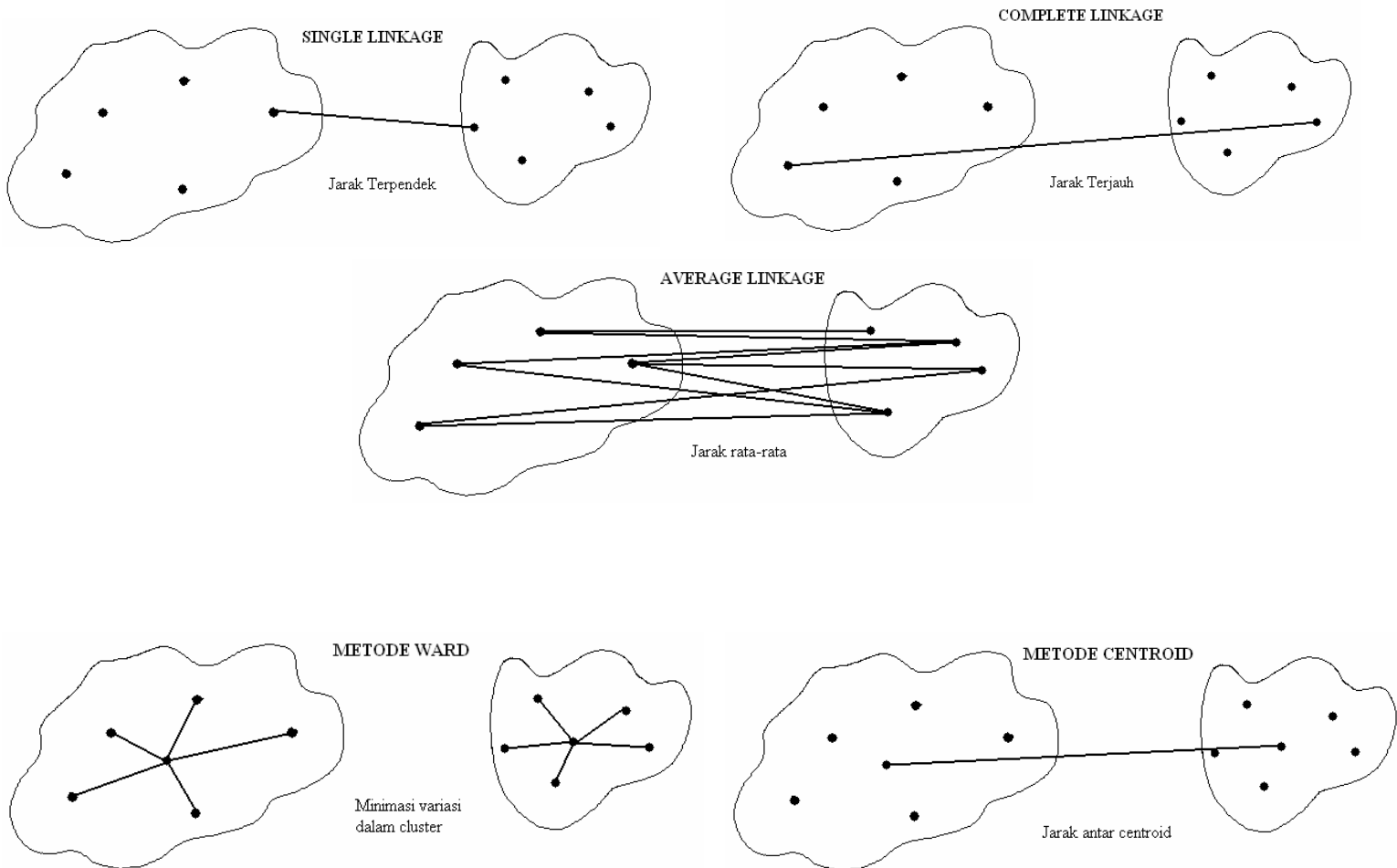
- Peneliti telah menetapkan jumlah cluster yang akan dibentuk
- Didasarkan pada ukuran variabilitas error (deviasi dari pusat cluster)
- Didasarkan pada kemudahan interpretasi

**INTERPRETASI CLUSTER YANG TERBENTUK**

- Setelah cluster terbentuk tugas analisis yang berikutnya adalah mendeskripsikan masing-masing cluster tersebut.
- Salah satu cara mendeskripsikan cluster yang terbentuk adalah dengan menggunakan centroid atau nilai rata-rata masing-masing variabel dari tiap objek yang berada dalam cluster tersebut
- Selanjutnya tiap cluster dapat diberi nama
- Penamaan cluster yang baik harus memperhatikan tujuan studi

**VALIDASI**

- Membandingkan hasil yang diperoleh dengan algoritma yang berbeda
- Membagi sampel menjadi dua
- Menggunakan variabel yang tidak digunakan untuk mengelompokkan
- Simulasi



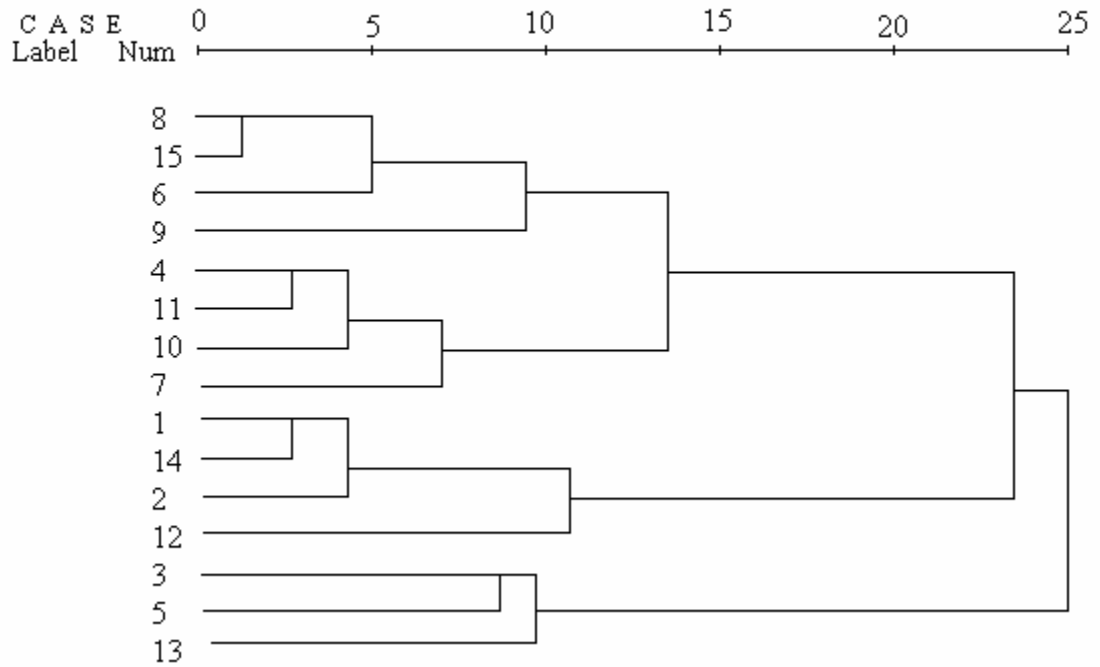
Tabel 1. Data Perusahaan yang akan dikelompokkan

Perusahaan ke-	Jumlah Karyawan	Pendapatan (milyar Rp.)	Usia	Jumlah Produk
1	43	1	7	3
2	54	2.7	7	4
3	49	3.1	3.5	7
4	43	2.52	2.5	2
5	52	5.4	4	7
6	29	1.8	3	6
7	37	3.4	4	2
8	27	2.6	4.5	5
9	20	2	2	3
10	34	1.2	3	3
11	48	1.3	2.5	2
12	53	0.9	6	6
13	41	4.2	6	5
14	47	2.3	7	2
15	30	3.3	5	4

Tabel 2. Agglomeration Schedule Pengelompokkan 15 Industri

Stage	Cluster-Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	8	15	0.773	0	0	6
2	4	11	1.178	0	0	4
3	1	14	1.524	0	0	5
4	4	10	2.187	2	0	7
5	1	2	2.488	3	0	11
6	6	8	2.693	0	1	9
7	4	7	3.323	4	0	12
8	3	5	3.546	0	0	10
9	6	9	4.263	6	0	12
10	3	13	4.515	8	0	14
11	1	12	4.726	5	0	13
12	4	6	5.621	7	9	13
13	1	4	9.483	11	12	14
14	1	3	10.525	13	10	0

Rescaled Distance Cluster Combine

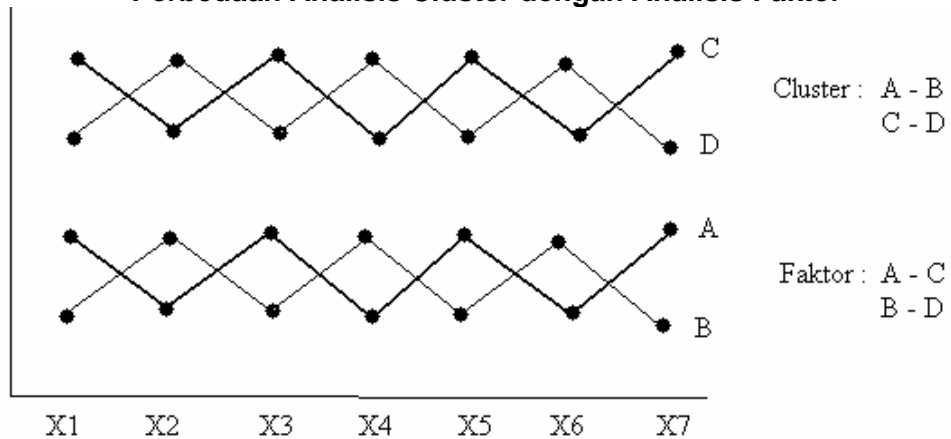


Dendrogram Pengelompokan 15 Industri

Tabel 3. Nilai Centroid dari masing-masing Cluster

Cluster	Jumlah Karyawan	Pendapatan (Milyar Rp.)	Usia Perusahaan	Jumlah Produk
Cluster 1 (4,6,7,8,9,10,11,15)	33.50	2.27	3.31	3.38
Cluster 2 (1,2,12, dan 14)	49.25	1.73	6.75	3.75
Cluster 3 (3,5, dan 13)	47.33	4.23	4.50	6.33

Perbedaan Analisis Cluster dengan Analisis Faktor



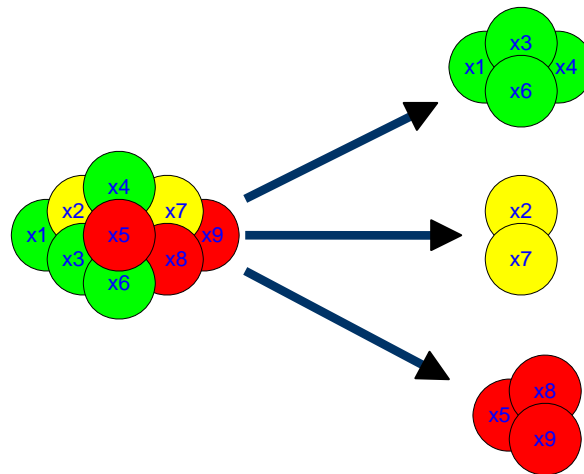
## TEKNIK-TEKNIK MULTIVARIAT

ANALISIS FAKTOR• **Tujuan:**

1. mereduksi data (jumlah variabel)
  - menseleksi variabel
  - membentuk variabel baru (*surrogate variables*)
2. mengidentifikasi satu set dimensi yang tidak tampak (tidak mudah teramati)

• **Ide dasar:**

Membagi variansi dari suatu variabel menjadi *common variance* dan *specific variance*. *Common variance* adalah variasi nilai variabel yang disebabkan oleh suatu *common factor*. Beberapa variabel yang besar dipengaruhi oleh suatu *common factor* akan dapat dikelompokkan.

• **Ilustrasi:**

Sebuah perusahaan ingin mengetahui faktor-faktor apakah yang mempengaruhi para pelanggannya untuk tetap berhubungan bisnis dengan perusahaan tersebut. Untuk itu disebarkan kuesioner yang berisikan penilaian terhadap 8 variabel pemasaran:

- |                            |                             |
|----------------------------|-----------------------------|
| X1 = kecepatan pengantaran | X5 = pelayanan              |
| X2 = tingkat harga         | X6 = image tenaga penjualan |
| X3 = fleksibilitas harga   | X7 = kualitas produk        |
| X4 = image perusahaan      |                             |

Setelah data dikumpulkan dan diolah dengan analisis faktor, didapatkan 2 buah faktor:

F1 = menyusun variabel X1, X2, X3, X5, dan X7

F2 = menyusun variabel X4 dan X6

• **Isu-isu dalam analisis faktor:**

1. input data (skala harus sama atau matriks korelasi)
2. kecukupan data secara *overall* (angka KMO > 0.5)
3. kecukupan data tiap variabel (diagonal matriks korelasi anti-image)
4. jumlah faktor yang diekstraksi
5. metode ekstraksi dan metode rotasi
6. penggunaan untuk teknik lanjutan (misalkan: regresi linier berganda, analisis diskriminan, atau analisis cluster)
7. eksploratori vs konfirmatori

- **Contoh-contoh penerapan analisis faktor:**
  - pengelompokan atribut produk
  - pengelompokan variabel-variabel yang mempengaruhi motivasi karyawan
  - pengelompokan indikator-indikator ekonomi
  - dan lain-lain

### ANALISIS CLUSTER

- **Tujuan:**

mengelompokkan sekumpulan data obyek ke dalam beberapa kelompok yang memiliki karakteristik tertentu
- **Ide dasar:**
  1. mengukur jarak antar obyek
  2. mengelompokkan obyek dengan jarak yang berdekatan (obyek yang similar)
- **Ilustrasi:**

Sebuah jaringan hotel meminta para tamu mengisi kuesioner dengan memberikan tingkat kepentingan terhadap variabel-variabel berikut:

  - kualitas restoran (X1)                      - keramahan pegawai (X4)
  - kualitas kamar (X2)                        - harga (X5)
  - kelengkapan fasilitas (X3)               - kemudahan akses (X6)

Setelah dilakukan analisis cluster didapatkan terdapat 3 kelompok tamu:

**Kelompok A** terdiri atas 60% responden memiliki nilai rata-rata yang tinggi pada variabel X1 dan X3 serta rendah pada X5

**Kelompok B** terdiri atas 30% responden memiliki nilai rata-rata yang tinggi pada X2, X4, dan X6 serta rendah pada X5

**Kelompok C** terdiri atas 10% responden memiliki nilai rata-rata yang tinggi pada X5 dan X6 serta nilai rendah pada X1 dan X3

Masing-masing kelompok responden mewakili segmen pasar dari hotel tersebut yang dapat diinterpretasikan sebagai berikut:

**Kelompok A** mewakili segmen tamu yang menuntut kualitas pelayanan tinggi dan tidak sensitif terhadap harga. Segmen ini dapat berupa para wisatawan.

**Kelompok B** mewakili segmen tamu menengah yang lebih mementingkan *comfort*.

**Kelompok C** mewakili segmen tamu yang mengutamakan kepraktisan. Segmen ini dapat berupa para *businessman*.
- **Perbedaan dengan analisis faktor**
  - analisis faktor mengelompokkan variabel
  - analisis cluster mengelompokkan objek
  - analisis faktor mengelompokkan berdasarkan kesamaan perilaku
  - analisis cluster mengelompokkan berdasarkan kedekatan jarak
- **Isu-isu dalam analisis cluster**
  - variabel yang digunakan untuk mengelompokkan
  - asumsi tidak terdapat multikolinearitas
  - teknik pengelompokan
- **Teknik-teknik pengelompokan:**
  1. metode hierarkis
  2. metode nonhierarkis

- **Contoh-contoh penerapan:**
  - penentuan segmen konsumen
  - pengelompokan industri
  - pengelompokan jenis pegawai
  - pengelompokan tipe produk (misal:bir)
  - dan lain-lain

### ANALISIS KONJOIN

- **Tujuan**

Analisis konjoin digunakan untuk memahami bagaimana responden membangun preferensi terhadap suatu produk atau jasa. Dari sini dapat diramalkan preferensi konsumen atas suatu konsep produk baru dan juga perkiraan *market share*-nya
- **Ide Dasar**

Analisis konjoin didasarkan pada premis sederhana bahwa konsumen mengevaluasi nilai atau utilitas total suatu produk dengan mengkombinasikan sejumlah utilitas yang disediakan oleh masing-masing atribut secara terpisah.  
Jadi analisis konjoin akan mengestimasi bobot masing-masing atribut dan nilai utilitas atas suatu level atribut.
- **Ilustrasi**

Sebuah bank akan menerbitkan produk kartu kredit baru. Bank tersebut telah mengidentifikasi atribut-atribut kartu kredit sebagai:

  - suku bunga      - limit kredit
  - iuran tahunan   - tenggang bebas bunga

Selanjutnya bank tersebut menetapkan bahwa pada masing-masing atribut dapat dibuat 3 level pilihan. Misalkan untuk suku bunga: 2.25%, 2.5%, atau 2.75%.

Selanjutnya responden diminta untuk mengurutkan beberapa konsep kartu kredit berdasarkan preferensinya. Misalkan ranking 1 adalah konsep A: bunga 2.25%, iuran Rp. 180 ribu, limit 18 juta, dan tenggang 40 hari; ranking 2 konsep H, dst.

Selanjutnya dengan analisis konjoin akan didapatkan bobot kepentingan dari masing-masing atribut. Misalkan bobot suku bunga adalah 43% dan bobot iuran tahunan 8%. Juga dapat diestimasi nilai utilitas dari masing-masing level, misal untuk suku bunga 2.75% utilitasnya - 3.32.

Dari sini akan dapat diramalkan bagaimana penerimaan konsumen atas suatu konsep produk Z, misalnya. Juga dapat dihitung *market share*-nya.
- **Isu-isu dalam analisis konjoin**
  - teknik pengumpulan data: metode *full-profile* vs metode *trade-off*
  - teknik-teknik estimasi bobot dan nilai level
- **Contoh penerapan:**
  - pengembangan produk baru
  - meramalkan pangsa pasar suatu produk

## MULTIDIMENSIONAL SCALING

- **Tujuan**  
Mengidentifikasi dimensi-dimensi apa saja yang digunakan konsumen untuk mempersepsikan suatu produk dan selanjutnya menentukan posisi suatu produk berdasarkan tiap dimensi tersebut serta pada akhirnya menentukan strategi positioning berdasarkan peta posisi dari produk-produk yang sudah ada.
- **Ide dasar**  
Dalam benak konsumen terdapat sebuah peta perseptual di mana setiap produk memiliki tempat tertentu dalam peta tersebut. Peta perseptual tersebut adalah multidimensi (dua atau lebih dimensi). Dengan menginputkan persepsi konsumen atas kemiripan pasangan-pasangan (konsep) produk, maka peta perseptual ini dapat diperkirakan.
- **Ilustrasi**  
Seorang manajer perusahaan akan meluncurkan sebuah minuman botol. Untuk itu ia ingin melihat ‘peta’ posisi dari produk minuman botol yang sudah ada di pasaran. Mula-mula dikumpulkan data berupa nilai kemiripan dari beberapa pasang minuman botol. Misalkan: Coke vs Pepsi, Sprite vs Pepsi, Coke vs Sosro, Pepsi vs Sosro, dst.. Selanjutnya, dengan menggunakan teknik MDS, didapatkan peta 2 dimensi yang menggambarkan posisi relatif dari tiap minuman botol yang dievaluasi. Selanjutnya si Manajer memperkirakan *apa sebenarnya* dua dimensi tersebut. Juga si manajer dapat ‘menemukan’ tempat yang masih kosong dalam peta perseptual tersebut yang ia rencanakan akan isi dengan produk baru perusahaannya.
- **Isu-isu dalam MDS**
  - penilaian secara *overall* atau per atribut
  - metode agregasi
  - jumlah dimensi
- **Contoh penerapan:** strategi pemasaran

## ANALISIS DISKRIMINAN

- **Tujuan:**  
Menentukan apakah terdapat kelompok-kelompok objek yang berbeda dan selanjutnya menentukan variabel-variabel apa sajakah yang dapat membedakan kelompok-kelompok objek tersebut serta menentukan suatu prosedur untuk meramalkan akan masuk ke kelompok manakah sebuah obyek pengamatan.
- **Ide dasar:**  
Setiap objek dikelompokkan dalam beberapa kategori (misalnya: membeli vs tidak membeli; pertumbuhan tinggi, sedang, stagnan, menurun; dll). Selanjutnya terdapat pula data berkaitan dengan nilai variabel-variabel dari masing-masing objek (misal: pendapatan, ukuran keluarga, tingkat pendidikan, dll). Analisis diskriminan akan mencoba menentukan variabel manakah yang akan membedakan suatu objek sehingga ia masuk dalam kategori membeli sedangkan objek yang lain masuk dalam kategori tidak membeli.
- **Ilustrasi:**
  - Seorang pegawai Deperindag ingin membuat kebijakan yang mendukung perkembangan SME di Kota Bandung. Untuk itu ia ingin mengetahui faktor apa sajakah yang mempengaruhi kemajuan sebuah SME.

- Pada tahap awal ia mengumpulkan data dari 30 SME di Kota Bandung. Dari tiap SME dikumpulkan data berupa nilai kompetensi perusahaan yang berkaitan dengan faktor-faktor SDM, keuangan, teknologi produksi, logistik, dan lain-lain.
- Selanjutnya berdasarkan kinerja keuangan dan pertumbuhan, ia kelompokkan ke-30 SME tersebut ke dalam 3 kelompok performansi: tinggi, sedang, dan rendah. .
- Berikutnya, ia menggunakan analisis diskriminan dan disimpulkan bahwa variabel yang mempengaruhi performansi sebuah SME antara lain akses informasi pasar, penguasaan teknologi produk, dan kekuatan modal.
- Berdasarkan hasil studi ini, ia mengusulkan kepada Pemda untuk membuat kantor pemasaran bersama, training-training untuk meningkatkan penguasaan teknologi, serta pameran yang akan mempertemukan SME dengan calon pembeli dan pemodal.
- **Isu-isu dalam Analisis Diskriminan**
  - asumsi normal
  - asumsi tidak terdapat multikolinearitas
  - antar kelompok *mutually exclusive*
  - matriks variansi-kovariansi pada masing-masing kelompok identik
  - ukuran sampel
  - validas (*holdout sample*)
- **Contoh Penerapan:**
  - persetujuan aplikasi kredit
  - penentuan kebijakan
  - proyeksi/peramalan

## KASUS

### IDENTIFIKASI SEGMENT DAN PREFERENSI KONSUMEN BENGKEL MOTOR DI BANDUNG

Sepeda motor merupakan salah satu alat transportasi yang populer di Indonesia. Setiap sepeda motor tentu saja memerlukan perawatan berkala dan juga sewaktu-waktu harus diganti suku cadangnya. Perawatan dan penggantian suku cadang ini pada umumnya dilakukan di bengkel. Di kota Bandung terdapat banyak sekali bengkel motor, mulai dari bengkel resmi suatu merk motor sampai bengkel yang bersifat umum untuk semua jenis merk. Walaupun secara sekilas bengkel resmi tampaknya akan lebih diminati konsumen karena adanya jaminan kualitas servis, tetapi kenyataan di lapangan menunjukkan bahwa terdapat pula bengkel resmi yang usahanya lesu. Di sisi lain tampak pula beberapa bengkel umum yang maju usahanya.

Dari fenomena ini kemudian muncul pertanyaan yang ingin dicoba dijawab melalui penelitian ini yaitu:

- Hal-hal apa saja yang menjadi pertimbangan konsumen dalam memilih bengkel untuk merawat dan/atau memperbaiki sepeda motornya?
- Bagaimana segmentasi konsumen bengkel sepeda motor ini?

STRUCTURAL EQUATION MODELING (SEM)

PENGANTAR

- Sampai dengan saat ini, teknik-teknik yang dipelajari hanya dapat meneliti hubungan tunggal dalam satu waktu.
- *Structural Equation Modeling* (SEM) merupakan suatu teknik analisis multivariat yang dapat digunakan untuk menganalisis beberapa hubungan dependen sekaligus secara simultan.
- Analisis secara simultan ini dilakukan agar didapatkan efisiensi secara statistikal (*statistical efficiency*).

$$\hat{X}_1, \dots, \hat{X}_5 = \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = \frac{0,05}{5} \quad \text{Jika secara simultan, untuk memperoleh probabilitas yang sama utk tiap persamaan } (\alpha)$$

$$\hat{X}_1, \dots, \hat{X}_5 = \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 0,05 \quad \text{Jika secara tunggal}$$

- Contoh persamaan-persamaan dalam SEM adalah sebagai berikut :
 

$Y_1$	=	$X_1 + X_2$
$Y_2$	=	$X_1 + X_2$
$Y_3$	=	$Y_1 + Y_3$
(metrik)		(metrik, nonmetrik)
- Pada perkembangannya model-model SEM menjadi sangat bervariasi, sehingga terjadi kekaburan apakah suatu model termasuk dalam model SEM atau tidak
- Secara umum terdapat dua karakteristik dasar model SEM :
  1. estimasi hubungan dependen saling mempengaruhi secara majemuk
  2. kemampuan untuk menyatakan konsep yang tidak teramati (variabel laten) dan memperhitungkan nilai eror pengukuran dalam proses estimasi.

Variabel Latent

- Tinjauan variabel laten (yang diamati) vs variabel manifest (yang muncul/terlihat), yaitu mengukur variabel laten melalui variabel manifes harus dengan mengukur reliabilitas (ketepatan alat ukur).
- Konsep reliabilitas : ukuran bagaimana suatu variabel independen bersifat “bebas error”
- Reliabilitas vs validitas
- Pengaruh eror pengukuran (*measurement error*) :
 
$$\beta_{y|x} = \beta_s \times \rho_x$$
 dengan  $\beta_{y|x}$  adalah koefisien regresi yang teramati  
 $\beta_s$  adalah koefisien regresi sebenarnya  
 $\rho_x$  adalah reliabilitas dari variabel prediktor
- Dalam SEM dikenal adanya **measurement model**, yaitu bagaimana variabel laten disusun oleh variabel-variabel manifes → konfirmatori analisis faktor
- Measurement model menspesifikasikan hubungan antara variabel laten dengan variabel manifes.
- Konsep independensi/dependensi tunggal dapat diwakili oleh beberapa variabel (pertanyaan)
- Selanjutnya dari variabel (pertanyaan) dapat ditetapkan nilai dari konsep independen/dependen tersebut (misalnya penggunaan *summated scale* atau skala campuran).

**Contoh SEM Sederhana**

Suatu perusahaan menginginkan peningkatan retensi dan produktivitas pegawainya. Upaya ini dapat dilakukan dengan pemahaman yang lebih baik atas motivasi dan attitude pegawai terhadap perusahaan. Peneliti telah mengidentifikasi 3 attitude yang dirasakan paling penting :

1. job satisfaction
2. organizational commitment
3. probability of employee turnover

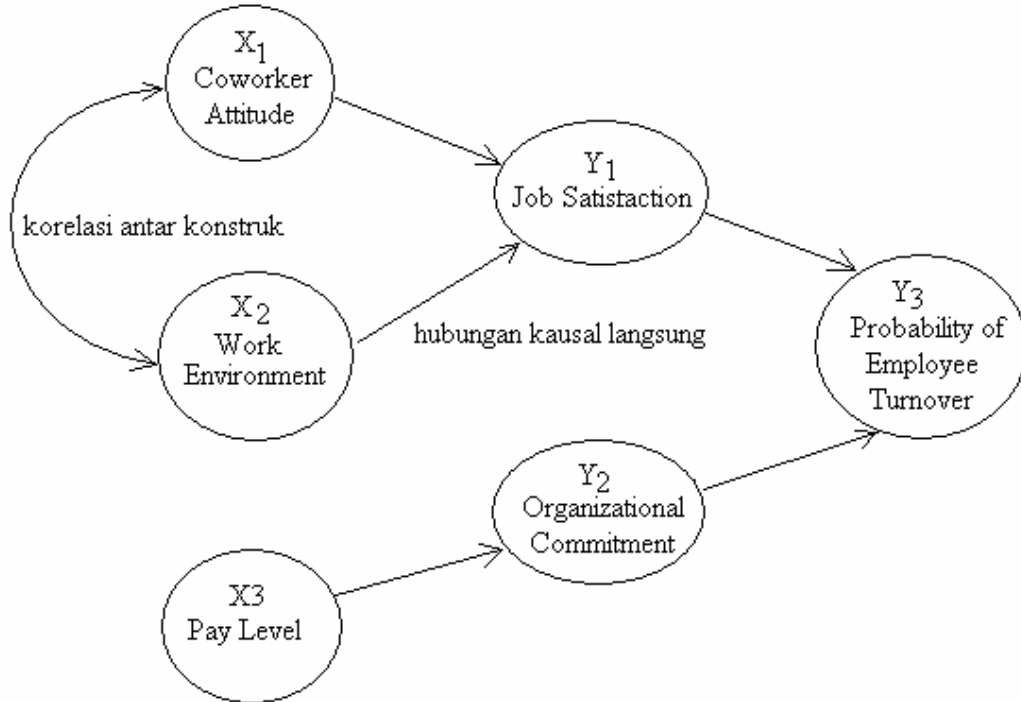
Ketiga attitude ini dipengaruhi oleh 3 hal :

1. Coworker attitude
2. Work environment
3. Pay level

Hubungan antar tiap attitude faktor pengaruhnya dapat dituliskan dalam hubungan dependen (structural model) sebagai berikut :

$$\begin{aligned} \text{Job sat.} &= \text{co. att} + \text{work env.} \\ \text{Org. comm.} &= \text{job sat.} + \text{pay level} \\ \text{Prob. turn.} &= \text{job sat.} + \text{org. comm.} \end{aligned}$$

Persamaan-persamaan ini dapat digambarkan dalam sebuah **path diagram**

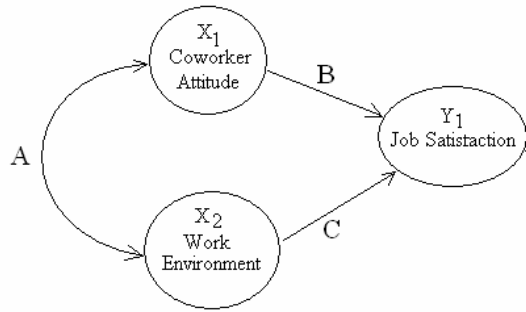


Structural Model (Path Diagram)

**Path Diagram**

- Elemen dalam path diagram disebut dengan konstruk (*construct*) yaitu suatu konsep yang memiliki dasar teoritis dan bertindak sebagai *building block* yang mendefinisikan hubungan-hubungan
- Panah lurus dengan satu kepala panah menunjukkan hubungan kausal langsung antar konstruk
- Panah lengkung menunjukkan korelasi (yang belum tentu ada sebab akibat) antar konstruk (interdependen/autokorelasi).
- Panah lurus dengan dua kepala panah menunjukkan hubungan **nonrecursive** atau **reciprocal**

**Contoh Aplikasi**



Korelasi Bivariat

	X <sub>1</sub>	X <sub>2</sub>	Y <sub>1</sub>
X <sub>1</sub>	1.0		
X <sub>2</sub>	0.5	1.0	
Y <sub>1</sub>	0.6	0.7	1.0

**Korelasi sebagai Compound Paths**

- Korelasi<sub>x<sub>1</sub>,x<sub>2</sub></sub> = A
- Korelasi<sub>x<sub>1</sub>,y<sub>1</sub></sub> = B + AC
- Korelasi<sub>x<sub>2</sub>,y<sub>1</sub></sub> = C + AB

Penyelesaian untuk mencari koefisien struktural

A = 0.5

B + AC = 0,6 → B + 0.5 C = 0.6

C + AB = 0.7 → C + 0.5 B = 0.7

Dari persamaan kedua dan ketiga akan didapat :

B = 0.33

C = 0.53

**Peran Teori dalam SEM**

- Teori dapat didefinisikan sebagai : “sekumpulan hubungan sistematis yang memberikan penjelasan konsisten dan komprehensif dari suatu fenomena/kasus”
- Hubungan (path) dalam SEM sepenuhnya dapat dispesifikasikan oleh peneliti sehingga dalam hal ini sangat penting struktur model didukung oleh teori yang kuat
- Dalam SEM sangat mungkin terjadi fenomena overfitting

**Strategi Pemodelan**

- Karena SEM sangat fleksibel, berbagai model dapat dihasilkan sebagai output analisis. Karena itu diperlukan strategi pemodelan.
- Terdapat 3 strategi umum :
  1. strategi pemodelan konfirmatori
    - Di sini peneliti menspesifikasikan sebuah model dan SEM digunakan untuk mengakses signifikansi statistik (apakah bekerja atau tidak).
    - Kelemahan : confirmation bias (tidak semu).  
SEM cenderung mengkonfirmasi walaupun hasilnya belum tentu baik.
    - Konsekuensi : peneliti belum “membuktikan” model yang diajukan tetapi hanya menetapkan bahwa strategi ini merupakan salah satu dari beberapa model yang dapat diterima.
  2. strategi model bersaing
    - di sini peneliti membandingkan beberapa model sekaligus dan memilih mana yang paling “kuat”

- salah satu contoh adalah program TETRAD yang menggunakan input matrik korelasi dan model inisial. Selanjutnya program akan mencari pola-pola hubungan dan selanjutnya mengisolasi pola yang mendapat “dukungan empiris”
  - kelemahan : dianggap terlalu black box (tidak jelas proses nya).
3. strategi pengembangan model
- didasarkan pada premis bahwa “teori hanya dapat memberikan titik awal bagi pengembangan model sah secara teoritis yang dapat didukung secara empirik”
  - Jadi, SEM digunakan untuk mendapatkan *insight* untuk selanjutnya dilakukan respisifikasi model.

### LANGKAH-LANGKAH DALAM SEM :

#### 1. Mengembangkan model berdasarkan teori

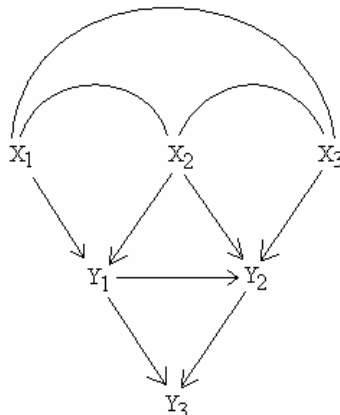
- SEM didasarkan pada **hubungan kausal**
- Penentuan ada tidaknya hubungan sebab-akibat antar dua variabel tidak terletak pada hasil perhitungan analitis tetapi pada **justifikasi teoritis** yang mendukung analisis tersebut.
- Terdapat empat kriteria untuk menyatakan adanya suatu hubungan sebab-akibat :
  1. Hubungan (pengaruh) yang cukup diantara dua variabel
  2. Pengaruh sementara dari sebab dan akibat.
  3. Kurangnya variabel-variabel sebab alternatif.
  4. Dasar teoritis untuk hubungan tersebut.
- Salah satu kesalahan umum dalam pengembangan teori berbasis model adalah **spesifikasi kesalahan**.

#### 2. Membuat path diagram

- Istilah dan cara membuat path diagram telah dibahas pada pertemuan terdahulu.
- Sebagai tambahan, konstruk (*constructs*) dapat dibagi dalam 2 kelas :
  - Konstruk eksogen : dikenal pula sebagai variabel independen dan tidak terdapat panah dengan kepala panah tunggal yang masuk ke construct ini
  - Konstruk endogen : dapat diprediksi oleh satu atau lebih construct yang lain
- Terdapat dua asumsi dalam menyusun path diagram :
  - Semua hubungan adalah *causal relationship* (hubungan sebab akibat)
  - Semua hubungan adalah *linear*

#### 3. Mengkonversi path diagram

- Sebelum masuk dalam spesifikasi model, peneliti perlu mendefinisikan :
  1. **Structural equations** (structural model) → menghubungkan constructs
    - Structural model mendefinisikan hubungan antara konstruk endogen dengan konstruk eksogen atau konstruk endogen



- Dari path diagram tersebut dapat disusun persamaan model struktural sebagai berikut :

$$\begin{array}{rclclcl}
 \text{Var. endogen} & = & \text{Var. eksogen} & + & \text{Var. endogen} & + & \text{Error} \\
 Y_1 & = & b_1X_1 + b_2X_2 & & & & + \epsilon_1 \\
 Y_2 & = & & b_3X_2 + b_4X_3 & + b_5Y_1 & + b_6Y_3 & + \epsilon_2 \\
 Y_3 & = & & & + b_7Y_1 + b_8Y_2 & & + \epsilon_3
 \end{array}$$

**2. Measurement model**

- menentukan suatu variabel apa saja yang menjadi ukuran suatu konstruk
- Measurement model adalah model yang digunakan untuk mendefinisikan konstruk berdasarkan variabel atau indikator tertentu
- Setiap konstruk “dibangun” oleh indikator/variabel tertentu. Dalam hal ini terdapat *indicator loadings on constructs*
- Rule of thumb : 5 – 7 / indicator construct
- Measurement model vs. analisis faktor

Variabel	Analisis Faktor			Measurement Model		
	Faktor 1	Faktor 2	Faktor 3	Construct A	Construct B	Construct C
V <sub>1</sub>	L <sub>11</sub>	L <sub>12</sub>	L <sub>13</sub>	L <sub>1</sub>		
V <sub>2</sub>	L <sub>21</sub>	L <sub>22</sub>	L <sub>23</sub>	L <sub>2</sub>		
V <sub>3</sub>	L <sub>31</sub>	L <sub>32</sub>	L <sub>33</sub>		L <sub>3</sub>	
V <sub>4</sub>	L <sub>41</sub>	L <sub>42</sub>	L <sub>43</sub>		L <sub>4</sub>	
V <sub>5</sub>	L <sub>51</sub>	L <sub>52</sub>	L <sub>53</sub>			L <sub>5</sub>

**3. Kelompok matriks**

- Mengindikasikan semua hubungan hipotetik diantara konstruk suatu variabel

- Isu-isu penting :
  - Berapa jumlah indikator per konstruk ?
  - Isu reliabilitas
  - Two step vs single step

**4. Memilih jenis matrix input dan meninjau masalah riset**

- Asumsi model :
  - Observasi bersifat independen
  - Sampel nya acak (*random sampling*)
  - Semua hubungan bersifat linier
  - Berdistribusi normal (*nonskewed distribution*)
- Missing data :
  - Missing data dapat mengakibatkan efek mendalam pada SEM
  - Isu missing data cukup kompleks, namun untuk kepraktisan sebaiknya input SEM digunakan data yang lengkap. Perlakuan terhadap missing data dapat dilakukan pada *data screening (preanalysis)* sebagaimana telah dijelaskan pada perkuliahan minggu ke-2.
- Matriks kovariansi vs korelasi
  - Kovariansi : memberikan perbandingan yang valid antar populasi yang berbeda tetapi kesulitan dalam interpretasi jika unit pengukurannya berbeda

- Korelasi : sesuai digunakan jika tujuan penelitian hanya untuk mengerti pola hubungan antar konstruk tetapi tidak untuk menjelaskan total variansi dari suatu konstruk.
- Jika peneliti ingin menguji suatu teori maka harus digunakan input kovariansi
- Ukuran sampel :
  - Terdapat empat faktor yang mempengaruhi ukuran sampel :
  - 1. Model misspecification  
Ukuran sampel harus diperbesar jika terdapat kemungkinan misspecification
  - 2. Model size  
Minimum rasio adalah 5 responden per parameter yang diestimasi. Kalau memungkinkan rasio ini adalah 10 : 1
  - 3. Tingkat kejauhan dari normalitas  
Jika data menjauhi normal, rasio harus ditingkatkan menjadi 15 : 1
  - 4. Prosedur estimasi  
Metode estimasi menggunakan maximum likelihood membutuhkan sampel minimum 100. namun ukuran yang terlalu besar menyebabkan model terlalu sensitif.
- Rekomendasi ukuran sampel :  
Sebagai patokan, pada umumnya ukuran sampel 200 buah sudah mencukupi. Jika ada kemungkinan misspesifikasi, hubungan cukup kompleks, atau non normalitas, jumlah ini dapat ditambah.
- Metode dan proses :
  - Metode yang umum : maximum likelihood
  - Proses :
    1. direct estimation
    2. bootstrapping
    3. simulation
    4. jackknifing
- 5. Membuat identifikasi model, menguji estimasi dan kecocokan model, dan interpretasi model
  - Dalam SEM dikenal *identification problem* yaitu ketidakmampuan model yang diajukan untuk membuat estimasi yang unik.
  - Permasalahan ini terkait dengan *degree of freedom* dalam SEM :  

$$d.o.f = \frac{1}{2} [(p+q)(p+q+1)] - t$$

p : jumlah indikator endogen  
 q : jumlah indikator eksogen  
 t : jumlah koefisien yang diestimasi dalam model yang diusulkan
  - Just-identified terjadi jika  $d.o.f = 0 \rightarrow$  tidak terdapat generalizability
  - Under-identified terjadi jika  $d.o.f < 0$
  - Over-identified terjadi jika  $d.o.f > 0 \rightarrow$  tujuan yang harus dicapai
  - Untuk menyelesaikan permasalahan identifikasi dapat digunakan pedoman berikut :
    - Buatlah model teoritik dengan jumlah koefisien yang minimum
    - Tentukan variansi kesalahan pengukuran (*measurement error*) dari konstruk jika memungkinkan
    - Tentukan sembarang koefisien struktural yang telah diketahui dapat digunakan..
    - Hilangkan variabel-variabel yang memperumit masalah.

6. Modifikasi model → mengevaluasi kriteria *goodness-of-fit* (kebaikan data)

- Sebelum mengevaluasi model lebih lanjut, peneliti harus mengamati adanya **offending estimates** (perkiraan yang mengganggu/merusak) untuk kemudian memperbaikinya terlebih dahulu.
- Offending estimates meliputi :
  - Variansi error negatif
  - Standardized coefficient melewati atau sangat dekat dengan 1
  - Standard error untuk suatu nilai yang diestimasi menjadi sangat besarBiasanya perbaikan terhadap salah satu offending estimates dapat memperbaiki yang lain, sehingga jangan semua dulu diperbaiki.
- Terdapat 3 ukuran *goodness-of-fit* :
  1. absolute fit measures : hanya menaksir kecocokan model secara keseluruhan, tidak terdapat *adjustment* terhadap kemungkinan adanya overfitting
  2. incremental fit measures : membandingkan model yang diusulkan dengan model lain yang telah dispesifikasikan oleh peneliti.
  3. parsimonious fit measures : meng-*adjust* ukuran fitness model dengan jumlah koefisien yang diestimasi
- *Measurement model fit* dapat diukur dengan reliabilitas indikator-indikator dalam menyusun konstruk (misalnya *Alpha's Cronbach*). Dalam melakukan uji Alpha indikator-indikatornya harus unidimensional, sehingga harus diuji unidimensionalitas nya.
- *Structural model fit* dapat dilihat dari signifikansi koefisien-koefisien yang diestimasi

7. Interpretasi dan modifikasi model

- Standardized vs unstandardized coefficients
- Model respecification : proses menambahkan atau mengurangi parameter yang diestimasi dari model awal (orisinal)
- Respesifikasi model hanya boleh dilakukan dalam kerangka teoritis dan model yang baru harus divalidasi silang

**TUGAS BACA**

- Baca dua ilustrasi penggunaan SEM dalam buku Hair, et.al. (2002) halaman 616 s.d. 643
- Pada pertemuan berikutnya akan diadakan quiz berdasarkan dua ilustrasi tersebut.