

ANALISIS MULTIVARIAT

Pertemuan 1

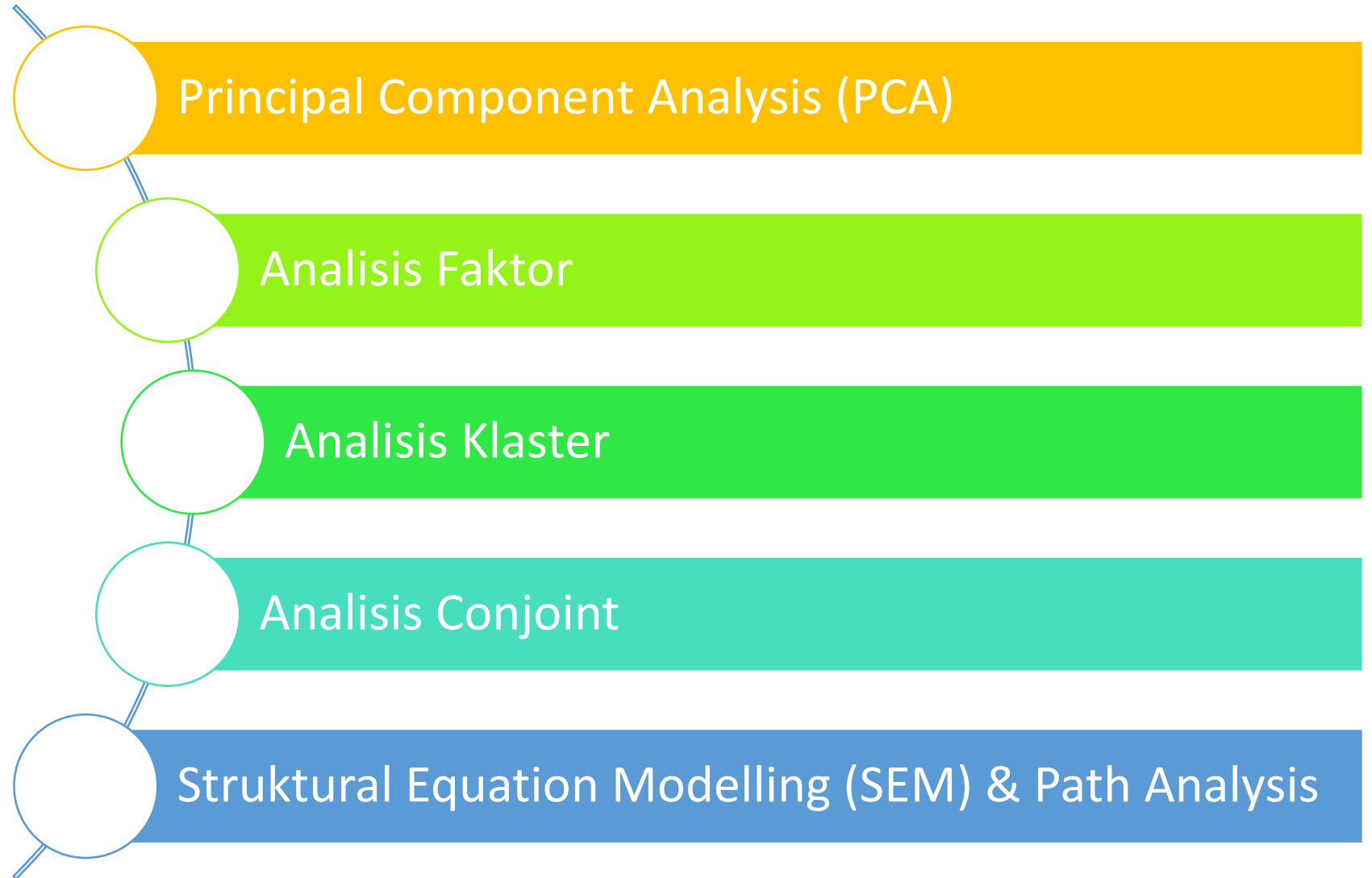
Wahyu Sri Utami,S.Si.,M.Sc.

Program Studi Sains Data
Universitas Teknologi Yogyakarta

Materi Kuliah



Materi Kuliah



Istilah Penting Dalam Analisis Data

- **Elemen/unit sampling** → sesuatu yang menjadi objek penelitian

Ex. Orang (pelanggan, staf, turis,dsb); Unit Organisasi (negara, departemen, Perguruan tinggi, pasar, Perusahaan,dsb); benda , dsb.

- **Karakteristik/Atribut** → hal-hal yang dimiliki oleh elemen
- **Variabel** → sesuatu yang nilainya berubah-ubah menurut waktu atau berbeda menurut elemen/tempat. Umumnya Karakteristik merupakan variable.

Ex. **Nasabah Bank** memiliki variable: jumlah tabungan, gaji, usia, tingkat kepuasan, sikap, motivasi, loyalitas

Perusahaan memiliki variable: jumlah modal, jumlah karyawan, jumlah produksi, jumlah penjualan, jumlah pajak yang dibayar, jumlah sumbangan sosial,dsb.

- **Populasi** → Kumpulan yang lengkap dari keseluruhan elemen yang sejenis namun dapat dibedakan berdasarkan karakteristiknya.
- **Sampel** → Bagian dari populasi
- **Sensus** → cara pengambilan data untuk menghasilkan data parameter
- **Sampling** → metode pengumpulan sampel untuk menghasilkan data perkiraan atau estimator
- **Sampling Error** → kesalahan yang terjadi pada data estimator yang disebabkan oleh penelitian yang tidak menyeluruh. Berguna untuk mengukur tingkat ketelitian data estimator.

Segi	Survei Sampel	Sensus
Tenaga	<ul style="list-style-type: none"> Jumlah relatif sedikit Dapat dipilih yang berkualitas 	<ul style="list-style-type: none"> Jumlah sangat besar Lebih sulit untuk memilih yang berkualitas seluruhnya
Waktu	<ul style="list-style-type: none"> Lebih cepat 	<ul style="list-style-type: none"> Lebih lama
Biaya	<ul style="list-style-type: none"> Lebih murah 	<ul style="list-style-type: none"> Lebih mahal
Kedalaman dan kualitas data	<ul style="list-style-type: none"> Biasanya kualitas data lebih baik Pertanyaan yang lebih sulit bisa dipergunakan 	<ul style="list-style-type: none"> Kualitas data kurang baik, hal ini akibat dari kualitas tenaga pengumpul Pertanyaan sederhana
Penyajian Data	<ul style="list-style-type: none"> Data tidak bisa disajikan sampai ke tingkat yang paling rendah 	<ul style="list-style-type: none"> Data bisa disajikan sampai ke tingkat yang paling rendah, karena semua unit dalam populasi dikumpulkan
Kesalahan (Error)	<ul style="list-style-type: none"> Adanya kesalahan sampel Adanya kesalahan bukan dari sampel, namun relatif kecil 	<ul style="list-style-type: none"> Tidak ada kesalahan sampel Adanya kesalahan bukan dari sampel yang besar

Pengukuran

- Pengukuran adalah pemberian nilai/angka dengan aturan tertentu pada atribut suatu elemen.
- Contoh: Johnny diminta untuk memberikan penilaian tentang kepuasan terhadap layanan suatu Restoran dengan Skala Likert 5 butir

5 : sangat puas

4 : puas

3 : netral

2 : tidak puas

1 : sangat tidak puas

Penskalaan (Scalling)

- Penskalaan adalah penempatan nilai dalam suatu garis bilangan untuk memudahkan dalam melakukan perbandingan.
- Dapat dilakukan dengan alat ukur seperti timbangan, meteran, thermometer, dsb. Selain itu, juga dapat digunakan garis bilangan untuk mengurutkan skala Likert.

- Alat ukur dikatakan **Valid** jika mampu mengukur apa yang seharusnya diukur.
- Timbangan yang rusak, pertanyaan yang tidak jelas adalah contoh alat ukur yang tidak valid.
- Alat ukur dikatakan **Handal (Reliabel)** jika digunakan berkali-kali dalam kondisi yang bervariasi memberikan hasil yang sama.

Tipe Data

- **Nominal** : tidak bisa diurutkan, tidak dapat dilakukan operasi hitung

Ex. Jenis kelamin, agama, suku bangsa, Alamat, jawaban Ya/Tidak, dsb

- **Ordinal** : dapat diurutkan, tidak dapat dilakukan operasi hitung

Ex. Skala Likert, Nilai Akhir kuliah, ranking kelas, dsb.

Kedua Tipe Data ini dikatakan sebagai skala Non Metrik / Kategorik

Tipe Data

- **Interval** : bisa diurutkan, dapat dilakukan operasi hitung, Titik Nol tidak tetap

Ex. Rentang suhu, Rentang Nilai, dsb

- **Rasio** : dapat diurutkan, dapat dilakukan operasi hitung, Mempunyai titik nol tetap

Ex. Gaji, Nilai, BB, TB, Harga barang , dsb.

Kedua Tipe Data ini dikatakan sebagai skala Metrik

Analisis Univariat

- Misalkan Nasabah suatu bank ditanya tentang Jumlah tabungannya, penghasilan per bulan, umur, tingkat pendidikan, dan jumlah anggota keluarga yang ditanggung tetapi tidak dikaitkan dengan variable lain maka hal ini merupakan variable Tunggal (univariat)
- Variabel Univariat hanya dapat dilakukan analisis untuk melihat gambaran tentang jumlah, rata-rata, persentase, modus, Quartil, Variansi, dsb.
- Selain itu, analisis univariat dapat digunakan untuk melihat distribusi dari data tersebut.

- **Contoh Analisis Univariat:**

1. Uji Z
2. Uji t-student
3. Uji paired t-test, dsb

Analisis Bivariat

- Analisis bivariat melibatkan dua variable yaitu X dan Y.
- Analisis bivariat dapat digunakan untuk melihat keeratan hubungan dari variable X dan Y
- Selain itu, analisis bivariat juga dapat digunakan untuk mengetahui hubungan/pengaruh variable X terhadap variable Y, serta dapat digunakan untuk meramalkan nilai Y di masa depan dengan nilai X tertentu.

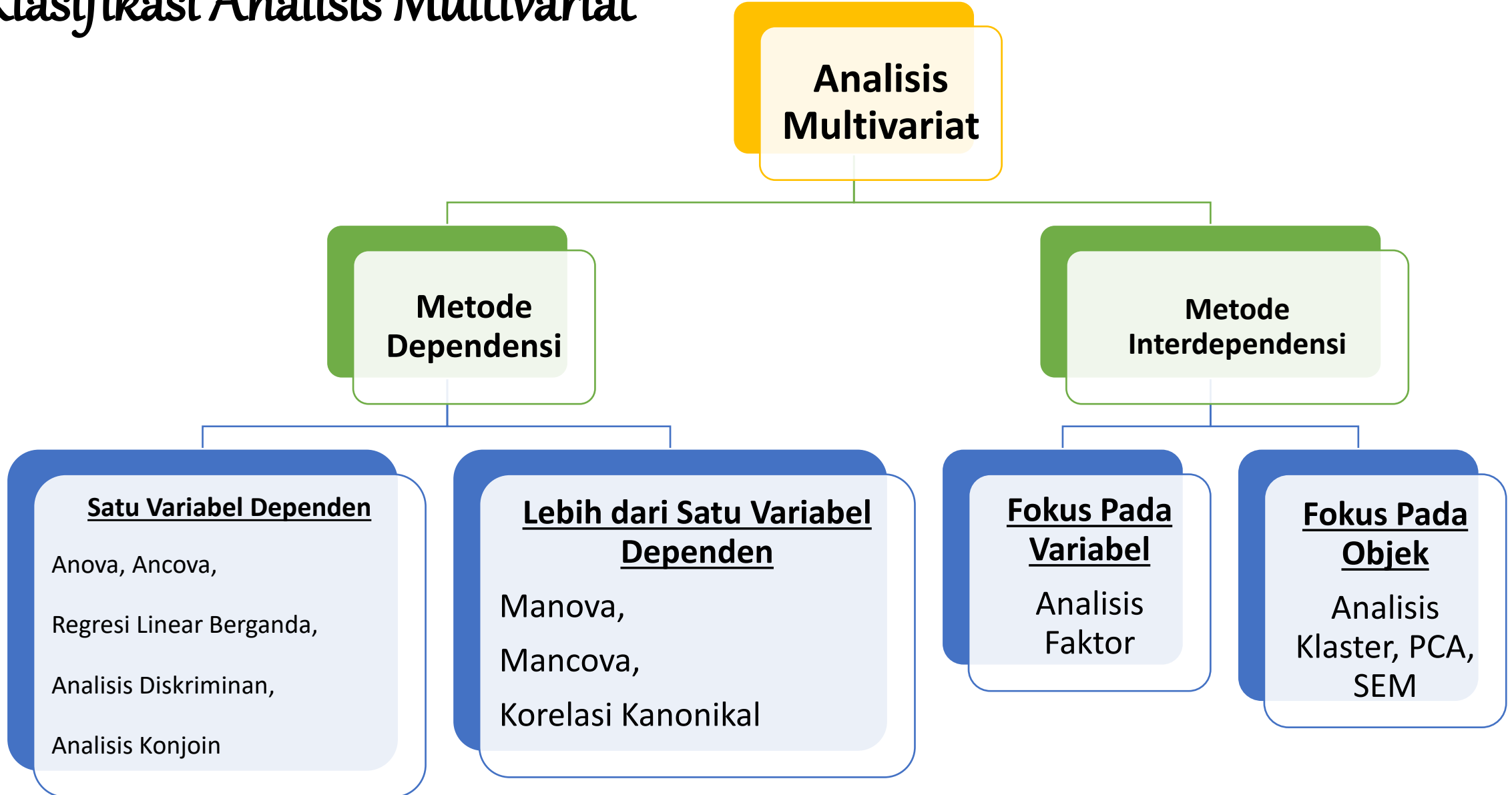
- **Contoh Analisis Bivariat:**

1. Korelasi (r)
2. Regresi Linear Sederhana: $Y = a + bX + \varepsilon$
3. Koefisien Determinasi (r^2)

Analisis Multivariat

- Dalam berbagai kasus seringkali ditemukan penyebabnya terdiri atas lebih dari 1 factor.
- Diberikan Y sebagai variable terikat (dependen) dan $X_1, X_2, X_3, \dots, X_k$ sebagai variable bebas (independent)
- Misal dalam suatu penjualan (Y) terjadi penurunan yang disebabkan karena biaya promosi (X_1), harga jual (X_2), mutu pelayanan (X_3), persaingan produk (X_4), dsb.

Klasifikasi Analisis Multivariat



Analisis Dependensi

- Tujuan:

1. memprediksi nilai variable terikat Y berdasarkan nilai > 2 variable bebas
2. Mengetahui hubungan antara variable Y dengan >2 variable bebas
3. Mengetahui pengaruh nilai variable bebas X_i dengan variable terikat Y

Analisis Interdependensi

- Tujuan:

Mengelompokkan suatu set variable menjadi kelompok yang lebih sedikit jumlahnya dan masing-masing kelompok membentuk variable yang baru yang disebut factor.(Reduksi Dimensi)

Hubungan Antar Metode Dependensi Multivariat

- Misalkan Y adalah variable terikat dan X_i adalah variable bebas.

1. Regresi Linear Berganda

$Y = a_0 + b_1X_1 + b_2X_2 + \cdots + b_nX_n$	
Y berskala metrik	X_i bisa berskala metrik atau nonmetrik

→ Untuk peramalan nilai Y

→ Untuk mengetahui hubungan variable X_i dengan Y

Hubungan Antar Metode Dependensi Multivariat

2. Analisis Diskriminan Berganda

$Y = b_1X_1 + b_2X_2 + \dots + b_nX_n$	
Y berskala nonmetrik	X_i bisa berskala metrik

→ Untuk meramalkan peluang bahwa suatu objek penelitian akan masuk dalam kelompok tertentu (variable terikat) berdasarkan beberapa variable bebas.

→ Peneliti harus menentukan fungsi diskriminan yang dapat membedakan objek tertentu masuk ke kelompok mana.

Hubungan Antar Metode Dependensi Multivariat

3. Analisis Variansi (ANOVA)

$Y = b_1X_1 + b_2X_2 + \cdots + b_nX_n$	
Y berskala metrik	X_i bisa berskala nonmetrik

→ Untuk mengetahui dampak beberapa variable bebas yang non-metrik (treatment) terhadap variable tak bebas yang metrik.

→ Pengaruh yang terjadi ditunjukkan dengan adanya perbedaan rata-rata

Hubungan Antar Metode Dependensi Multivariat

4. Analisis Variansi Multivariat (MANOVA)

$Y_1 + Y_2 + Y_3 + \cdots + Y_n = b_1X_1 + b_2X_2 + \cdots + b_nX_n$	
Y berskala metrik	X_i bisa berskala nonmetrik

→ Untuk mengetahui dampak beberapa variable bebas yang non-metrik (treatment) terhadap beberapa variable tak bebas yang metrik.

→ Pengaruh yang terjadi ditunjukkan dengan adanya perbedaan rata-rata

Hubungan Antar Metode Dependensi Multivariat

5. Korelasi Kanonikal

$Y_1 + Y_2 + Y_3 + \cdots + Y_n = b_1X_1 + b_2X_2 + \cdots + b_nX_n$	
Y berskala metrik atau non-metrik	X_i bisa berskala metrik atau nonmetrik

→ sebagai generalisasi dari regresi linear berganda

→ Untuk mengkorelasikan secara simultan (bersamaan) beberapa variable Y dan beberapa variable X.

→ prinsipnya: mencari bobot untuk variable X dan Y sehingga diperoleh korelasi sederhana yang maksimum (sekuat mungkin) antara suatu set variable Y dan X.

Hubungan Antar Metode Dependensi Multivariat

6. Analisis Conjoint

$Y = b_1X_1 + b_2X_2 + \dots + b_nX_n$	
Y berskala metrik atau nonmetrik	X_i bisa berskala nonmetrik

→ Untuk memberikan suatu ukuran kuantitatif mengenai kepentingan relative suatu atribut terhadap atribut yang lain dari suatu produk.

Hubungan Antar Metode Interdependensi Multivariat

1. Structural Equation Modelling (SEM)

$Y_1 = X_{11} + X_{12} + \cdots + X_{1n}$ $Y_2 = X_{21} + X_{22} + \cdots + X_{2n}$ \vdots $Y_m = X_{m1} + X_{m2} + \cdots + X_{mn}$	
Y_i berskala metrik	X_{ij} bisa berskala metrik atau nonmetrik

→ Untuk mengetahui pola hubungan antara variable dan indikatornya, variable yang satu dengan yang lainnya serta kesalahan pengukuran secara langsung untuk memperoleh gambaran menyeluruh mengenai suatu model.

Hubungan Antar Metode Interdependensi Multivariat

2. Analisis Faktor

- Untuk menemukan variable baru (factor) yang jumlahnya lebih sedikit dibandingkan dengan jumlah variable aslinya yang tidak saling berkorelasi (tidak multicollinearity).
- Faktor tersebut harus mengandung informasi sebanyak mungkin dari variable asli
- Analisis factor terdiri atas PCA dan common factor analysis

Hubungan Antar Metode Interdependensi Multivariat

3. Analisis Klaster

- Untuk mengelompokkan elemen yang mirip sebagai objek penelitian menjadi kluster yang berbeda dan mutually exclusive.

Hubungan Antar Metode Interdependensi Multivariat

4. Penskalaan Multidimensional

- Untuk membentuk penilaian terhadap similaritas atau preferensi ke dalam jarak yang diwakili dalam ruang multidimensional.
- Teknik ini akan memposisikan objek A dan B sedemikian sehingga jarak antar objek dalam ruang multidimensional menjadi lebih pendek daripada jarak dua pasang objek yang lain.

Review

1. Berikan 3 contoh analisis multivariate dalam kehidupan sehari-hari!
2. Misalkan diketahui model bivariate $\hat{Y} = 0,5 + 1,25X$. Jelaskan makna dari model tersebut!
3. Hasil penghitungan koefisien korelasi menunjukkan hasil $r = 0,9$ dan $r^2 = 0,81$. Apa makna angka-angka tersebut?Jelaskan!
4. Mengapa analisis multivariat jauh lebih realistis untuk memecahkan masalah-masalah sosial dibandingkan analisis bivariat? Jelaskan!