

ANALISIS MULTIVARIAT

sesi-1: Simple Linear regression

**PROGRAM STUDI S1 PWK
UNIVERSITAS GADJAH MADA
2005**

Referensi:

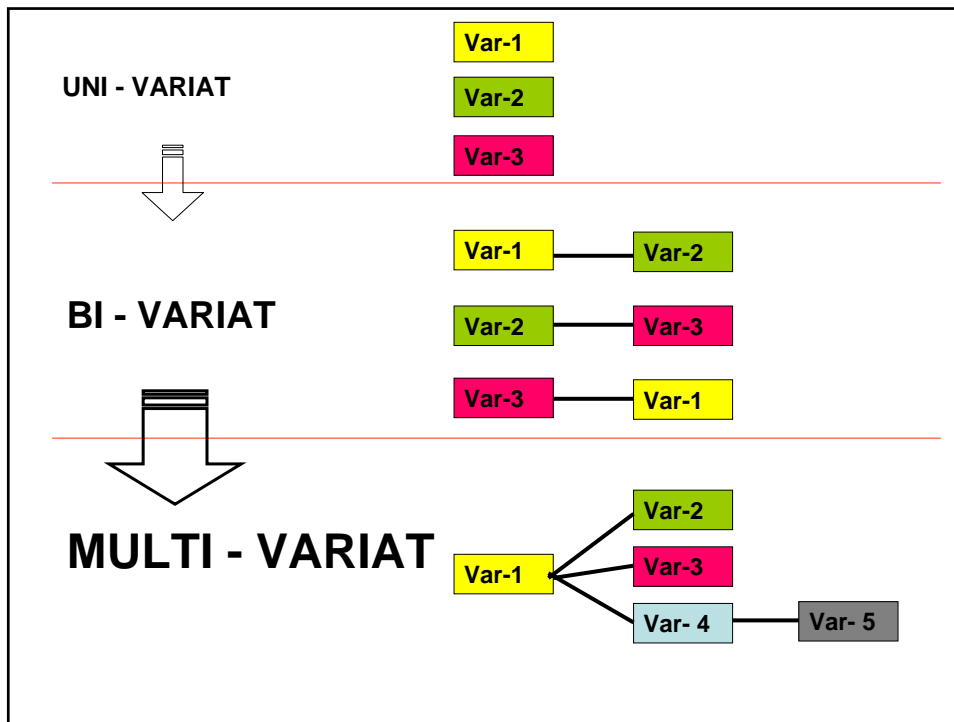
Everitt, B.S., 1996, Making Sense of Statistics in Psychology: A Second-Level Course, Oxford University Press, Oxford.

Kachigan, S.K., 2002, Statistical Analysis: An Interdisciplinary Introduction to Univariate & Multivariate Methods, Radius Press, New York.

Kline P., 1994, An Easy Guide to Factor Analysis, Routledge, London.

Napa J.A., 1995, Metode Statistik dan Ekonometri, Liberty, Yogyakarta.

Supranto J., 2004, Analisis Multivariat: Arti & Interpretasi, Rineka Cipta, Jakarta.



PENGANTAR

MENGAPA MULTIVARIAT?

Karena tidak semua gejala itu hanya didasarkan pada hubungan dua variabel saja.

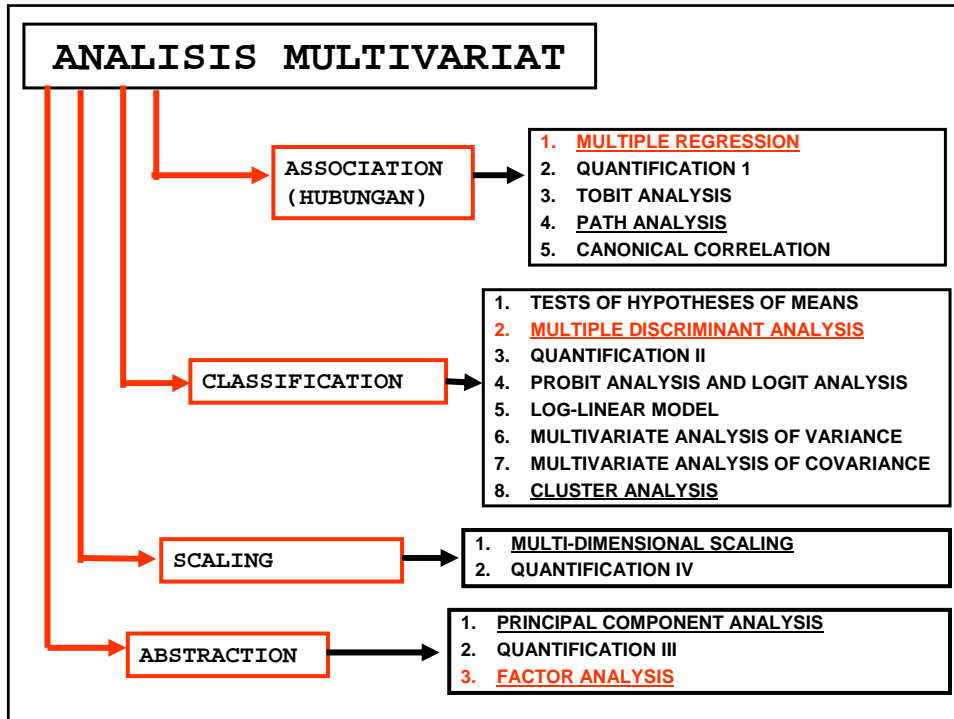
Contoh:

-Harga tanah tidak hanya ditentukan oleh lokasi yang dekat dengan jalan raya, tetapi oleh faktor lain misal dekat kampus, dekat pasar, kesuburan tanah, bentuk persilnya.

-INDEKS PRESTASI tidak hanya ditentukan oleh lama belajar, tetapi juga IQ, EQ, ...



Sehingga diperlukan multivariat → mengkaitkan banyak variabel yang secara logis berkait.



SEBELUM MULTIVARIAT, PERLU MEMAHAMI TERLEBIH DAHULU:

SIMPLE LINEAR REGRESSION

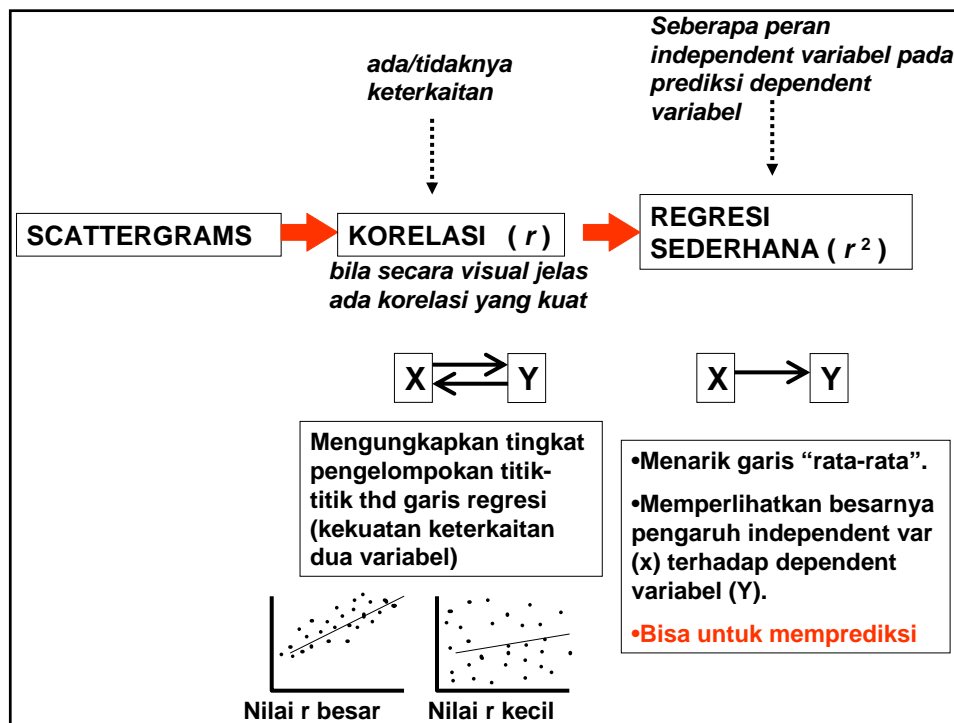
(ini sebenarnya bivariat, tetapi menjadi dasar pemahaman bagi kita tentang multivariat)

Kemudian baru:

MULTIPLE LINEAR REGRESSION

SIMPLE LINEAR REGRESSION (REGRESI SEDERHANA)

- Tujuan: untuk menemukan seberapa besar pengaruh perubahan variabel independen terhadap variabel dependen. Besarnya pengaruh dilihat koefisien regresinya.
- Kedua variabel adalah QUANTITATIVE.
- REGRESI TIDAK DAPAT untuk mengukur KETERKAITAN. Bila korelasi (keterkaitan) kecil maka REGRESI tidak ada artinya. → **KORELASI DIHITUNG DULU BARU REGRESI.**
- Merupakan statistik BIVARIATE dan merupakan cikal bakal MULTIVARIATE.



REGRESI SEDERHANA

Contoh

Adakah dan seberapa kaitan antara jumlah penduduk (X) dengan jumlah mobil pribadi (Y)?

Apabila pada tahun 2010 jumlah penduduk menjadi 1000 orang, berapakah jumlah mobil pada saat itu.

Tahun	Jml Penduduk (X)	Jml mobil (Y)
1980	125	4
1985	345	6
1990	390	12
1995	456	14
2000	564	18
2005	654	24

Hitung dulu Pearson Correlation Coefficientnya (r) :

Dulu dipelajari dengan rumus di bawah ini:

$$r = \frac{n \cdot \sum x_i y_i - (\sum x_i) (\sum y_i)}{\left[\sqrt{n \cdot \sum x_i^2 - (\sum x_i)^2} \right] \left[\sqrt{n \cdot \sum y_i^2 - (\sum y_i)^2} \right]}$$

Rumus di atas sebenarnya adalah Covariance dari x dan y dibagi dengan perkalian standar deviasi x dan standar deviasi y.

$$\text{Jadi } r = \frac{\text{Covariance } xy}{(\text{stand.dev. } X) (\text{stand.dev. } Y)} \quad \text{atau} \quad \frac{S_{xy}}{\sqrt{S_x} \cdot \sqrt{S_y}}$$

Catatan: Covariance adalah fundamental index untuk mengukur besarnya hubungan antara dua variabel kuantitatif.

TANPA MERUBAH FORMAT PERHITUNGAN TABEL, MAKA PROSES MENGHITUNG r BISA DILAKUKAN.

Bila sudah mendapat hasil r maka kita bisa menghitung r^2 (kuadratkan nilai r!)

BUKAN r^2 dulu, kemudian $r = \sqrt{r^2}$; bila hal ini maka tanda + dan - tidak jelas.

Tahun	Jml Penddk (X)	Jml mobil (Y)	X ²	Y ²	X.Y
1980	125	4	15625	16	500
1985	345	6	119025	36	2070
1990	390	12	152100	144	4680
1995	456	14	207936	196	6384
2000	564	18	318096	324	10152
2005	654	24	427716	576	15696
n = 6	2534	78	1240498	1292	39482

$$r = \frac{n \cdot \sum x_i y_i - (\sum x_i) (\sum y_i)}{\sqrt{n \cdot \sum x_i^2 - (\sum x_i)^2} \sqrt{n \cdot \sum y_i^2 - (\sum y_i)^2}}$$



Alternatif lain:

$$r = \frac{S_{xy}}{\sqrt{S_x} \cdot \sqrt{S_y}} = \frac{\sum x \cdot y - (\sum x) (\sum y) / n}{\sqrt{\sum x^2 - (\sum x)^2 / n} \sqrt{\sum y^2 - (\sum y)^2 / n}}$$

Tahun	Jml Penddk (X)	Jml mobil (Y)	X ²	Y ²	X.Y
1980	125	4	15625	16	500
1985	345	6	119025	36	2070
1990	390	12	152100	144	4680
1995	456	14	207936	196	6384
2000	564	18	318096	324	10152
2005	654	24	427716	576	15696
n = 6	2534	78	1240498	1292	39482

$$\bar{x} = 2534 / 6 = 422,33$$

$$\bar{y} = 78/6 = 13$$

$$S_{xy} = \sum x \cdot y - (\sum x) (\sum y) / n = 39482 - (2534)(78) / 6 = 6540$$

$$S_x = \sum x^2 - (\sum x)^2 / n = 1240498 - (2534)^2 / 6 = 170305$$

$$S_y = \sum y^2 - (\sum y)^2 / n = 1292 - (78)^2 / 6 = 278$$

$$r = \frac{S_{xy}}{\sqrt{S_x} \cdot \sqrt{S_y}} = \frac{6540}{\sqrt{170305} \cdot \sqrt{278}}$$

$r = 0,950476$ cek dengan SPSS

$r^2 = 0,903405$ dianggap kuat keterkaitannya

•Tahap perhitungan r dan r² sudah dilakukan di atas.

•Lalu bagaimana bila kita akan memprediksinya? Misal tahun 2010 jumlah penduduknya jadi 1000 jiwa. Lalu berapa jumlah mobilnya?

Rumus regresi sederhana $Y = a + b X$

$$b = S_{xy} / S_x$$

$$= 6540 / 170305$$

$$= 0,038402$$

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

$$= 13 - 0,038402 \cdot 422,333$$

$$= - 3,218314$$

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

Y_i : variabel dependen ke- i

X_i : variabel independen ke- i

ε_i : variabel pengganggu ke- i

β_0 : konstansta

β_1 : koefisien regresi

Jadi, persamaan garis regresinya =

$$Y = - 3,218 + 0,038 x$$

Selanjutnya, jumlah mobil tahun 2010 bila jml penduduk 1000 jiwa =

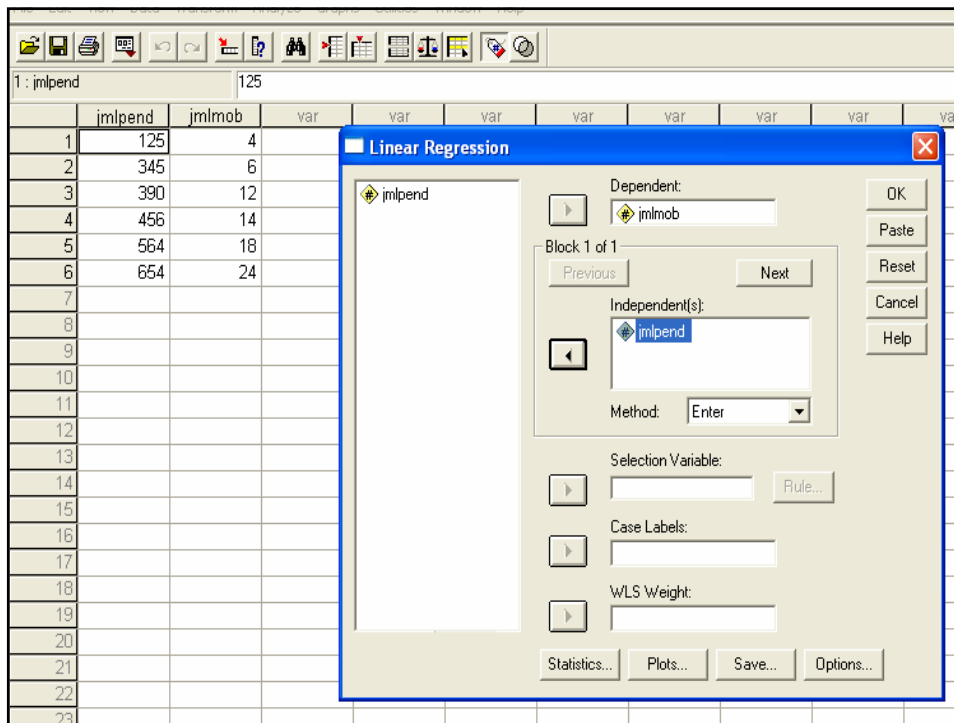
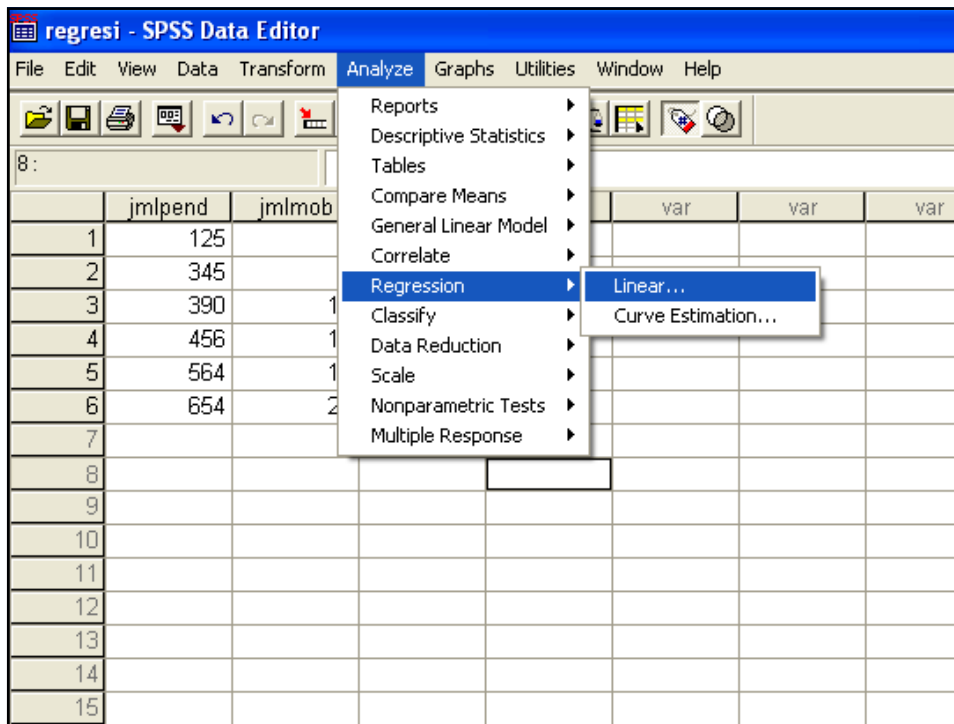
$$Y = - 3,218 + 0,038 (1000)$$

$$= - 3,218 + 38$$

$$= 34,7 \rightarrow \text{jadi jumlah mobil tahun 2010} = 35 \text{ unit.}$$

HASIL PERHITUNGAN SPSS

	jmlpend	jmlmob	var	var	var	v
1	125	4				
2	345	6				
3	390	12				
4	456	14				
5	564	18				
6	654	24				
7						
8						
9						
10						
11						
12						



Regression

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	jmpend ^a	.	Enter

a. All requested variables entered.
b. Dependent Variable: jmlmob

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.950 ^a	.903	.879	2.591

a. Predictors: (Constant), jmpend

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	251.147	1	251.147	37.410	.004 ^a
	Residual	26.853	4	6.713		
	Total	278.000	5			

a. Predictors: (Constant), jmpend
b. Dependent Variable: jmlmob

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-3.218	2.855		-1.127	.323
	jmpend	.038	.006	.950	6.116	.004

a. Dependent Variable: jmlmob

Annotations:

Apa arti **residual**? Beberapa pustaka menyebut SSE (*sum of square error*) yang menunjukkan variasi kesalahan kuadrat yang tidak dapat dijelaskan oleh garis regresi (*unexplained variation*)

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

Semakin kecil variabel pengganggu atau standard error, berarti kenyataan akan mendekati harapan yang terkandung dalam teori.

Menghitung SSE / RESIDUAL =

$$SSE = \sum Y_i^2 - a \sum Y_i - b \sum X_i Y_i$$

$$= 1292 - (-3,218314) (78) - (0,038402) (39482)$$

$$= 26,840728 \rightarrow \text{ini bisa juga untuk menghitung } F \text{ atau } t \text{ hitung.}$$

$$F = t^2 = \frac{b^2 \cdot \sum X_i^2}{SSE / (n-2)}$$

t – test berguna untuk menguji persamaan regresi untuk tiap variabel
F – test berguna untuk menguji persamaan multiple regresi untuk keseluruhan model.

BAGAIMANA MENGHITUNG t (hitung): *→ cek pd populasi*

$$t_{n-2} = \frac{r \sqrt{n-2}}{\sqrt{1-r^2}} \quad d.f = n-2$$

$$= \frac{0,95 \sqrt{6-2}}{\sqrt{1-0,95^2}} = 6,0848$$

⇒ Setelah ini cek t - TABLE

Coefficients ^a					
Model		Unstandardized Coefficients		Standardized Coefficients	t
		B	Std. Error	Beta	
1	(Constant)	-3.218	2.855		-1.127
	jmpend	.038	.006	.950	6.116

a. Dependent Variable: jmlmob