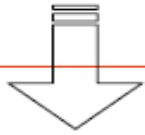


Statistik Bisnis 2 Multivariate – Haryoso Wicaksono, S.Si., M.M., M.Kom.

PENGANTAR MULTIVARIATE

UNI - VARIAT

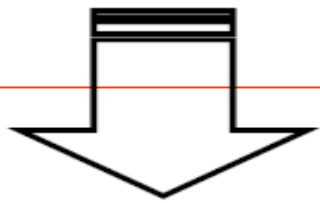


Var-1

Var-2

Var-3

BI - VARIAT



Var-1

Var-2

Var-2

Var-3

Var-3

Var-1

MULTI - VARIAT

Var-2

Var-1

Var-3

Var-4

Var-5

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.

ANALISIS MULTIVARIAT

ASSOCIATION
(HUBUNGAN)

1. MULTIPLE REGRESSION
2. QUANTIFICATION 1
3. TOBIT ANALYSIS
4. PATH ANALYSIS
5. CANONICAL CORRELATION

CLASSIFICATION

1. TESTS OF HYPOTHESES OF MEANS
2. MULTIPLE DISCRIMINANT ANALYSIS
3. QUANTIFICATION II
4. PROBIT ANALYSIS AND LOGIT ANALYSIS
5. LOG-LINEAR MODEL
6. MULTIVARIATE ANALYSIS OF VARIANCE
7. MULTIVARIATE ANALYSIS OF COVARIANCE
8. CLUSTER ANALYSIS

SCALING

1. MULTI-DIMENSIONAL SCALING
2. QUANTIFICATION IV

ABSTRACTION

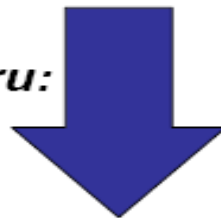
1. PRINCIPAL COMPONENT ANALYSIS
2. QUANTIFICATION III
3. FACTOR ANALYSIS

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

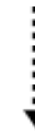
(REGRESI SEDERHANA)

- 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.*

ada/tidaknya
keterkaitan



Seberapa peran
independent variabel pada
prediksi dependent
variabel

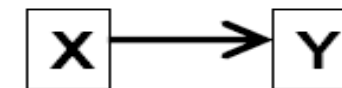
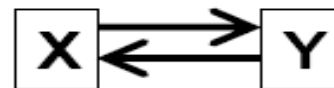


SCATTERGRAMS

KORELASI (r)

REGRESI
SEDERHANA (r^2)

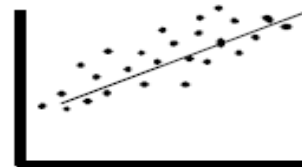
bila secara visual jelas
ada korelasi yang kuat



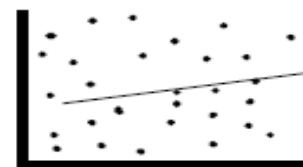
Mengungkapkan tingkat
pengelompokan titik-
titik thd garis regresi
(kekuatan keterkaitan
dua variabel)

- Menarik garis “rata-rata”.
- Memperlihatkan besarnya pengaruh independent var (x) terhadap dependent variabel (Y).

•Bisa untuk memprediksi



Nilai r besar



Nilai r kecil

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

*Pengantar Multivariate [DataSet2] - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help



7 : Jum_Mobil

	Jum_Penduduk	Jum Mobil	var
1	125	4	
2	345	6	
3	390	12	
4	456	14	
5	564	18	
6	654	24	
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			

Linear Regression

Dependent: Jum_Mobil

Block 1 of 1

Independent(s): Jum_Penduduk

Method: Enter

Selection Variable: Rule...

Case Labels:

WLS Weight:

Statistics... Plots... Save... Options...

OK Paste Reset Cancel Help

Linear Regression: Statistics

Regression Coefficients

- Estimates
- Confidence intervals
- Covariance matrix

Residuals

- Durbin-Watson
- Casewise diagnostics
 - Outliers outside: 3 standard deviations
 - All cases

Model fit

- R squared change
- Descriptives
- Part and partial correlations
- Collinearity diagnostics

Continue Cancel Help

→ Regression

[DataSet2]

Descriptive Statistics

	Mean	Std. Deviation	N
Jum_Mobil	13,00	7,457	6
Jum_Penduduk	422,33	184,556	6

Correlations

		Jum_Mobil	Jum_Penduduk
Pearson Correlation	Jum_Mobil	1,000	,950
	Jum_Penduduk	,950	1,000
Sig. (1-tailed)	Jum_Mobil	.	,002
	Jum_Penduduk	,002	.
N	Jum_Mobil	6	6
	Jum_Penduduk	6	6

Variables Entered/Removed^b

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

a. All requested variables entered.

b. Dependent Variable: Jum_Mobil

Regression

Variables Entered/Removed^b

Model	Variables Entered	Variables Removed	Method
1	jmlpend ^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), jmlpend

ANOVA^b

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), jmlpend

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
	jmlpend	.038	.006	.950	6.116	.004

a. Dependent Variable: jmlmob

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.

**SB2 MV 5 BUKU LATIHAN SPSS -
STATISTIK MULTIVARIAT [PUSTAKA MAYA]**

Buku Latihan SPSS Statistik Multivariat

Apa yang dimaksud dengan analisis Multivariat?

Secara umum, Analisis Multivariat atau Metode Multivariat berhubungan dengan metode-metode statistik yang secara bersama-sama (simultan) melakukan analisis terhadap lebih dari dua variabel pada setiap obyek atau orang.

Jadi bisa dikatakan analisis multivariat merupakan perluasan dari analisis univariat (seperti uji t) atau bivariat (seperti korelasi dan regresi sederhana).

Sebagai contoh, jika dilakukan analisis regresi sederhana, dengan satu variabel Y dan satu variabel X, maka analisis seperti itu dikatakan bivariat, karena ada dua (bi) variabel, X dan Y. Sedang jika dilakukan analisis regresi berganda, dengan satu variabel Y dan dua variabel X (X1 dan X2), maka analisis sudah bisa dikatakan *multi*-variat, karena ada tiga variabel (yang berarti, X1 dan X2).

Apa yang dimaksud dengan 'variat' dalam pengertian 'multivariat'?

Variat bisa didefinisikan sebagai suatu kombinasi linier dari variabel-variabel dengan bobot variabel yang ditentukan secara empiris.

Sebagai contoh, ada persamaan regresi berganda:

$$\text{Nilai variat} = w_1.X_1 + w_2.X_2 + w_3.X_3 + \dots + w_n.X_n$$

Di sini X_n adalah variabel yang telah ditentukan oleh peneliti, sedang w_n adalah hasil dari proses multivariat. Nilai variat adalah hasil dari proses perkalian dan penjumlahan w dan X , yang menghasilkan suatu nilai variat tertentu.

Apakah dalam analisis Multivariat perlu dilakukan pemilahan jenis Data?

Data yang secara statistik bisa dibagi menjadi Data metrik dan non metrik tetap perlu dilakukan pada analisis multivariat. Hal ini disebabkan banyak metode multivariat yang justru mengharuskan ciri-ciri data tertentu.

Apa beda data metrik dengan data non metrik?

Data metrik adalah data yang didapat dengan jalan *mengukur* dan bisa mempunyai desimal. Seperti Tinggi Badan, yang bisa saja bernilai 170 cm atau 178,45 cm (desimal). Data metrik akan dikategorikan sebagai data interval atau data rasio.

Sedang data non metrik adalah data yang didapat dengan jalan *menghitung*, tidak mempunyai desimal serta dilakukan dengan kategorisasi. Seperti Jenis Kelamin, yang diberi kode 1 untuk 'Pria' dan 2 untuk 'Wanita'. Data non metrik akan dikategorikan sebagai data nominal atau data ordinal.

Multivariat dan Komputer

Apakah keunggulan Analisis Multivariat dibandingkan dengan Analisis Univariat/Bivariat?

Seperti telah dijelaskan di atas, secara prinsip Multivariat adalah perluasan dari Univariat dan Bivariat, di mana jika Uni atau Bivariat hanya menghitung maksimal dua variabel, Multivariat menghitung lebih dari dua variabel. Dalam praktek Multivariat, semua variabel tersebut dianalisis secara simultan atau bersamaan.

Perbedaan tersebut adalah juga keunggulan bagi Multivariat. Hal ini disebabkan banyak penelitian atau fenomena yang secara alamiah melibatkan banyak variabel. Sebagai contoh, jika ingin dianalisis perilaku konsumen terhadap Rumah Mewah. Perilaku ini sebenarnya melibatkan banyak variabel, seperti Harga Rumah, Motivasi beli, Pengaruh lingkungan, Pendapat teman atau keluarga, alternatif Sistem pembayaran, Kualitas Rumah dan sebagainya. Atau tingkat pertumbuhan sebuah tanaman, yang melibatkan variabel seperti Jumlah pupuk yang diberikan, Curah Hujan, tingkat keasaman tanah, jenis tanah, Intensitas Sinar Matahari dan sebagainya. Kedua contoh di atas tidak bisa diselesaikan dengan menggunakan analisis univariat atau bivariat, karena keterbatasan kedua analisis tersebut.

Jika demikian, mengapa Analisis Multivariat tidak digunakan sejak dahulu dalam menyelesaikan masalah-masalah yang terkait dengan perhitungan Statistik?

Oleh karena waktu itu Komputer (Software) belum mencapai taraf kemajuan seperti sekarang, di lain sisi hampir semua perhitungan Multivariat tidak bisa atau sulit sekali jika dilakukan secara manual. Dengan berkembangnya kualitas Software, seperti SPSS, sekarang dimungkinkan penggunaan berbagai metode Multivariat dalam praktek statistik.

Apakah perhitungan Multivariat harus menggunakan bantuan komputer?

Oleh karena metode Multivariat melibatkan banyak variabel, maka perhitungannya menjadi jauh lebih kompleks dibandingkan analisis yang hanya menggunakan satu atau dua variabel. Dalam hal ini, walaupun ada berbagai teori, seperti teori faktor, teori diskriminant dan sebagainya, namun dalam praktek, teori-teori tersebut akan sangat sulit diaplikasikan jika tanpa menggunakan komputer.

Apakah SPSS hanya satu-satunya program komputer statisitk yang bisa digunakan untuk pengolahan data multivariat?

Selain SPSS, sebenarnya ada program lain yang bisa digunakan untuk mengolah data multivariat, seperti software SAS. Hanya saja SPSS selain program statistik paling populer di Indonesia dan juga di dunia, SPSS juga jauh lebih praktis digunakan oleh kaum awam dibandingkan dengan program SAS.

TEORI UJI DATA.....	4
MODUL 1 Uji Data (1) Missing Value Analysis	6
MODUL 2 Uji Data (2) Perlakuan Terhadap Missing Value	19
MODUL 3 Uji Data (3) Outlier	23
MODUL 4 Uji Data (4) Normalitas	34
MODUL 5 Uji Data (5) Homoskedastisitas	39
MODUL 6 Uji Data (6) Linieritas	43

Apa tujuan Uji Data untuk analisis Multivariat?

Uji Data pada prinsipnya bertujuan untuk memastikan bahwa berbagai metode multivariat (cluster analysis, factor analysis dan lainnya) bisa digunakan pada data tertentu. Dengan demikian, hasil proses multivariat bisa diinterpretasi dengan tepat.

Bagaimana jika Data yang akan diproses tidak dilakukan pengujian terlebih dahulu?

Pengabaian Uji Data bisa berakibat biasanya kesimpulan yang diambil, atau bahkan metode multivariat tidak bisa diproses. Seperti jika Data (yang terdiri atas banyak variabel) mempunyai banyak *missing value* (data yang hilang atau tidak ada isinya). Jika data tersebut dipaksa untuk tetap diproses, output yang dihasilkan bisa sangat berbeda dibandingkan jika data tidak ada yang hilang (*missing*). Pada beberapa jenis data yang sangat banyak mengandung *missing value*, proses multivariat bahkan tidak bisa dilakukan.

Ada berapa macam Uji Data?

Uji Data bisa dilakukan dengan empat cara:

- Pengujian dengan menggunakan Grafik, seperti untuk menguji bentuk kenormalan sebuah distribusi data, menguji sebaran dua variabel untuk korelasi dan sebagainya.
- Pengujian adanya Missing Data, yakni menguji apakah data yang tidak lengkap atau ada data yang hilang akan mempengaruhi data secara keseluruhan.
- Pengujian adanya Outlier (data yang sangat ekstrim), yang mungkin keberadaan data Outlier akan mengganggu keseluruhan data.
- Pengujian beberapa asumsi metode-metode Multivariat, seperti Uji Normalitas Data, Uji Linieritas dan sebagainya.

Pada beberapa modul untuk Uji Data, pengujian secara grafik dilakukan sebagai pelengkap uji lainnya.

Modul 1

UJI DATA (1) MISSING VALUE ANALYSIS

Missing Data atau Missing Value adalah informasi yang tidak tersedia untuk sebuah subyek (kasus). Dalam terminologi SPSS, missing data adalah adanya sel-sel kosong pada satu atau beberapa variabel. Missing Data terjadi karena informasi untuk sesuatu tentang obyek tidak diberikan, sulit dicari atau memang informasi tersebut tidak ada.

Sebagai contoh, pada Data Gaji Responden atau Usia Responden, bisa saja ada Responden yang karena alasan pribadi tidak mau menyebutkan Gaji ataupun Usianya. Hal ini berakibat adanya data yang kosong pada kolom Gaji atau Usia.

Missing Data pada dasarnya tidak bermasalah bagi keseluruhan data, apalagi jika jumlahnya hanya sedikit, misal hanya sekitar 1% dari seluruh data. Namun jika persentase data yang hilang tersebut cukup besar, maka perlu dilakukan pengujian apakah data yang mengandung banyak missing tersebut masih layak diproses lebih lanjut ataukah tidak.

KASUS 1:

Perhatikan isi file `MISSING VALUE.sav` yang isinya sama persis dengan file `DISKRIMINANT.sav` (akan dibahas di modul lain).

File `MISSING VALUE` berisi data 75 konsumen yang digolongkan berdasarkan banyaknya air minum mineral yang dikonsumsinya, yakni:

- **SEDIKIT:** konsumen termasuk sedikit mengkonsumsi air mineral. Kode untuk tipe ini adalah 0.
- **BANYAK:** konsumen termasuk banyak mengkonsumsi air mineral. Kode untuk tipe ini adalah 1.

NB: Kode untuk data kategori bisa dilihat dengan menekan `CTRL+T`, lalu lihat bagian (kolom) `VALUES`.

Sedang data lain adalah data bertipe rasio, seperti Usia konsumen, Berat dan Tinggi badan konsumen, Income konsumen, Jam Kerja konsumen dalam sehari, serta kegiatan Olahraga konsumen (jam) dalam sehari.

missing value [DataSet1] - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

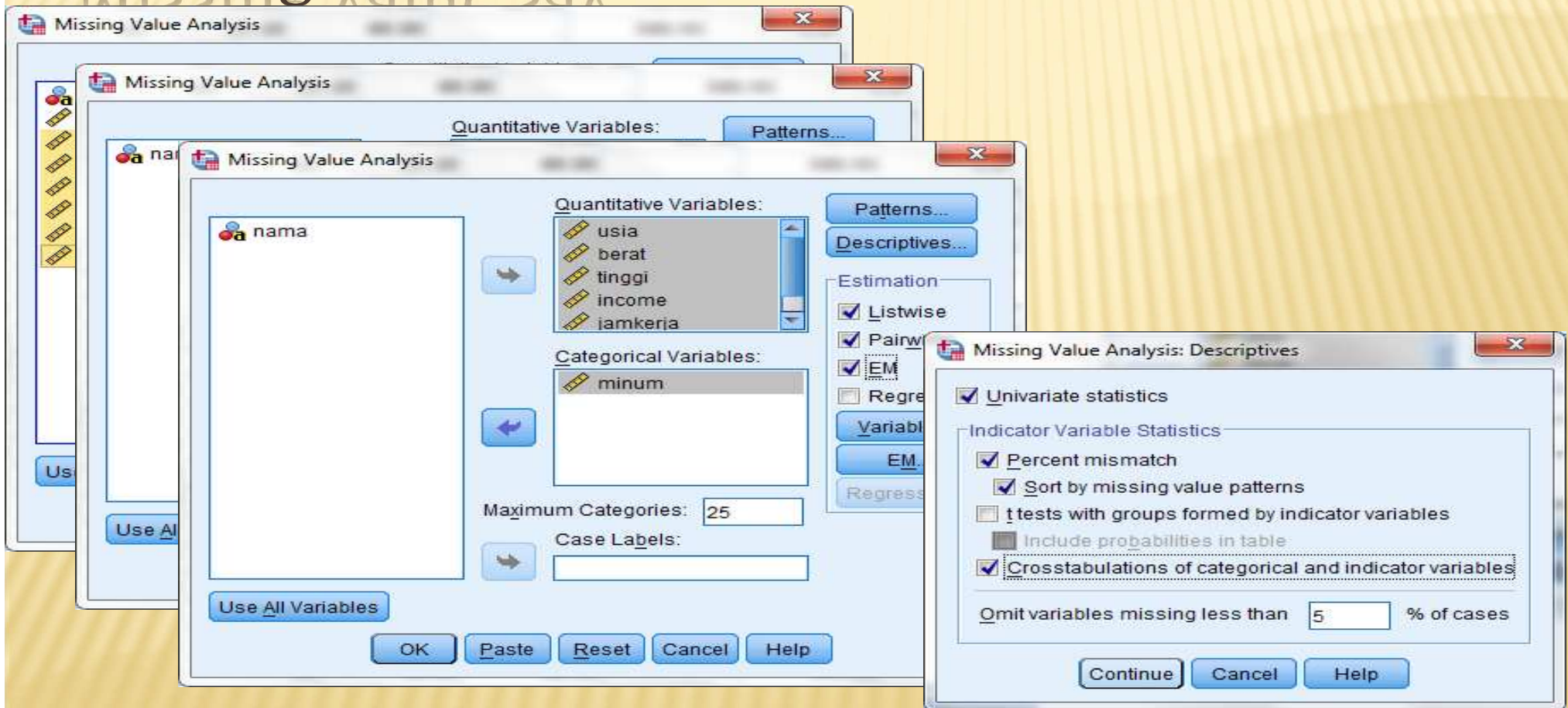
1 : nama RUSDI

	nama	minum	usia	berat	tinggi	income	jamkerja	olahraga
1	RUSDI	.00	40,00	.	154,00	680,00	5,33	3,0
2	NINA	.00	30,00	70,00	157,00	700,00	5,30	3,6
3	LANNY	.00	25,00	60,00	.	580,00	5,27	3,5
4	CITRA	.00	26,00	75,00	160,00	600,00	5,33	3,0
5	DINA	.00	.	50,00	159,00	700,00	5,50	3,5
6	SISKA	1,00	28,00	62,00	158,00	440,00	5,00	2,2
7	LUSI	.00	29,00	.	160,00	580,00	5,07	2,9
8	LENNY	.00	40,00	52,00	165,00	.	5,13	4,0
9	RUDI	1,00	35,00	68,00	150,00	700,00	5,17	3,5
10	ROBY	.00	36,00	70,00	152,00	720,00	5,23	3,6
11	BAMBANG	.00	.	50,00	.	780,00	5,33	3,9
12	YUNUS	.00	30,00	62,00	155,00	600,00	5,30	3,0
13	LESTARI	.00	34,00	60,00	157,00	680,00	5,27	2,9
14	ERNI	1,00	35,00	.	160,00	700,00	5,33	4,0
15	ESTI	1,00	.	62,00	165,00	580,00	5,50	3,5
16	HANY	1,00	30,00	51,00	162,00	600,00	5,00	3,6
17	HESTY	.00	35,00	80,00	157,00	700,00	5,33	3,9
18	SUSAN	1,00	22,00	52,00	154,00	440,00	5,30	.
19	LILIS	.00	40,00	72,00	155,00	800,00	5,27	3,4
20	LITA	1,00	41,00	45,00	164,00	820,00	5,33	.

Data View Variable View

SPSS Processor is ready

Missing value.sav



OUTPUT SPSS

Univariate Statistics

	N	Mean	Std. Deviation	Missing		No. of Extremes ^a	
				Count	Percent	Low	High
usia	66	30.1364	6.36792	9	12.0	0	0
berat	69	54.9130	10.74400	6	8.0	0	0
tinggi	71	160.1268	6.88669	4	5.3	0	7
income	73	618.9041	123.64332	2	2.7	0	0
jamkerja	73	5.2630	.12602	2	2.7	0	0
olahraga	73	3.048	.6377	2	2.7	0	0
minum	75			0	.0		

a. Number of cases outside the range (Q1 - 1.5*IQR, Q3 + 1.5*IQR).

Summary of Estimated Means

	usia	berat	tinggi	income	jamkerja	olahraga
Listwise	29.9057	55.5094	159.8302	609.4340	5.2478	2.996
All Values	30.1364	54.9130	160.1268	618.9041	5.2630	3.048
EM	30.2384	54.9249	160.1303	620.4232	5.2640	3.048

Modul 2

UJI DATA (2) PERLAKUAN TERHADAP MISSING VALUE

Jika pada modul Missing Data (Value) bagian pertama dijelaskan pengujian kerandoman Missing Data yang ada, modul berikut melanjutkan penanganan Data yang missing dan ternyata bersifat random.

Jika missing value terbukti random, dalam arti missing value yang terjadi tidak disengaja dan tidak mengacu keadaan tertentu (misal missing hanya pada variabel usia), maka berbagai perlakuan (treatment) bisa dilakukan pada data-data yang missing.

Penanganan terhadap Missing Value bisa bervariasi, seperti membuang baris (kasus) yang mengandung missing value, menghapus variabel (kolom) yang mengandung missing value dan sebagainya.

Salah satu cara yang populer adalah bukan menghilangkan baris atau kolom yang mengandung data missing, namun justru *mengisi sel (data) yang missing dengan nilai tertentu yang dianggap bisa mendekati kenyataan sebenarnya jika data terisi*. Hal ini lebih baik dan rasional daripada membuang satu baris (data konsumen) hanya karena usia konsumen tidak terdata, atau bahkan satu variabel hanya karena satu dua sel tidak terisi. Cara mengisi data yang missing bisa bermacam-macam, dan yang populer adalah mengisi dengan *rata-rata keseluruhan data*. Sebagai contoh, jika akan mengisi data usia yang hilang, cari rata-rata usia konsumen secara keseluruhan, kemudian mengisi setiap data missing dari variabel usia dengan angka rata-rata tersebut.

HASIL TRANSFORM → REPLACE MISSING VALUE

Result Variables

	Result Variable	N of Replaced Missing Values	Case Number of Non-Missing Values		N of Valid Cases	Creating Function
			First	Last		
1	usia_1	9	1	75	75	SMEAN(usia)
2	berat_1	6	1	75	75	SMEAN(berat)
3	tinggi_1	4	1	75	75	SMEAN (tinggi)
4	income_1	2	1	75	75	SMEAN (income)
5	jamkerja_1	2	1	75	75	SMEAN (jamkerja)
6	olahraga_1	2	1	75	75	SMEAN (olahraga)