Berurusan dengan Outliers

Wahyu Widhiarso | Fakultas Psikologi UGM | Tahun 2001

Outliers adalah nama bagi subyek, subyek yang unik. Yang unik-unik ini kadang bisa mengacaukan. Nilainya jauh dari rata-rata kebanyakan orang lain. Secara statistik ini bisa dihilangkan. Ini nih contoh visual bagi yang namanya *outliers*.



Gambar A Outliers pada Distribusi Normal Gambar B Outliers pada Korelasi

Cara di atas adalah cara mengidentifikasi outliers dengan cara meraba-raba saja. Ada juga cara yang lebih praktis. Misalnya dengan melihat nilai-nilai yang ekstrim.



Tetapi hanya 5 subyek yang paling atas dan paling bawah saja yang ditampilkan. Yaitu mereka yang memiliki *Ekstrem value*

B. Ouliers : Jauh Dari Garis Korelasi

Outliers pada acara ini adalah mereka yang jauh dari garis korelasi. Pada gambar B di atas, ada satu subyek outliers. Ini nih cara mengidentifikasikannya.

Sebelum menganalisis. Beri satu kolom tambahan pada data anda...yaitu nomor subyek. Ini contohnya...



Keterangan

- 1. Cook's Distance : ukuran pengaruh subyek pada model. Nilai Cook's >1 adalah subyek outliers
- 2. Mahal's Distance : ukuran jarak nilai subyek dari garis yang dikehendaki. Makin jauh jarak dari pusat garis korelasi, makin jelek dia bukan? Untuk jumlah sampel 30 (N=30), nilai di atas 11 perlu dipertimbangkan untuk di del, sedangkan N=100, nilai di atas 15 yang di del. Ukuran ini Barnet dan Lewis (1978) yang bilang Iho...bukan aku. Kalau aku sih, 2 subyek yang nilainya yang paling besar dan ekstrem yang aku del. Terserah anda Iha !
- 3. Leverage Distance : nilai yang mendekati 1 saja yang di del.
 - Aturan yang saya rekomendasikan adalah Mahal Distance kemudian Cook's distance

Ini adalah contoh outputnya. Coba lihat ternyata ada nilai Mahal's dan Cook's yang sangat besar !

Nilai Mahal's nya ada yang di atas 11!

Kita lihat subyek berapa yang memiliki nilai itu....Kembalilah pada Windows SPSS yang memuat daftar nomor subyek dan nilai subyek...

Residuals Statistics ^a									
	Minimum	Maximum	Mean	Std. Deviation	N				
Predicted Value	13,19	24,04	19,50	3,016	38				
Std. Predicted Value	-2,091	1,507	,000	1,000	38				
Standard Error of Predicted Value	1,794	9,377	2,802	1,320	38				
Adjusted Predicted Value	-39,41	23,03	17,90	10,037	38				
Residual	-17,68	21,96	,00,	10,696	38				
Std. Residual	-1,608	1,997	,000,	,973	38				
Stud. Residual	-1,631	3,535	,052	1,119	38				
Deleted Residual	-18,20	74,41	1,60	16,322	38				
Stud. Deleted Residual	-1,672	4,345	,079	1,212	38				
Mahal. Distance	,011	25,927	1,947	4,239	38				
Cook's Distance	,000	11,094	,315	1,797	38				
Centered Leverage Value	,000	,701	,053	,115	38				

a. Dependent Variable: nomor



Tingkat Kecerewetan

- 1. Coba tekan menu <u>analyze</u> \rightarrow <u>descriptive</u>
- 2. Masukkan data yang hendak diidentifikasi pada kotak variables

Descriptives	×	tas - SPSS	S Data Ed alyze <u>G</u> raphs	l itor <u>U</u> tilities <u>W</u> i	ndow <u>H</u> elp
nomor [nomor] Variable(s):	ОК		= <u>[?</u> #	<u> 作 一 田</u>	
🐨 Skala Ngerumpi (rumpi)	Paste	0,221896201424995			
	Barrat	rumpi	jealous	zrumpi	zjealous
	Tieser	10	51	- ,52211	- ,22282
	Cancel	9	47	-,77011	-,72090
	Help	9	48	-,77011	- ,59638
		9	45	-,77011	- ,96994
		9	46	-,77011	- ,84542
Save standardized values as variables	Options	9	47	-,77011	-,72090
1		9	45	-,77011	- ,9699
Liceates and saves one 2-score variable for each selected		9	60	-,77011	,8978
to the first seven characters of original variable names.	4	8	45	-1,01811	- ,96994
		8	48	-1,01811	- ,59638
Mombust Nilsi 7	, <u> </u>	8	47	-1,01811	-,72090
			1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1	
Takan katak ini agar tian variahal ada nilai 7 nya	32 32	8	47	-1,01811	-,72090
Tekan kotak ini agar tiap variabel ada nilai Z-nya.	12 32 13 33	8	47	-1,01811 -1,26611	- ,72090 -1 ,34350
Tekan kotak ini agar tiap variabel ada nilai Z-nya. Tekan OK. Lalu kembalilah pada Windows SPSS	12 32 13 33 14 34	8 7 7	47 42 41	-1,01811 -1,26611	- ,72090 -1 ,34350 -1 ,46802
Tekan kotak ini agar tiap variabel ada nilai Z-nya. Tekan OK. Lalu kembalilah pada Windows SPSS utama. Soalnya Keluarnya nilai Z nanti pada	32 32 33 33 34 34 35 35	8 7 7 25	47 42 41 41	-1,01811 -1,26611 -1,20011 3,19792	- ,72090 -1 ,34350 -1 ,46802 -1 ,4 `802
Tekan kotak ini agar tiap variabel ada nilai Z-nya. Tekan OK. Lalu kembalilah pada Windows SPSS utama. Soalnya Keluarnya nilai Z nanti pada	12 32 13 33 14 34 15 35 16 36	8 7 7 25 17	47 42 41 41 68	-1,01811 -1,26611 -1,20011 3,19792 1,21390	-,72090 -1,34350 -1,46802 -1,47802 -1,47802
Tekan kotak ini agar tiap variabel ada nilai Z-nya. Tekan OK. Lalu kembalilah pada Windows SPSS utama. Soalnya Keluarnya nilai Z nanti pada	12 32 13 33 14 34 35 35 36 36 37 37	8 7 7 25 17 16	47 42 41 41 68 67	-1,01811 -1,26611 3,19792 1,21390	-,72090 -1,34350 1,46802 -1,4 802 1,5 401 1,76949
Tekan kotak ini agar tiap variabel ada nilai Z-nya. Tekan OK. Lalu kembalilah pada Windows SPSS utama. Soalnya Keluarnya nilai Z nanti pada	12 32 13 33 14 34 15 35 16 36 17 37 18 38	8 7 7 25 17 16 15	47 42 41 41 68 67 65	-1,01811 -1,26611 3,19792 1,21390 	-,7209(-1,34350 -1,4680, -1,4680, -1,4680, -1,4680, -1,4680, -1,4680, 1,76949 1,52045

"Jika dataku banyak.. Masak aku mecicili satu persatu tiap kotak kecil di SPSS untuk mencari nilai Z diatas 3. Mataku Bisa Burem gara-gara Statistik!" (kata seorang teman)

Ulangi cara di atas..Menu <u>Analyze</u> → <u>Descriptive</u> → masukkan nilai Z ke kotak <u>Variables</u> Kotak <u>Save Standard Value</u> gak usah dihidupkan. Lalu muncul output seperti ini..

	Descriptive Statistics			
Descriptives Variable(s): OK Stale Neuroni (nume) Stale Neuroni (nume)		score: Skala Ngerump	Zscore: Skala Kecemburuan	
Skala Kecemburuan (je Skala Kecemburuan (je	Ν	38	38	
_ <u>Beset</u>	Minimum	1,00011	-1,46802	
Cancel	Maximum	3,19792	2,51660	
Нер	Mean	,0000000	,0000000	
Save standardized values as variables	Std. Deviation	1,0000000	1,00000000	

Ini nih ternyata ada **Outliers** pada data kita. Kalau pengen cari ya..urutkan z score melalui menu <u>Data</u> \rightarrow <u>Sort Case</u> lalu urutkan data anda berdasarkan *Z score* variabel yang ingin diketahui outliernya