

Modul 3

UJI DATA (3) OUTLIER

Data Outlier adalah data yang secara nyata berbeda dengan data-data yang lain. Sebagai contoh, data dari 100 tinggi badan orang di Asia, ternyata ada data dengan tinggi badan 210 centimeter. Tinggi badan tersebut jelas bersifat 'ekstrim' dibanding rata-rata tinggi badan orang asia pada umumnya, misal sekitar 160 centimeter. Data 210 centimeter inilah yang disebut **data outlier**.

Data Outlier bisa terjadi karena beberapa sebab:

1. Kesalahan dalam pemasukan data. Terkait dengan contoh di atas, ternyata data yang seharusnya 150 centimeter diinput 210 centimeter.
2. Kesalahan pada pengambilan sampel. Terkait dengan contoh di atas, mereka yang diambil sampel Tinggi Badan adalah para pemain Basket profesional, yang memang tinggi badan mereka di atas rata-rata.
3. Memang ada data-data ekstrim yang tidak bisa dihindarkan keberadaannya. Terkait dengan contoh di atas, memang ada banyak orang Asia yang *benar-benar* mempunyai Tinggi Badan 210 centimeter

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
usia	75	20.00	45.00	30.4267	6.27344
berat	75	35.00	120.00	55.6933	13.82905
tinggi	75	148.00	190.00	160.3200	7.56264
income	75	400.00	1000.00	623.7333	131.75298
jamkerja	75	5.00	5.50	5.2653	.12538
olahraga	75	2.0	4.5	3.053	.6312
Valid N (listwise)	75				

Modul 4

UJI DATA (4) NORMALITAS

Tujuan Uji Normalitas adalah ingin mengetahui apakah distribusi sebuah data mengikuti atau mendekati distribusi normal, yakni distribusi data dengan bentuk lonceng (bell shaped). Data yang 'baik' adalah data yang mempunyai pola seperti distribusi normal, yakni distribusi data tersebut tidak menceng ke kiri atau menceng ke kanan.

Uji normalitas pada multivariat sebenarnya sangat kompleks, karena harus dilakukan pada seluruh variabel secara bersama-sama. Namun uji ini bisa juga dilakukan pada setiap variabel, dengan logika bahwa *jika secara individual masing-masing variabel memenuhi asumsi normalitas, maka secara bersama-sama (multivariat) variabel-variabel tersebut juga bisa dianggap memenuhi asumsi normalitas.*

Lihat Output berupa tabel:

	Kolmogorov-Smirnov ^a		
	Statistic	df	Sig.
USIA	.087	75	.200*
BERAT	.165	75	.000

*. This is a lower bound of the true significance.
a. Lilliefors Significance Correction

Kriteria pengujian:

- Angka signifikansi (SIG) $>0,05$, maka data berdistribusi normal
- Angka signifikansi (SIG) $<0,05$, maka data tidak berdistribusi normal

Jika sebuah variabel mempunyai sebaran data yang tidak normal, perlakuan yang dimungkinkan agar menjadi normal adalah:

- Menambah jumlah data. Seperti pada kasus, bisa dicari 20 atau 30 atau sejumlah data baru untuk menambah ke 75 data Berat Badan konsumen yang sudah ada. Kemudian dengan jumlah data yang baru, dilakukan pengujian sekali lagi.
- Menghilangkan data yang dianggap penyebab tidak normalnya data. Seperti pada variabel BERAT, jika dua data yang outlier dibuang, yakni berat 100 dan 120, kemudian diulang proses pengujian, mungkin data bisa menjadi normal. Jika belum normal, ulangi pengurangan data yang dianggap penyebab ketidaknormalan data. Namun demikian, pengurangan data harus dipertimbangkan, apakah tidak mengaburkan tujuan penelitian karena hilangnya data yang seharusnya ada.
- Dilakukan transformasi data, misal mengubah data ke logaritma atau ke bentuk natural (LN) atau bentuk lainnya, kemudian dilakukan pengujian ulang.
- Data diterima apa adanya, memang dianggap tidak normal dan tidak perlu dilakukan berbagai *treatment*. Untuk itu, alat analisis yang dipilih harus diperhatikan, seperti untuk multivariat mungkin factor analysis tidak begitu mementingkan asumsi kenormalan. Atau pada kasus statistik univariat, bisa dilakukan alat analisis non parametrik.

Modul 5

UJI DATA (5) HOMOSKEDASTISITAS

Uji Homoskedastisitas pada prinsipnya ingin menguji apakah sebuah grup (data kategori) mempunyai varians yang sama di antara anggota grup tersebut. Jika varians sama, dan ini yang seharusnya terjadi, maka dikatakan ada Homoskedastisitas. Sedangkan jika varians tidak sama, dikatakan terjadi Heteroskedastisitas.

Alat untuk menguji Homoskedastisitas bisa dibagi dua, yakni dengan alat analisis Levene Test, atau dengan Analisis Residual yang berupa grafik. Kasus berikut akan membahas pengujian Homoskedastisitas dengan menggunakan Levene Test.

Test of Homogeneity of Variance

		Levene Statistic	df1	df2	Sig.
TINGGI	Based on Mean	4.244	1	73	.043
	Based on Median	3.789	1	73	.055
	Based on Median and with adjusted df	3.789	1	71.641	.056
	Based on trimmed mean	4.143	1	73	.045

JAMKE

Proses pengujian untuk variabel TINGGI:

Menentukan Hipotesis:

Ho: kedua varians populasi adalah identik

H1: kedua varians populasi adalah tidak identik

NB: Yang dimaksud dua populasi adalah populasi konsumen yang minum SEDIKIT, dan mereka yang minum BANYAK. Dengan demikian, jumlah populasi disesuaikan dengan kategori data yang ada pada variabel yang dimasukkan pada bagian FACTOR LIST.

Kriteria Pengujian:

Jika Probabilitas (SIG) > 0,05, maka Ho diterima

Jika Probabilitas (SIG) < 0,05, maka Ho ditolak

Test of Homogeneity of Variance

		Levene Statistic	df1	df2	Sig.
TINGGI	Based on Mean	4.244	1	73	.043
	Based on Median	3.789	1	73	.055
	Based on Median and with adjusted df	3.789	1	71.641	.056
	Based on trimmed mean	4.143	1	73	.045
JAMKERJA	Based on Mean	2.822	1	73	.097

Kriteria Pengujian:

Jika Probabilitas (SIG) > 0,05, maka Ho diterima

Jika Probabilitas (SIG) < 0,05, maka Ho ditolak

Keputusan:

Pada baris TINGGI dari tabel output di atas, dan dengan dasar Mean, didapat angka SIG adalah 0,043. Oleh karena angka SIG. < 0,05 maka Ho ditolak. Hal ini berarti varians dari data Tinggi Konsumen yang minum sedikit air mineral berbeda secara nyata dengan data Tinggi Konsumen yang minum banyak air mineral. Dapat disimpulkan, telah terjadi heteroskedastisitas pada variabel TINGGI dengan dasar grup MINUM.

Modul 6

UJI DATA (6) LINIERITAS

Linieritas adalah keadaan di mana hubungan antara variabel dependen dengan variabel independen bersifat linier (garis lurus) dalam range variabel independen tertentu. Sebagai contoh, hubungan antara Kecepatan Lari seseorang (variabel dependen) yang tergantung pada Usia orang tersebut (variabel bebas atau independen). Secara umum dapat dikatakan bahwa makin tinggi usia seseorang, maka lari orang tersebut cenderung semakin cepat, yang jika direpresentasikan pada grafik, akan terdapat garis ke kanan atas. Namun sebenarnya, hal itu benar hanya pada range usia tertentu, misal antara 17 tahun sampai 40 tahun. Di atas 40 tahun mungkin kecepatan lari seseorang berbanding terbalik dengan usianya, yakni makin tinggi usia orang tersebut, makin lambat larinya.

Linieritas bisa diuji dengan menggunakan *scatter plot* (diagram pencar) seperti yang digunakan untuk deteksi data outlier, dengan memberi tambahan garis regresi. Oleh karena *scatter plot* hanya menampilkan hubungan dua variabel saja, maka jika terdapat lebih dari dua data, pengujian akan dilakukan dengan berpasangan tiap dua data.