

Pertemuan 6

Husni Mubarak Ramadhan

2023-03-03

2 Statistik deskripsi dengan R

Sebelum memulai dengan konsep dasar analisis data, seseorang harus menyadari berbagai jenis data dan cara untuk mengatur data dalam file komputer.

2.1 Beberapa istilah dasar **Populasi** – agregat subjek (makhluk, benda, kasus, dan sebagainya). Untuk studi tertentu, *populasi target* harus ditentukan: pada subjek mana kita akan menggeneralisasi atau menggunakan hasilnya?

Sampel – kumpulan subjek *dalam penelitian*. Secara umum, sampel harus representatif untuk populasi target.

Observasi – unit studi atau *subjek* atau individu. Seringkali manusia, terkadang juga hewan, tumbuhan atau apa pun.

Variabel – kualitas atau kuantitas, diukur atau dicatat untuk setiap subjek dalam sampel (usia, jenis kelamin, tinggi badan, berat badan, tingkat merokok, dll.).

Dataset – seperangkat nilai dari semua variabel yang menarik bagi semua individu dalam penelitian ini. Hasil numerik yang diperoleh dari dataset akan digunakan untuk menarik kesimpulan tentang populasi target.

2.2 Organisasi data Himpunan data sebagian besar diatur (dan disimpan sebagai file komputer) dalam bentuk *matriks data*.

Matriks data yang mewakili jenis kelamin (1-laki-laki; 0-perempuan), usia, tidak. anak-anak, berat (kg), dan tinggi (cm) 7 orang:

No	Jenis Kelamin	Umur	Nomor anak	Berat	Tinggi
1	0	57	1	65	158
2	1	70	3	100	175
3	0	45	0	71	162
4	0	38	2	58	164
5	0	25	1	81	170
6	1	50	4	68	172
7	1	61	0	85	179

Setiap baris matriks semacam itu mewakili satu pengamatan. Semua baris memiliki panjang yang sama: data yang sama telah direkam untuk semua individu. Setiap kolom mewakili satu variabel. Misalnya, Berat adalah nama variabel, yang mewakili berat badan (dalam kg) seseorang.

2.3 Jenis data • Data numerik

-Data diskrit – variabel hanya dapat mengambil nilai bilangan bulat (0, 1, 2 dll.)

Contoh: jumlah anak, jumlah teman

-Data kontinu – setiap nilai bernomor nyata (seringkali dalam rentang tertentu) adalah contoh yang mungkin: berat badan, usia

• Data kualitatif (non-numerik, kategoris)

-Data nominal: kategori tidak berurutan contoh: golongan darah, warna mata

-Data ordinal atau terurut: urutan kategori contoh: tingkat merokok, sikap (baik-sedang-buruk)

Pengkodean numerik dari data nominal atau pesanan tidak membuat data numerik!

2.4 Meringkas/menyajikan data Data kontinu/diskrit Statistik lokasi ringkasan: rata-rata, median. Rata-rata sampel adalah rata-rata aritmatika data. Ini dapat dihitung, dengan menjumlahkan semua nilai data dan membagi jumlah dengan ukuran sampel total.