

BAB I

PENDAHULUAN STATISTIKA

1.1 Beberapa istilah dasar

Kita sangat akrab dengan kata “statistik” dalam kehidupan sehari-hari, bahkan di negara kita terdapat lembaga negara yang bernama Badan Pusat Statistik (BPS). Kita juga sering mendengar istilah “observasi”, “data”, “sensus”, “sample”, “populasi” dan lain-lain. Mirip dengan kata statistik, terdapat kata “statistika” seperti terlihat pada judul bab ini di atas. Berikut definisi beberapa istilah tersebut:

STATISTIKA adalah kumpulan metoda yang digunakan untuk merencanakan eksperimen, mengambil data, dan kemudian menyusun, meringkas, menyajikan, menganalisa, menginterpretasikan dan mengambil kesimpulan yang didasarkan pada data tersebut.

DATA adalah hasil observasi atau pengamatan yang telah dikumpulkan. Data dapat berupa hasil pengukuran; misalnya data tinggi dan berat badan, hasil pengelompokan; misalnya jenis kelamin, hasil jawaban responden terhadap suatu questioner; misalnya tingkat kepuasan.

POPULASI adalah koleksi lengkap semua elemen yang akan diselidiki. Suatu koleksi dikatakan lengkap jika ia memuat semua subjek yang akan diselidiki.

SENSUS adalah koleksi data dari semua anggota dalam populasi.

SAMPEL adalah sebagian koleksi anggota yang dipilih dari populasi.

STATISTIKA DESKRIPTIF adalah statistika yang berkaitan dengan analisis dan deskripsi suatu grup sebagai populasinya, tanpa melakukan penarikan kesimpulan apapun untuk komunitas yang lebih luas dari grup tersebut.

STATISTIKA INFERENSI adalah statistika yang mencoba untuk membuat suatu deduksi atau kesimpulan pada populasi dengan menggunakan sampel dari populasi tersebut.

Sebagai contoh, suatu lembaga survey melakukan wawancara terhadap 2350 penduduk Indonesia untuk mengetahui tingkat kepuasan terhadap kinerja pemerintah. Dalam hal ini sebanyak 2350 penduduk merupakan **sampel** dan keseluruhan penduduk Indonesia sekitar 230 juta jiwa adalah **populasinya**. Kalau tidak salah, setiap 5 tahun sekali pemerintah melakukan sensus ekonomi atau sensus pertanian. Pada kegiatan sensus semua kepala keluarga didata dan data yang terkumpul disebut sensus atau **data sensus**.

Pengumpulan data dengan cara sensus membutuhkan biaya, waktu dan tenaga yang banyak. Untuk alasan efisiensi, dalam banyak kasus pola atau kelakuan populasi cukup dipelajari melalui sampelnya. Nantinya, hasil analisis pada sampel ini digunakan untuk memberikan kesimpulan pada populasi asalnya. Agar dapat diharapkan kesimpulan yang valid maka sampel yang diambil haruslah representatif, artinya ia benar-benar mewakili populasinya. Sampel yang tidak valid akan melahirkan kesimpulan yang menyimpang dari keadaan yang sesungguhnya.

Pemilu atau pilkada di Indonesia dilakukan untuk mengetahui aspirasi dari semua pemilih. Jadi pemilu merupakan proses sensus untuk populasi pemilih; walaupun kenyataannya tidak semua data populasi dapat diperoleh karena banyaknya “golput”. Sedangkan, lembaga survey yang melakukan perhitungan cepat atau “quick count” adalah melakukan proses sampling, artinya data hanya diambil dari sebagian TPS yang tersebar dengan cara sedemikian rupa sehingga data yang diperoleh “dipercaya” dapat mewakili para pemilih semuanya. Hasilnya sangat cepat diperoleh dikarenakan data yang diambil hanya sebagian kecil dari data sesungguhnya. Keakuratan kesimpulan yang diambil bergantung pada kualitas sampel yang diambil dan metoda analisis data yang digunakan.

Ingat, dalam sistem sampling terdapat faktor kesalahan yang sudah diperhitungkan sejak awal. Diantara faktor kesalahan ini adalah sampling error yang merupakan ukuran peluang ketidakmiripan sampel dengan populasinya. Juga, metoda yang digunakan dalam melakukan analisis data selalu didasarkan pada teori probabilitas. Artinya tidak ada kesimpulan apapun dalam statistik yang bersifat eksak; semuanya mempunyai peluang kejadian sebaliknya. Sangat dimungkinkan beberapa lembaga survey perhitungan cepat pilkada memberikan kesimpulan yang berbeda satu sama lainnya; terutama bila keadaan sesungguhnya hanya memberikan selisih yang sangat tipis. Masih ingat dengan kasus pilkada Jawa Timur beberapa waktu yang lalu?

1.2 Tipe-tipe data

Pada bagian sebelumnya kita telah mendefinisikan sampel dan populasi. Keduanya dibedakan berdasarkan proses melakukan observasi. Untuk membedakan antara data sampel dan data populasi biasanya digunakan istilah **statistik** dan **parameter**.

PARAMETER adalah suatu ukuran numerik yang menggambarkan karakter suatu populasi.

STATISTIK adalah ukuran numerik yang menggambarkan karakter suatu sampel.

CONTOH

1. Berdasarkan sensus ekonomi tahun 2010 terdapat 35% rumah tangga di Indonesia tergolong miskin. Nah, angka 35% ini adalah parameter karena ia diperoleh dari populasi yaitu semua rumah tangga di Indonesia.
2. Berdasarkan hasil survey terhadap 50 orang mahasiswa pendidikan matematika UNMUH Ponorogo angkatan 2008/2009 diperoleh bahwa rata-rata NEM matematika mereka adalah 6.75. Angka 6.75 ini adalah statistik karena ia diberikan oleh sampel yang terdiri dari 50 orang mahasiswa tersebut.

Selain data yang berbentuk angka seperti contoh di atas, terdapat pula data dalam bentuk kategori. Kedua bentuk data ini didefinisikan secara formal sebagai berikut :

DATA KUANTITATIF adalah data yang menggambarkan hasil perhitungan atau hasil pengukuran.

DATA KUALITATIF atau **DATA KATEGORI** adalah data yang dapat dipisahkan dalam beberapa kategori atau kelompok yang dibedakan oleh karakter bukan numerik.

CONTOH

1. **Data kuantitatif:** tinggi badan, nilai NEM, temperatur dalam derajat celsius, besar penghasilan.
2. **Data kualitatif:** jenis kelamin, profesi, temperatur dalam rasa (dingin, panas sejuk).

Selanjutnya, data kuantitatif dibedakan atas data diskrit dan data kontinu.

DATA DISKRIT adalah data yang banyak kemungkinannya berhingga atau terbilang.

DATA KONTINU adalah data yang banyak kemungkinannya takterbilang.

CONTOH

1. **Data diskrit:** jam kerja dalam sehari (kemungkinannya: adalah 1, 2, 3, ... , 24), banyak telur yang dihasilkan oleh ayam betina, banyak hari libur dalam setiap bulan.
2. **Data kontinu:** temperatur udara di berbagai tempat (kemungkinannya: semua nilai yang ada pada interval, misalnya dari -20 derajat celsius sampai dengan 50 derajat celsius).

1.3 Level Pengukuran

Cara umum yang digunakan untuk mengklasifikasikan data adalah ditentukan oleh empat macam level pengukuran, yaitu level **nominal**, **ordinal**, **interval** dan **rasio**. Dalam statistika terapan, level pengukuran data merupakan faktor penting dalam menentukan prosedur dan metoda statistika yang digunakan.

LEVEL NOMINAL dicirikan oleh data yang terdiri atas nama-nama, label, atau kategori. Data seperti ini tidak dapat diurutkan seperti dari atas ke bawah, atau sebaliknya.

CONTOH: Berikut adalah contoh-contoh yang mengilustrasikan pengukuran level nominal:

1. **Ya, tidak, tidak tahu:** biasanya diberikan pada lembar kuesioner.
2. **Warna:** warna mobil yang dimiliki oleh dosen UNMUH Ponorogo (hitam, merah, putih, biru, dan lain-lain).

Data-data yang diperoleh pada level ini tidak dapat diurutkan. Data ini tidak dapat digunakan untuk kalkulasi, misalnya $Ya + tidak\ tahu = ???$, $merah + hitam = ???$ tidak dapat dilakukan.

LEVEL ORDINAL data yang diperoleh pada level ini dapat disusun dalam urutan tertentu, tetapi selisih nilai-nilainya tidak dapat ditentukan atau bahkan tidak bermakna sama sekali.

CONTOH: Berikut adalah contoh data yang diperoleh dari pengukuran level ordinal

1. Nilai akhir pada KHS mahasiswa yang diberikan oleh pak Julian HERNADI: E, D, C, B-, B, A-, A. Nilai-nilai ini dapat diurutkan, misalnya nilai A lebih baik dari nilai B, tetapi seberapa besar selisih antara A dan B tidak dapat ditentukan. Jelasnya A- B tidak bermakna.
2. *Transparency International* Indonesia (TII) baru-baru ini mengumumkan ranking indeks persepsi korupsi (IPK) untuk 50 kota yang ada di Indonesia. Dari ke 50 kota tersebut, Yogyakarta menduduki kota terbersih pada ranking pertama, disusul

Palangkaraya pada ranking kedua, Banda Aceh pada ranking ketiga dan seterusnya sampai Kupang pada ranking ke 50 atau terkorup. Data ranking di sini merupakan level pengukuran ordinal. Walaupun ada angka di sini namun selisih antara ranking 2 dan ranking 1, bila ditulis dalam bentuk $2-1 = 1$ tidak mempunyai makna sama sekali.

LEVEL INTERVAL seperti level ordinal dengan sifat tambahannya adalah selisih antara dua data mempunyai makna. Tetapi level ini tidak mempunyai **titik nol** alami sebagai titik awal.

CONTOH: Berikut inidata dalam level interval

1. **TEMPERATUR:** suhu badan 36 derajat celsius dan 37 derajat celsius merupakan contoh data dalam level interval. Nilai-nilai ini dapat diurutkan dan selisihnya dapat ditentukan dengan jelas, dalam contoh ini selisihnya adalah 2 derajat celsius. Tetapi secara alami tidak ada titik nol dimana suhu atau temperatur ini dimulai. Suhu 0 derajat tidak berarti tidak ada panas. Tidaklah benar mengatakan bahwa suhu badan 40 derajat celsius panasnya 2 kali lipat dari suhu badan 20 derajat celsius.
2. **TAHUN:** tahun 542, 1000, 2000, 2008 merupakan data dalam level interval. Data ini dapat diurutkan dan dapat diketahui selisih antara 2 tahun sebarang, namun ia tidak ada titik nol alami. Artinya, waktu tidak dimulai dari tahun 0 dan tahun 0 hanya sebagai titik nol buatan manusia sebagai ganti titik nol alami yang menyatakan “tidak ada waktu”.

LEVEL RASIO seperti level interval namun ia mempunyai titik nol alami sebagai titik awal. Data dari level rasio, data data dapat dibandingkan (selisih) dan dirasioakan (pembagian).

CONTOH: Berikut inidata dalam level rasio

1. **HARGA:** harga-harga buku teks mahasiswa merupakan data level rasio dimana harga 0 rupiah menunjukkan tidak ada harga alias gratis.
2. **BOBOT:** berat badan manusia merupakan data level rasio dimana berat 0 kg menyatakan tidak ada bobot.
3. **Indeks persepsi korupsi (IPK):** ketika belum diranking, IPK yang dikeluarkan oleh TII masih dalam bentuk skor skala 10 dengan ketelitian 2 digit dibelakang koma, misalnya Yogyakarta dengan IPK 6.43, Palangkaraya dengan IPK 6.10, Banda Aceh dengan IPK 5.87 dan seterusnya Kendari dengan IPK 3.39, terkecil Kupang dengan IPK 2.97. Di sini nilai 0 menunjukkan kriteria terkorup “di dunia dan akhirat”

SOAL-SOAL LATIHAN:

1. [STATISTIK dan PARAMETER] Identifikasilah apakah nilai (angka) berikut sebagai parameter atau statistik.
 - a. Dewan Perwakilan Rakyat (DPR) saat ini terdiri dari 150 perempuan dan 350 pria.
 - b. Sebuah sampel mahasiswa dipilih diperoleh bahwa rata-rata waktu belajar mandiri mereka dalam seminggu adalah 15.2 jam.
 - c. Dalam mempelajari semua penumpang Titanic yang berjumlah 2223 orang, ditemukan 706 orang selamat pada saat kapal tenggelam.

2. [DATA KONTINU dan DATA DISKRIT] Bedakan apakah nilai (angka berikut) sebagai data kontinu atau data diskrit.
 - a. Gaji yang diperoleh oleh pekerja Indonesia di luar negeri mencapai 3.000.000,- rupiah setiap bulannya.
 - b. Dalam 1560 orang pria yang disurvei ditemukan 38% dari mereka adalah perokok aktif.
 - c. Suatu sampel terdiri dari sejumlah mobil, ditemukan bahwa rata-rata beratnya adalah 1500 kg.

3. [LEVEL PENGUKURAN] Tetapkan level yang paling cocok (nominal, ordinal, interval, rasio) untuk pengukuran berikut.
 - a. Tinggi badan pemain sepak bola.
 - b. Temperatur saat ini di dalam kelas.
 - c. Rating suatu acara televisi: “fantastik, baik, cukup, kurang, tidak diterima”.
 - d. Nomor punggung pemain basket.
 - e. Nomor telepon pada buku telepon.
 - f. Majalah konsumen yang memberikan rating: “*best buy, recommended, not recommended*”.
 - g. Kode pos.

4. [SAMPEL – POPULASI] Tentukan yang mana sampel dan yang mana populasinya. Tentukan juga sampel mana yang paling mungkin sebagai representasi dari populasinya.
 - a. A reporter for *Newsweek* stands on a street corner and asks 10 adults if they feel that the current president is doing a good job.
 - b. Nielsen Media Research surveys 5000 randomly selected households and finds that among the TV sets in use, 19% are tuned to *60 Minutes* (based on data from *USA Today*).
 - c. In a Gallup poll of 1059 randomly selected adults, 39% answered “yes” when asked “Do you have a gun in your home?”
 - d. A graduate student at the University of Newport conducts a research project about how adult Americans communicate. She begins with a survey mailed to 500 of the adults that she knows. She asks them to mail back a response to this question: “Do you prefer to use e-mail or snail mail (the U.S. Postal Service)?” She gets back 65 responses,

5. [INTERPRETASI POOLING POLITIK] Misalkan sebuah lembaga survey meminta 200 orang responden tentang preferensi atau pilihan partai politik. Andaikan ada 4 parpol, masing-masing diberi nilai 0 (untuk partai ZERO), 1 (untuk partai ONE), 2 (untuk partai TWO) dan 3 (untuk partai THREE). Berdasarkan hasil survey, diperoleh nilai rata-rata pemilih adalah 0.95. Berikan interpretasi terhadap angka ini ?

6. [SKALA RATING MAKANAN] A group of students develops a scale for rating the quality of the cafeteria food, with 0 representing “neutral: not good and not bad.” Bad meals are given negative numbers and good meals are given positive numbers, with the magnitude of the number corresponding to the severity of badness or goodness. The first three meals are rated as 2, 4, and -5. What is the level of measurement for such ratings? Explain your choice.

