

PENDAHULUAN

Pada dasarnya, Statistika sudah diterapkan jauh ke belakang. Sebagai contoh, bangsa-bangsa di belahan dunia ini menggunakan data statistik untuk mendapatkan informasi deskriptif mengenai pajak, hasil pertanian, perang, kekuatan militer, bahkan sampai hasil pertandingan olah raga. Pada perkembangannya, dengan menerapkan teori peluang, statistika dapat digunakan untuk meneropong jauh di luar data yang telah dikumpulkan dan hasilnya dapat digunakan sebagai pengambilan keputusan atau pengambilan kebijakan melalui proses yang disebut sebagai prediksi,

Secara umum, **Statistika** merupakan suatu ilmu yang mempelajari tentang proses pengumpulan, penyajian, analisis, dan penafsiran data. **Statistika** juga dapat diartikan sebagai suatu ilmu yang mempelajari tentang bagaimana caranya mendapatkan suatu statistik. Sedangkan **statistik** dapat diartikan sebagai:

1. Suatu angka yang diperoleh dari hasil pengamatan/ pengukuran suatu karakteristik atau sifat dari sekumpulan benda tertentu
2. Suatu penduga parameter dari sebuah populasi

Secara luas, **statistika** adalah ilmu yang mempelajari dan mengusahakan agar data mempunyai makna dan **statistik** adalah hasil yang didapatkan berdasarkan proses statistika.

ISTILAH DALAM STATISTIKA

Ada beberapa istilah yang akan sering muncul ketika kita mempelajari statistika, yaitu:

1. **Populasi** merupakan kumpulan/ himpunan dari keseluruhan elemen yang menjadi obyek penelitian/penyelidikan.
2. **Sampel** merupakan bagian dari populasi
3. **Parameter** adalah suatu rangkuman berbentuk angka yang digunakan untuk mendeskripsikan/ menggambarkan sifat suatu populasi, misalnya nilai rata-rata (μ), simpangan baku (σ)
4. **Statistik** adalah suatu rangkuman numerik yang digunakan untuk menggambarkan suatu sampel, nilainya beragam sesuai dengan sampel acak yang telah terpilih, misalnya rata-rata (\bar{x}), simpangan baku (s).

DUA KELAS BESAR DALAM STATISTIKA

Terdapat dua kelas besar dalam statistika, yaitu:

1. Statistika Deskriptif (Exploratory Data Analysis)

Bidang statistika yang membahas cara atau metode mengumpulkan, menyederhanakan dan menyajikan data sehingga bisa memberikan informasi. Dalam statistika deskriptif belum sampai pada upaya menarik kesimpulan, tetapi baru sampai pada tingkat memberikan suatu bentuk ringkasan data sehingga masyarakat awam tentang statistika dapat memahami informasi yang terkandung dalam data.

2. Statistika Inferensia (Confirmatory Data Analysis)

Bidang statistika yang membahas cara-cara atau metode menarik kesimpulan dan mengukur keandalan kesimpulan tentang suatu populasi berdasarkan informasi yang bersumber dari suatu sampel.

URGENCY MEMPELAJARI STATISTIKA

Melalui statistika, data akan diolah menjadi informasi yang akan dijadikan sebagai bahan untuk membuat keputusan. Sehingga, sangat penting untuk mempelajari statistika, terutama dalam penelitian. Dalam sebuah penelitian, statistika sangat bermanfaat untuk:

1. Mengatur/ mengorganisasi data
2. Mendeskripsikan data
3. Membuat kesimpulan berdasarkan data

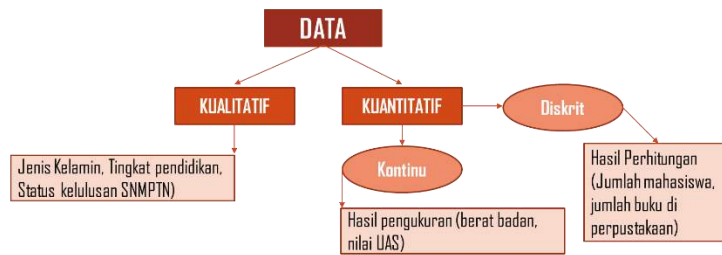
KETERBATASAN STATISTIKA

Disamping terdapat manfaat, statistika juga memiliki keterbatasan antara lain:

1. Statistik tidak sesuai untuk mengkaji fenomena kualitatif
2. Statistik tidak mengkaji individu
3. Statistik tidak menghasilkan kesimpulan yang tepat (exact)
4. Statistik hanya merupakan salah satu metode untuk mengkaji suatu masalah

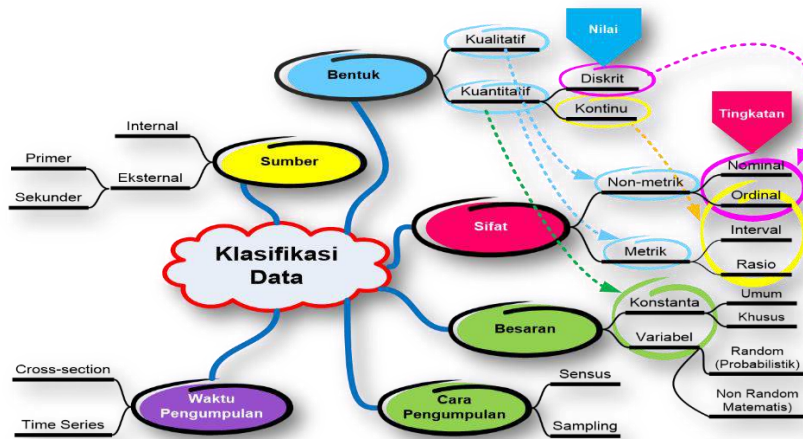
DATA

Data merupakan sekumpulan keterangan/ fakta yang berasal dari hasil pengukuran atau observasi yang dapat berupa sifat atau bilangan. Secara garis besar, data dapat dikelompokkan menjadi dua, yaitu kualitatif dan kuantitatif.



Gambar 1. Garis Besar Pengelompokan Data

Secara spesifik, klasifikasi data dapat dilihat pada gambar di bawah ini.



Gambar 2. Klasifikasi Data

VARIABEL DAN SKALA PENGUKURAN

Variabel/peubah adalah suatu ciri atau gejala suatu populasi atau sampel yang dapat mencapai lebih dari satu nilai. Proses pemberian simbol berupa angka terhadap setiap variasi karakteristik suatu variabel dari masing-masing unit yang diamati dalam statistik disebut pengukuran.

Jenis variabel terbagi menjadi variabel kualitatif (kategori) dan variabel kuantitatif. Sementara berdasarkan hasil pengukuran, variabel dibedakan menjadi variabel diskrit dan variabel kontinu. Adapun skala pengukuran yang digunakan untuk mengukur suatu variabel adalah:

1. Skala Nominal

Angka-angka yang disajikan pada skala nominal hanya sebagai nama penggolongan. Angka tersebut tidak mengukur besaran, tetapi hanya sebagai lambing.

2. Skala Ordinal

Suatu peubah ordinal adalah peubah nominal yang penggolongannya dapat digunakan menyusun suatu urutan yang bermakna.

3. Skala Interval

Angka-angka yang disajikan mempunyai sifat ordinal. Kelebihan dari skala ordinal, pada skala interval, dua angka yang berurutan memiliki jarak yang sama tapi angka-angka tersebut tidak dapat diperbandingkan.

4. Skala Rasio

Skala rasio merupakan skala pengukuran tertinggi. Pada skala rasio, data mempunyai sifat skala selang. Kelebihan pada skala rasio adalah nilai-nilai pada skala rasio dapat dibandingkan karena mempunyai nilai dasar yang mutlak.

Tabel 1. Perbedaan Skala Pengukuran

Skala	Ciri/ Karakteristik			
	Indikasi perbedaan	Indikasi arah perbedaan	Indikasi besaran perbedaan	Nol mutlak
Nominal	√			
Ordinal	√	√		
Interval	√	√	√	
Rasio	√	√	√	√

PROBABILITAS

1. Probabilitas merupakan besarnya kesempatan (kemungkinan) suatu peristiwa akan terjadi.
2. Probabilitas biasanya dinyatakan dalam angka antara 0 dan 1 yang menyatakan kemungkinan terjadinya atau tidaknya suatu peristiwa.
3. Probabilitas bekerja dari populasi menuju sample, sementara statistik sebaliknya.

2. Median

Median merupakan suatu nilai tengah yang membagi data menjadi dua bagian yang sama besar. Median diperoleh dengan cara mengurutkan terlebih dahulu nilai observasi yang diperoleh dari terkecil ke terbesar. Cara menentukan letak dan nilai median adalah:

$$\text{Letak median} = \frac{n + 1}{2}$$

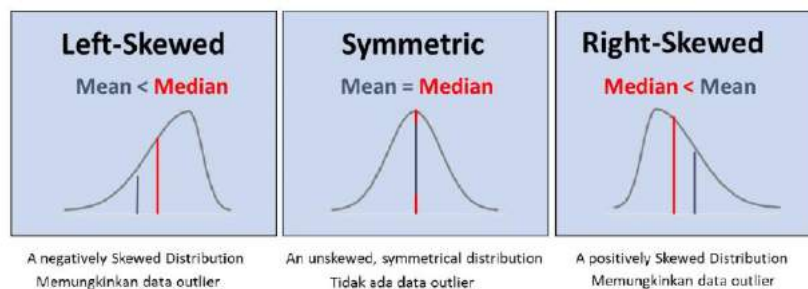
$$\text{Nilai Median} = X_{\text{letak median}}$$

- Jika letak median sesuai dengan urutan data, maka nilai median adalah sama dengan nilai diurutan tersebut.
- Jika letak median berada di antara dua urutan data, maka nilai median dicari dengan cara menjumlahkan nilai diantara letak median dan dibagi 2.

3. Modus

Modus adalah salah satu ukuran pemusatan data yang menunjukkan nilai observasi yang paling sering muncul atau menunjukkan nilai observasi yang memiliki frekuensi terbanyak

Berdasarkan mean, median, dan modus, kita dapat menentukan distribusi suatu data, yaitu:



Contoh 1.

Hitunglah ukuran pusat dari data berikut: 10, 9, 7, 8, 7, 6, 5, 9, 9, 9

Jawab:

Ukuran pusat ada tiga, yaitu mean, median dan modus.

a. mean

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} = \frac{10 + 9 + 7 + 8 + 7 + 6 + 5 + 9 + 9 + 9}{10} = \frac{79}{10} = 7.9$$

Cara dengan ms. excel gunakan fungsi “average”

b. Median

$$\text{Letak median} = \frac{n + 1}{2} = \frac{10}{2} = 5.5$$

Nilai median:

Urutan data: 5, 6, 7, 7, 8, 9, 9, 9, 9, 10

Nilai median di 5.5 adalah antara data ke-5 dan ke-6 -> $(8 + 9)/2 = 8.5$

Cara dengan ms. excel gunakan fungsi “quartile.exc”

c. Modus

Berdasarkan frekuensi nilai, maka yang menjadi modus adalah 9

Note:

Dalam suatu data bisa jadi tidak memiliki modus dan memiliki modus lebih dari satu, contohnya:

- Apabila terdapat **10 data yang semuanya mempunyai nilai yang berbeda**, maka **tidak memiliki modus**, seperti data ini 1, 2, 3, 4, 5, 6, 7, 8, 9 tidak mempunyai modus.
- Apabila terdapat 10 data yang seperti ini 10, 9, 8, 8, 8, 8, 5, 9, 9, 9, maka modulusnya adalah 8 dan 9 -
> Excel gunakan “mode.mult”

UKURAN LETAK

Untuk menentukan ukuran letak atau ukuran lokasi, data yang digunakan harus diurutkan terlebih dahulu. Dan, ukuran letak menunjukkan suatu nilai yang menjadi batas pembagian data menjadi beberapa bagian sama besar. Ukuran letak terbagi menjadi tiga, yaitu Kuartil, Desil, dan Persentil.

1. Kuartil

Kuartil, membagi data menjadi 4 bagian yang sama banyak, sehingga setiap bagian mempunyai 25% dari data yang ada.

$$\text{Letak } Q_i = \frac{i(n+1)}{4}$$

$$\text{Nilai } Q_i = \text{Data batas kiri} + \text{desimal letak } Q_i (\text{Data batas kanan} - \text{Data batas kiri})$$

2. Desil

Desil, membagi data menjadi 10 bagian yang sama banyak, sehingga setiap bagian mempunyai 10% dari data yang ada.

$$\text{Letak } D_i = \frac{i(n+1)}{10}$$

$$\text{Nilai } D_i = \text{Data batas kiri} + \text{desimal letak } D_i (\text{Data batas kanan} - \text{Data batas kiri})$$

3. Persentil

Persentil, membagi data menjadi 100 bagian yang sama banyak, sehingga setiap bagian mempunyai 1% dari data yang ada.

$$\text{Letak } P_i = \frac{i(n+1)}{100}$$

$$\text{Nilai } P_i = \text{Data batas kiri} + \text{desimal letak } P_i (\text{Data batas kanan} - \text{Data batas kiri})$$

Hubungan antara Ukuran Letak

- $Median (Q_2) = D_5 = P_{50}$
- $Q_1 = P_{25}$
- $Q_3 = P_{75}$

Contoh 2.

Hitung nilai Q_1, Q_3, D_5, P_{20} dari data berikut: 10, 9, 7, 8, 7, 6, 5, 9, 9, 9

Jawab:

urutan data: 5, 6, 7, 7, 8, 9, 9, 9, 9, 10

Mencari Q_1

$$\text{Letak } Q_1 = \frac{(10 + 1)}{4} = 2.75$$

- 2.75 menunjukkan bahwa nilai kuartil pertama ada di antara data ke-2 (batas kiri) dan ke-3 (batas kanan)
- Desimal dari 2.75 adalah 0.75

Data batas kiri = 6

Data batas kanan = 7

$$\text{Nilai } Q_1 = 6 + 0.75(7 - 6) = 6.75$$

Mencari Q_3

$$\text{Letak } Q_3 = \frac{3(10 + 1)}{4} = 8.25$$

- 8.25 menunjukkan bahwa nilai kuartil ketiga berada di antara data ke-8 (batas kiri) dan ke-9 (batas kanan)
- Desimal dari 8.25 adalah 0.25

Data batas kiri = 9

Data batas kanan = 9

$$\text{Nilai } Q_3 = 9 + 0.25(9 - 9) = 9$$

Mencari D_5

$$\text{Letak } D_5 = \frac{5(10 + 1)}{10} = 5.5$$

- 5.5 menunjukkan bahwa nilai Desil kelima berada di antara data ke-5 (batas kiri) dan ke-6 (batas kanan)
- Desimal dari 5.5 adalah 0.5

Data batas kiri = 8

Data batas kanan = 9

$$\text{Nilai } D_5 = 8 + 0.5(9 - 8) = 8.5$$

Mencari P_{20}

$$\text{Letak } P_{20} = \frac{20(10 + 1)}{100} = 2.2$$

- 2.2 menunjukkan bahwa nilai Persentil ke-20 berada di antara data ke-2 (batas kiri) dan ke-3 (batas kanan)
- Desimal dari 2.2 adalah 0.2

Data batas kiri = 6

Data batas kanan = 7

$$\text{Nilai } P_{20} = 6 + 0.2(7 - 6) = 6.2$$

UKURAN KERAGAMAN

Ukuran keragaman memberikan informasi tentang penyebaran data (penyimpangan data dari ukuran pusatnya).

1. *Range R*/ Jangkauan

Range, merupakan ukuran keragaman yang paling sederhana. Jangkauan ini menunjukkan selisih antara nilai maksimum dan minimum dari data observasi.

$$R = X_{max} - X_{min}$$

2. *Interquartile Range (IR)*/ Jangkauan antar Kuartil (JAK)

IR menunjukkan selisih antara kuartil ketiga dengan pertama dari data observasi.

$$IR = Q_3 - Q_1$$

3. Variansi dan Simpangan baku.

- Variansi (*variance*) adalah Rata-rata kuadrat penyimpangan data dari mean.
- Simpangan baku (*standard deviation*) adalah akar dari varians.
- Simpangan baku yang paling sering digunakan sebagai indikator yang menunjukkan keragaman data

Varians

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}$$

Simpangan Baku

$$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}}$$

Keterangan:

X_i : nilai pada data ke-i

\bar{X} : rata-rata data

n : banyaknya sampel/ data

4. Koefisien Variasi

- Merupakan ukuran keragaman relatif
- Selalu dalam bentuk persentase (%)
- Digunakan untuk melihat sebaran/distribusi data dari rata-rata hitungannya.
- Semakin kecil koefisien variasi, data semakin homogen

- Semakin besar koefisien variasi, data heterogeny

$$CV = \frac{S}{\bar{X}} \times 100\%$$

5. *Outlier/ Pencilan*

Pencilan menunjukkan nilai ekstrim dari data. Cara menentukan pencilan adalah sebagai berikut:

- Menentukan jangkauan antar kuartil (JAK) data
- Menentukan pagar dalam

$$PD_1 = Q_1 - \frac{3}{2}JAK$$

$$PD_2 = Q_3 + \frac{3}{2}JAK$$

- Menentukan pagar luar

$$PL_1 = Q_1 - 3.JAK$$

$$PL_2 = Q_3 + 3.JAK$$

- Apabila ada beberapa datum yang berada di antara PL_1 dan PD_1 atau PD_2 dan PL_2 , maka datum tersebut merupakan pencilan

Contoh 3.

Terdapat data: 10, 9, 7, 8, 7, 6, 5, 9, 9, 9

Carilah:

- R
- JAK
- Simpangan baku
- Koefisien Variasi
- Apakah ada outlier?

Jawaban:

- Range*

$$R = 10 - 5 = 4$$

- JAK

$$IR = Q_3 - Q_1 = 9 - 6.75 = 1.25$$

- Simpangan baku

$$\bar{X} = 7.9$$

$$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}}$$

$$= \sqrt{\frac{(5 - 7.9)^2 + (6 - 7.9)^2 + 2 \times (7 - 7.9)^2 + (8 - 7.9)^2 + 4 \times (9 - 7.9)^2 + (10 - 7.9)^2}{10 - 1}}$$

$$= \sqrt{\frac{22.9}{9}} = 1.5951$$

d. Koefisien Variasi

$$CV = \frac{S}{\bar{X}} \times 100\% = \frac{1.5951}{7.9} \times 100\% = 20.19\%$$

Koefisien variasi yang dihasilkan sebesar 20.19%, menunjukkan jika sebesar 20.19% data menyimpang dari rata-ratanya

e. Deteksi Pencilan

- Pagar dalam

$$PD_1 = Q_1 - \frac{3}{2}IR = 6.75 - 1.5(1.25) = 4.875$$

$$PD_2 = Q_3 + \frac{3}{2}IR = 9 + 1.5(1.25) = 10.875$$

- Pagar luar

$$PL = Q_1 - 3IR = 6.75 - 3(1.25) = 3$$

$$PL_2 = Q_3 + 3IR = 9 + 3(1.25) = 12.75$$

Karena tidak ada datum yang berada di antara PL_1 dan PD_1 atau PD_2 dan PL_2 , maka tidak ada pencilan pada data di atas

SOAL LATIHAN

Diberikan data usia sebagai berikut: 2, 9, 4, 5, 4, 3, 2, 4, 3, 3. Tentukanlah:

- Ukuran Pusat (Mean, median, modus)
- Ukuran Letak (Desil 3)
- Ukuran keragaman (Koefisien Variasi)
- Outlier (jika ada)

Jawaban:



Lembar Jawaban

