

# **BAB I**

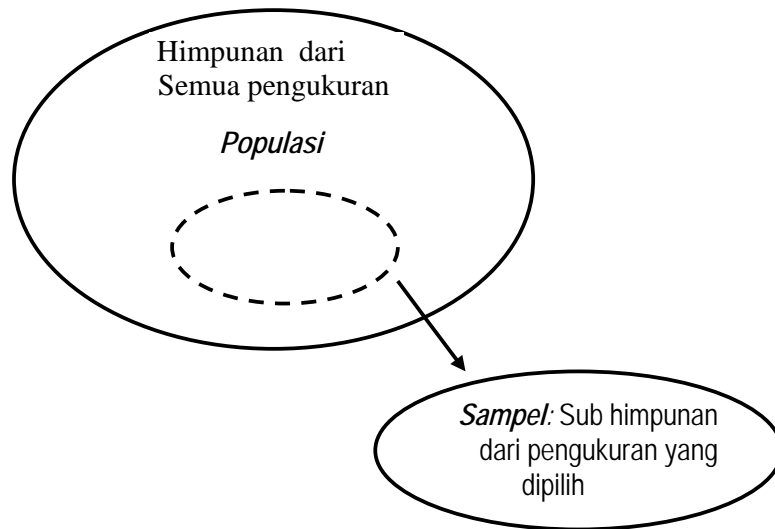
## **PENGANTAR STATISTIK DAN ANALISIS DATA**

### **1.1. Pengertian: Statistik inferensial, Sampel, Populasi, Disain eksperimen**

Pada awal tahun 1980 dan berlanjut sampai abad 21, industri di Amerika menekankan tentang **perbaikan kualitas**. Hal tersebut diilhami oleh kemajuan industri Jepang yang sangat pesat pada pertengahan abad 20. Keberhasilan industri di Jepang didasarkan pada **penggunaan metode statistik** dan **pola pikir statistik** pada personil manajemen perusahaan.

Penggunaan metode statistik bukanlah hal yang baru dalam industri, khususnya dalam kaitannya dengan pengumpulan informasi/data atau **data saintifik**. Terdapat perbedaan mendasar antara pengumpulan informasi saintifik dengan statistik inferensial. **Statistik inferensial** digunakan dalam proses mengambil keputusan dalam menghadapi ketidakpastian dan perubahan. Contoh ketidakpastian adalah kuat tekan beton dalam suatu pengujian tidak sama, walaupun dibuat dengan material yang sama. Dengan adanya kenyataan tersebut, maka metode statistik digunakan untuk menganalisis data dari suatu proses pembuatan beton tersebut sehingga diperoleh kualitas yang lebih baik. Statistik inferensial telah menghasilkan banyak metode analitis yang digunakan untuk menganalisis data. Dengan perkataan lain statistik inferensial tidak hanya mengumpulkan data, tetapi juga mengambil kesimpulan dari suatu sistem saintifik.

Informasi dikumpulkan dari suatu **sampel** atau kumpulan dari suatu **pengamatan** (observasi). Sedangkan **sampel** diambil dari **populasi** yang merupakan kumpulan (himpunan) yang mewakili semua pengukuran.



Contoh, sebuah perusahaan komputer berupaya menghilangkan kerusakan. Perusahaan mengambil 50 sampel komputer secara acak dari suatu proses. Disini, populasi adalah seluruh komputer yang diproduksi oleh perusahaan tersebut pada periode waktu tertentu. Setelah dilakukan perbaikan dalam proses produksi, perusahaan tersebut mengambil kembali 50 sampel. Kemudian dianalisis seberapa besar pengaruh perbaikan proses produksi terhadap pengurangan tingkat kerusakan komputer.

Terkadang seseorang meneliti hanya karakteristik tertentu dari objek yang diteliti. Misalkan, seorang insinyur ingin meneliti pengaruh kondisi proses, temperatur, kelembaban, banyaknya material tertentu terhadap **disain eksperimen** yang diinginkan. Dalam beberapa kasus penelitian tidak diperlukan disain eksperimen. Misal, seorang ingin meneliti faktor yang mempengaruhi kepadatan kayu dari suatu pohon. Dalam kasus ini yang dibutuhkan adalah **studi observasi** (pengamatan) langsung di lapangan karena faktor-faktor yang ada tidak bisa dipilih sebelumnya.

Kadang kalangan praktisi hanya ingin memperoleh beberapa jenis kesimpulan dari sampel data. Seperti: ukuran lokasi data (rata-rata, median, standar deviasi), variabilitas, distribusi, dll. Hal ini disebut dengan **statistik deskriptif**.

## 1.2 Prosedur sampling: pengumpulan data

Prosedur sampling adalah menentukan bagaimana sebuah sampel akan dipilih. *Simple random sampling* berarti setiap sampel tertentu dari ukuran sampel yang telah ditentukan memiliki peluang yang sama untuk dipilih. Bila sampel yang dipilih tidak memiliki peluang yang sama maka hasilnya akan bias. Sehingga sampel tersebut dinamakan **sampel yang bias** (*biased sample*). Contoh bila seseorang ingin meneliti tentang partai politik yang akan dipilih dan ia membuat kuesioner kepada 1000 sampel. Namun 1000 sampel tersebut hanya diambil diwilayah perkotaan, maka kemungkinan besar hasilnya tidak mencerminkan realitas yang ada. Karena pemilih perkotaan berbeda dengan pedesaan.

## 1.3 Ukuran lokasi: Rata-rata (mean) dan Nilai tengah (median) dari sampel

Rata-rata (mean) dari sampel dinyatakan sebagai:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

dimana  $n$  = jumlah pengukuran-pengukuran sampel

Contoh: Tentukan rata-rata dari pengukuran-pengukuran 2, 9, 11, 5, 6

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{2+9+11+5+6}{5} = 6,6$$

**Median** dari himpunan pengukuran  $x_1, x_2, x_3, x_4, \dots, x_n$  didefinisikan sebagai nilai dari  $x$  yang jatuh ditengah-tengah jika pengukuran-pengukuran disusun sesuai urutan besarnya. Jika jumlah pengukuran genap, kita pilih median sebagai nilai  $x$  yang terletak di tengah antara dua pengukuran-pengukuran tengah.

Contoh: tinjaulah pengukuran-pengukuran sampel sbb: 9, 2, 7, 11, 14.

Jika disusun dalam urutan besarnya 2, 7, 9, 11, 14. Maka dipilih 9 sebagai median.

Contoh: tinjaulah pengukuran-pengukuran sampel sbb: 9, 2, 7, 11, 14. 6

Jika disusun dalam urutan besarnya 2, 6, 7, 9, 11, 14. Maka kita memilih median sebagai nilai tengah antara 7 dan 9, yaitu 8.

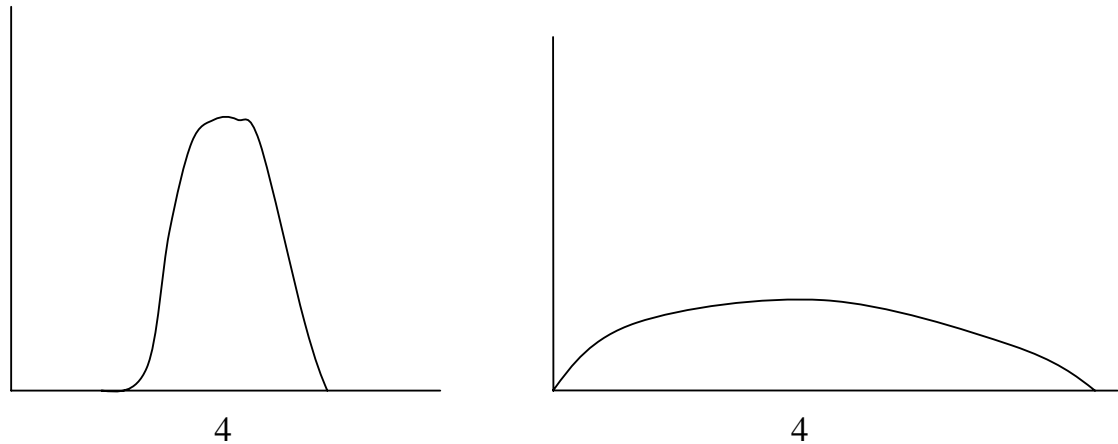
**Modus (*mode*)** dari himpunan  $n$  pengukuran-pengukuran  $x_1, x_2, x_3, x_4, \dots, x_n$  didefinisikan sebagai nilai dari  $x$  yang tampil dengan frekuensi tertinggi.

Contoh: tinjaulah pengukuran-pengukuran sampel sbb: 9, 2, 7, 11, 14. 7, 2, 7.

Karena 7 tampil tiga kali (paling banyak), maka modus adalah 7.

#### 1.4 Ukuran Perubahan (Variabilitas)

Langkah penting lainnya adalah menentukan ukuran perubahan (variabilitas) atau penyebaran (dispersi).



Gambar. Variabilitas atau dispersi dari data

Gambar diatas menunjukkan kedua distribusi memiliki median yang sama tapi terdapat perbedaan sangat besar dalam variabilitas pengukuran atas mean. Variasi (perbedaan) adalah suatu ukuran yang sangat penting dari data. Jika kita memproduksi alat yang butuh tingkat akurasi tinggi, maka variabilitas harus kecil sehingga sedikit produk yang cacat.

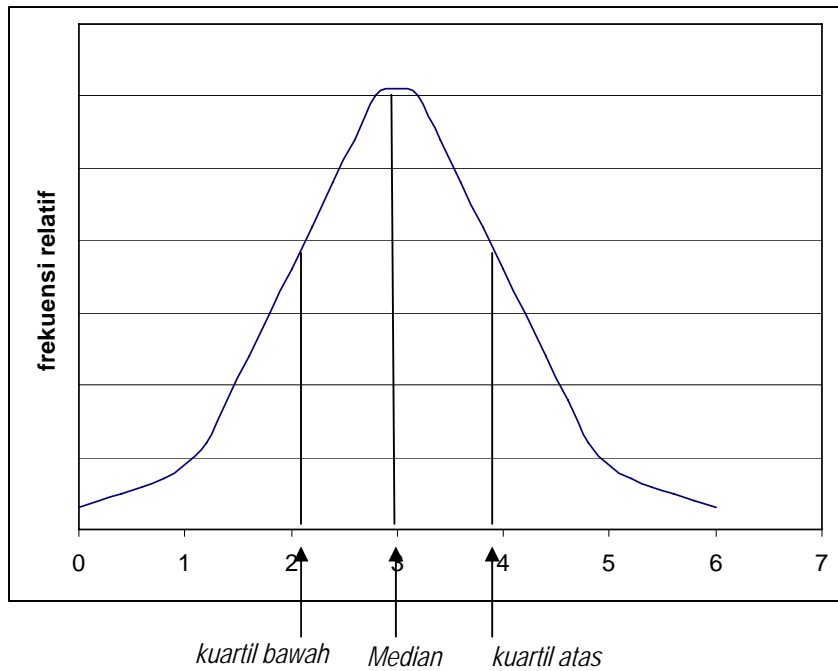
Ukuran paling sederhana dari variasi adalah rentang (*range*). **Rentang** dari himpunan pengukuran-pengukuran  $x_1, x_2, x_3, x_4, \dots, x_n$  didefinisikan sebagai beda (selisih) antara pengukuran terbesar dan pengukuran yang terkecil.

Contoh: bila dari hasil pengukuran diperoleh nilai 3, 4, 5, 9, 11, 2, 13; maka rentangnya adalah  $13-2 = 11$ .

Rentang memiliki keterbatasan dalam merepresentasikan ukuran dari variasi, hal ini disebabkan dengan rentang yang sama sangat mungkin variasinya berbeda. Untuk mengatasi keterbatasan rentang diperkenalkan istilah **kuartil** dan

**persentil.**

Bila  $x_1, x_2, x_3, x_4, \dots, x_n$  adalah himpunan dari  $n$  pengukuran yang disusun dalam urutan besarnya. **Kuartil bawah** adalah nilai dari  $x$  yang melebihi  $\frac{1}{4}$  dari pengukuran-pengukuran dan lebih kecil dari sisanya yang  $\frac{3}{4}$ . **Kuartil kedua** adalah **median**. **Kuartil atas** (kuartil ketiga) adalah nilai dari  $x$  yang  $\frac{3}{4}$  dari pengukuran-pengukuran dan lebih kecil dari  $\frac{1}{4}$ .



Jika data yang tersedia banyak lebih baik menggunakan **persentil**. Bila  $x_1, x_2, x_3, x_4, \dots, x_n$  adalah himpunan dari  $n$  pengukuran yang disusun dalam urutan besarnya. Persentil yang ke-  $p$  ialah nilai dari  $x$  dimana paling banyak  $p$  persen dari pengukuran-pengukuran akan lebih kecil dari pada nilai  $x$  dan paling banyak  $(100 - p)$  persen akan lebih besar.

Ukuran yang sering digunakan dalam variabilitas adalah varians dan standar deviasi.

**Varians suatu sampel** dari n pengukuran-pengukuran  $x_1, x_2, x_3, x_4, \dots, x_n$  didefinisikan sebagai jumlah kuadrat deviasi pengukuran terhadap mean  $\bar{x}$  dibagi dengan (n-1) dan dinyatakan dengan rumus berikut:

$$s^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1}$$

n - 1 disebut juga sebagai derajat kebebasan.

**Deviasi standar sampel** adalah akar positif dari varians.

$$s = \sqrt{s^2}$$

Contoh: dari hasil pengukuran terhadap 10 sampel diperoleh hasil sebagai berikut:

7,07 7,00 7,10 6,97 7,00  
7,03 7,01 7,01 6,98 7,08

rata-rata sampel:

$$\bar{x} = \frac{7,07 + 7,00 + 7,10 + \dots + 7,08}{10} = 7,0250$$

Varians:

$$s^2 = \frac{(7,07 - 7,025)^2 + (7,00 - 7,025)^2 + \dots + (7,08 - 7,025)^2}{9} = 0,001939$$

Deviasi standar sampel:

$$s = \sqrt{0,00193} = 0,0440$$

## 1.5 Metode Grafis dan Deskripsi Data

Suatu pemeriksaan diperoleh hasil berikut:

20,5	19,5	15,6	24,1	9,9
15,4	12,7	5,4	24,1	28,6
16,9	7,8	23,3	11,8	18,4
13,4	14,3	19,2	9,2	16,8
8,8	22,1	20,8	12,6	15,9

Bagaimana ke 25 data diatas tersebar dalam interval 5,4 sampai 28,6. Untuk

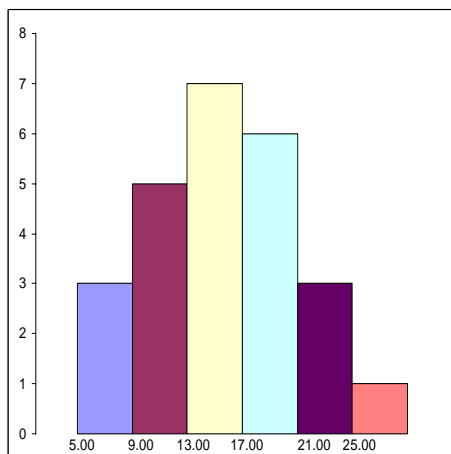
menjawab hal tersebut kita bagi interval diatas menjadi **subinterval** yang sama panjang. Subinterval sering disebut juga sebagai **kelas-kelas**, lazimnya 5 – 20 kelas. Jumlah pengukuran yang masuk dalam kelas tertentu disebut **frekuensi kelas ( $f_i$ )**. *Frekuensi relatif* kelas dinyatakan sebagai:

$$\text{Frekuensi relatif} = \frac{f_i}{n}$$

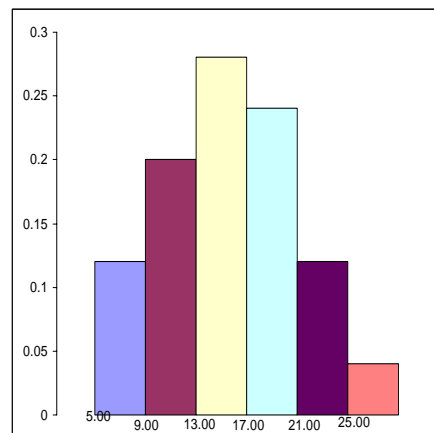
Tabel: frekuensi relatif

kelas (i)	kelas interval	Frekuensi (f)	Frekuensi relatif
1	5,00 - 8,99	3	0.12
2	9,00 - 12,99	5	0.2
3	13,00 - 16,99	7	0.28
4	17,00 - 20,99	6	0.24
5	21,00 - 24,99	3	0.12
6	25,00 - 28,99	1	0.04
Total		25	1

Tabel diatas dapat dinyatakan secara grafik dalam bentuk *histogram frekuensi* dan **histogram frekuensi relatif** (sering disebut sebagai **distribusi frekuensi**).



Gambar: Histogram frekuensi



Gambar: Histogram frekuensi relatif

Distribusi dikatakan simetris bila membagi menjadi dua bagian yang sama. Dan dikatakan menceng bila tidak simetris.



Contoh:

Pengujian 25 buah sampel menghasilkan data sebagai berikut:

5.2	6.0	7.5	8.0	10.0
10.8	10.5	9.2	7.4	6.5
8.0	9.0	12.5	11.3	7.0
11.7	8.5	5.5	9.3	9.5
6.5	7.5	6.5	8.1	11.5

- a. Gambarlah suatu histogram frekuensi relatif
- b. Hitung mean
- c. Hitung varians dan deviasi standar