

# Statistika

## A. PENDAHULUAN

**Statistika** adalah ilmu yang mempelajari pengambilan, penyajian, pengolahan, dan penafsiran data.

**Data** terdiri dari dua jenis, yaitu data kualitatif (sifat) dan data kuantitatif (angka).

## B. PENYAJIAN DATA

**Penyajian data** terdiri dari dua:

- 1) **Penyajian data tunggal**
- 2) **Penyajian data kelompok**

**Data tunggal** dapat disajikan dalam bentuk:

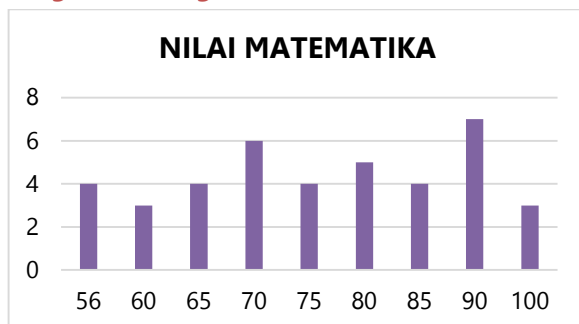
**Berjajar**

56	60	65	75	75	70	75
70	70	70	70	85	85	80
70	60	56	85	85	80	100
90	90	90	90	90	90	65
80	90	100	65	65	80	
56	56	60	75	80	100	

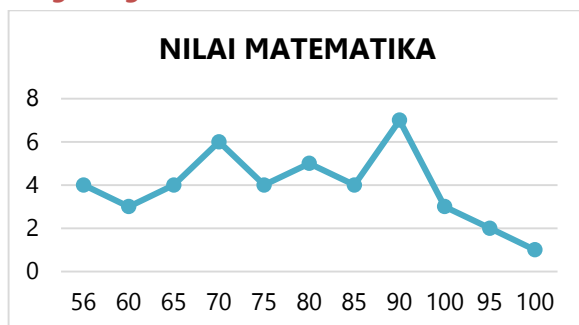
**Tabel distribusi frekuensi**

Nilai	Frekuensi
56	4
60	3
65	4
70	6
75	4
80	5
85	4
90	7
100	3

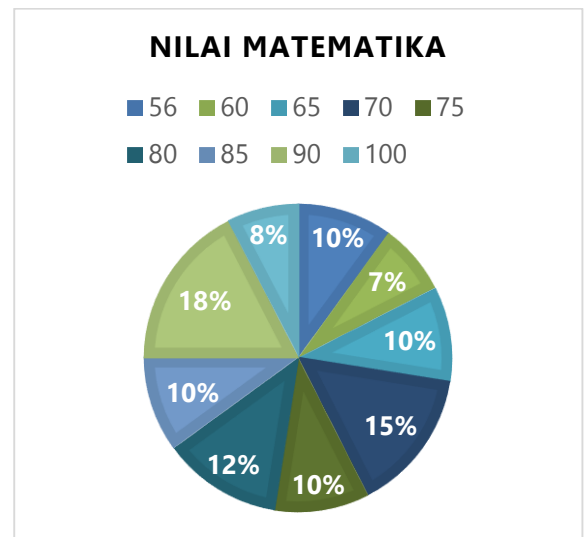
**Diagram batang**



**Diagram garis**



**Diagram lingkaran** (sudut atau presentase)



**Diagram batang-daun**

5	6666
6	0005555
7	0000005555
8	000005555
9	0000000
10	000

**Data tunggal** dapat diubah penyajiannya menjadi data kelompok, dengan cara berikut:

- 1) Penentuan **range/jangkauan** data.

$$R = x \text{ maks} - x \text{ min}$$

x maks = data terbesar  
x min = data terkecil  
R = 100 - 44 = 56

- 2) Penentuan **banyak kelas/kelompok** data yang akan dibuat.

$$k = 1 + 3,3 \cdot \log n$$

n = banyak data  
k = 1 + 3,3 · log 40  
k = 1 + 5,28 = 6,28 ≈ 6

- 3) Penentuan **panjang atau lebar kelas/kelompok**, yaitu interval data dari tiap kelompok.

$$c = \frac{R}{k}$$

c = 56 : 6  
c = 9,33 ≈ 9

Setelah dihitung, data majemuk dapat disajikan dalam bentuk:

#### Tabel distribusi frekuensi kumulatif/kelompok

Nilai	Frekuensi	
56-64	7	3 + 4
65-73	10	4 + 6
74-82	9	4 + 5
83-91	11	4 + 7
92-100	3	

Unsur-unsur dalam penyajian data majemuk berdasarkan pendekatan t.d. frekuensi kumulatif:

- 1) **Batas bawah** ( $B_B$ ), merupakan nilai terkecil dalam suatu interval.
- 2) **Batas atas** ( $B_A$ ), merupakan nilai terbesar dalam suatu interval.

Contoh: Pada interval 65-73, batas bawah adalah 65 dan batas atas adalah 73.

- 3) **Nilai tengah interval**, dengan rumus:

$$M = \frac{B_B + B_A}{2} \quad M = \frac{(65 + 73)}{2} = 69$$

- 4) **Tepi bawah**, dengan rumus:

$$T_B = B_B - \frac{1}{2} \text{ ketelitian data} \quad \begin{aligned} T_B &= 65 - \frac{1}{2} \cdot 1 \\ T_B &= 64,5 \end{aligned}$$

- 5) **Tepi atas**, dengan rumus:

$$T_A = B_A + \frac{1}{2} \text{ ketelitian data} \quad \begin{aligned} T_A &= 73 + \frac{1}{2} \cdot 1 \\ T_A &= 73,5 \end{aligned}$$

- 6) **Panjang kelas**, merupakan panjang interval kelas dengan rumus:

$$c = T_A - T_B \quad \begin{aligned} c &= 73,5 - 64,5 \\ c &= 9 \end{aligned}$$

Bentuk lain tabel distribusi frekuensi kelompok:

#### T.d. frekuensi kumulatif kurang dari ( $\leq$ )

Nilai yang digunakan adalah tepi atas tiap kelas.

Nilai	F. Kumulatif	
$\leq 64,5$	7	
$\leq 73,5$	17	7 + 10
$\leq 82,5$	26	17 + 9
$\leq 91,5$	37	26 + 11
$\leq 100,5$	40	37 + 3

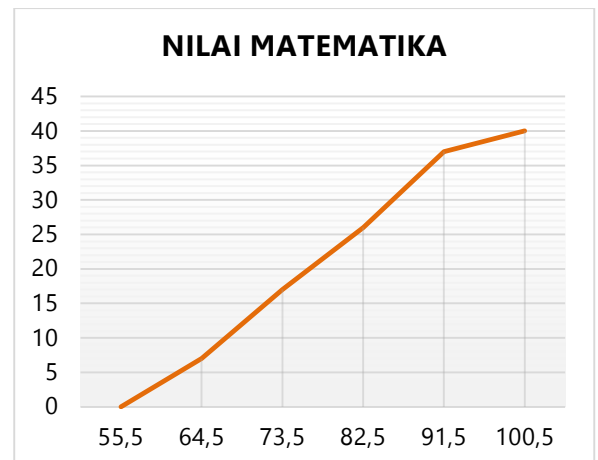
#### T.d. frekuensi kumulatif lebih dari ( $\geq$ )

Nilai yang digunakan adalah tepi bawah tiap kelas.

Nilai	F. Kumulatif	
$\geq 55,5$	40	
$\geq 64,5$	33	40 - 7
$\geq 73,5$	23	33 - 10
$\geq 82,5$	14	23 - 9
$\geq 91,5$	3	14 - 11

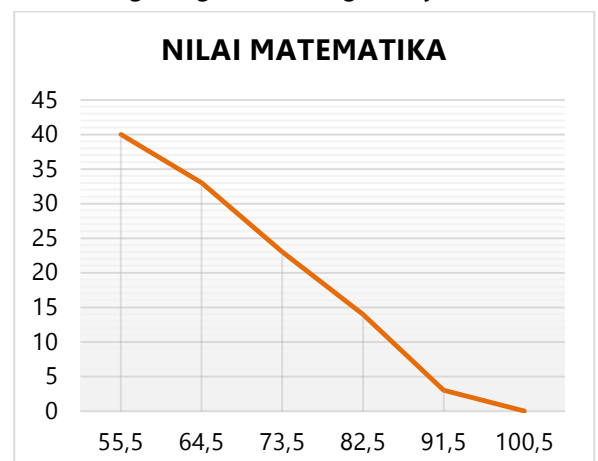
#### Ogif positif

Data yang digunakan untuk ogif positif berasal dari **tabel distribusi kumulatif kurang dari** dengan tambahan tepi bawah dari kelas terendah. Ciri dari ogif positif adalah grafiknya **menaik**.



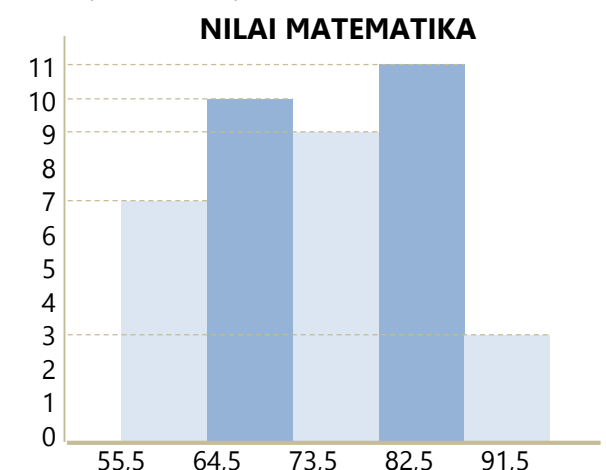
#### Ogif negatif

Data yang digunakan untuk ogif negatif berasal dari **tabel distribusi kumulatif lebih dari** dengan tambahan tepi atas dari kelas tertinggi. Ciri dari ogif negatif adalah grafiknya **menurun**.



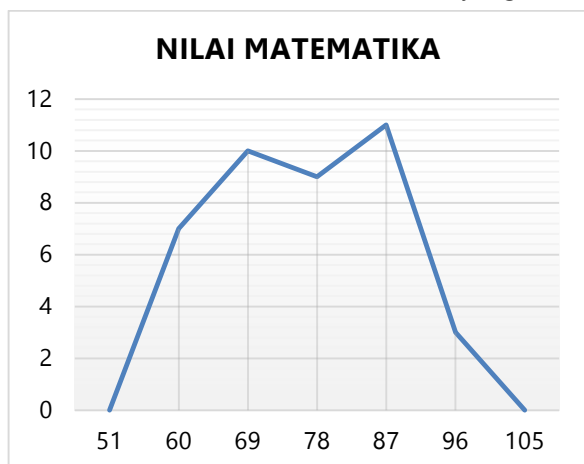
#### Histogram (diagram batang)

Data yang diperlukan histogram adalah tepi atas dan tepi bawah tiap kelas.



### Poligon frekuensi (diagram garis)

Data yang diperlukan poligon frekuensi adalah nilai tengah dari tiap kelas, dan nilai tengah satu kelas sebelum dan sesudah data kelas yang ada.



## C. PENGOLAHAN DATA TUNGGAL

**Pengolahan data tunggal** terdiri dari:

- Ukuran pemusatan data**, terdiri dari mean, modus, dan kuartil.
- Ukuran penyebaran data (dispersi)**, terdiri dari range, hamparan, simpangan kuartil, langkah, pagar luar, pagar dalam, simpangan rata-rata, ragam, dan simpangan baku.

## D. PEMUSATAN DATA TUNGGAL

**Mean** adalah nilai rata-rata hitung seluruh data yang ada.

$$\bar{x} = \frac{\sum x_i}{n} = \frac{\sum x_i \cdot f_i}{\sum f_i}$$

$x_i$  = data  
 $n$  = banyak data  
 $f_i$  = frekuensi data

**Mean** juga dapat dicari dengan nilai rata-rata sementara.

$$\bar{x} = \bar{x}_s + \frac{\sum d_i}{n} = \bar{x}_s + \frac{\sum d_i \cdot f_i}{\sum f_i}$$

$\bar{x}_s$  = rata-rata sementara, diambil dari salah satu data  
 $d_i$  = selisih data dengan rata-rata sementara ( $x_i - \bar{x}_s$ )

Contoh:

Dari data berikut: 114, 114, 115, 117, 117, 117, 119, 120, 121, 125, tentukan mean!

$$\bar{x} = \frac{114+114+115+\dots+125}{10} = 117,9$$

Misalnya jika rata-rata sementara yang dipilih adalah 117, maka:

-3 -3 -2 0 0 0 +2 +3 +4 +8  
 114 114 115 117 117 117 119 120 121 125

$$\bar{x} = 117 + \frac{-3-3-2+0+0+0+2+3+4+8}{10}$$

$$\bar{x} = 117 + \frac{9}{10} = 117,9$$

**Modus** adalah data yang paling sering muncul dari seluruh data yang ada setelah diurutkan.

Contoh: Pada data berikut,

1, 2, 3, 3, 3, 4, 5 modusnya 3.

1, 1, 2, 2, 3, 3, 4 modusnya 1, 2 dan 3.

1, 1, 2, 2, 3, 3 modusnya tidak ada.

**Kuartil** adalah batas-batas nilai yang terdapat pada data apabila sekelompok data telah diurutkan dan dibagi menjadi 4 bagian (3 batas).

**Kuartil** terbagi menjadi tiga:

- Kuartil bawah** ( $Q_1$ ), adalah nilai tengah data pada pertengahan data pertama.
- Kuartil tengah/median** ( $Q_2$ ), adalah nilai tengah seluruh data.
- Kuartil atas** ( $Q_3$ ), adalah nilai tengah data pada pertengahan data terakhir.

**Kuartil tengah/median** dapat ditentukan dengan rumus:

**Data ganjil**

(mediannya terletak pada satu data)

$$Q_2 = x \text{ ke } \frac{n+1}{2}$$

**Data genap**

(median terletak di antara dua data)

$$Q_2 = \frac{1}{2} [(x \text{ ke } \frac{n}{2}) + (x \text{ ke } \frac{n}{2} + 1)]$$

**Kuartil atas dan kuartil bawah** dapat ditentukan dengan rumus:

**Data ganjil**

$$Q_1 = x \text{ ke } \frac{1}{4} (n+1)$$

$$Q_3 = x \text{ ke } \frac{3}{4} (n+1)$$

**Data genap**

$$Q_1 = x \text{ ke } \frac{1}{4} (n+2)$$

$$Q_3 = x \text{ ke } \frac{3}{4} (n+2) - 1$$

**Batas-batas nilai** lain yang memiliki konsep sama dengan kuartil:

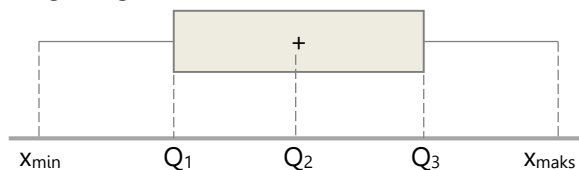
- Desil**, membagi data menjadi **10 bagian** (9 batas) dengan desil ke 5 sebagai median.

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

- Persentil**, membagi data menjadi **100 bagian** (99 batas), dengan persentil ke 50 sebagai median.

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

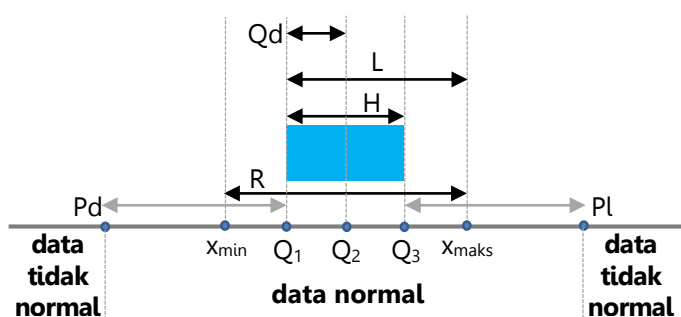
- Statistik lima serangkai adalah penyajian data berupa **diagram garis-kotak** atau **tabel** yang memuat data kuartil, batas bawah, dan batas atas. Diagram garis-kotak



Tabel

	Q <sub>2</sub>	
Q <sub>1</sub>		Q <sub>3</sub>
X <sub>min</sub>		X <sub>maks</sub>

## E. PENYEBARAN DATA TUNGGAL



- Range** adalah jangkauan dari seluruh data.

$$J = X_{maks} - X_{min}$$

- Hamparan** adalah **jangkauan antarkuartil** yang merupakan selisih kuartil atas dengan kuartil bawah.

$$H = Q_3 - Q_1$$

- Simpangan kuartil** adalah setengah dari hamparan.

$$Q_d = \frac{1}{2} H$$

- Langkah** adalah satu setengah kali dari hamparan.

$$L = \frac{3}{2} H$$

- Pagar dalam** adalah satu langkah dibawah kuartil bawah.

$$P_d = Q_1 - L$$

- Pagar luar** adalah satu langkah di atas kuartil atas.

$$P_l = Q_3 + L$$

- Pagar dalam dan pagar luar** berfungsi sebagai patokan untuk menyatakan suatu data normal atau tidak normal.

- Jika suatu data berada di luar pagar, maka data tersebut **tidak normal** atau **menyimpang** (sangat berbeda dari data yang lain).

- Simpangan rata-rata** adalah penyebaran dari nilai rata-rata.

$$S_R = \frac{\sum |x_i - \bar{x}|}{n} = \frac{\sum |x_i - \bar{x}| \cdot f_i}{\sum f_i}$$

- Ragam/varian** adalah jumlah kuadrat dari deviasi nilai-nilai data terhadap rata-rata.

$$R = S^2 = \frac{\sum (x_i - \bar{x})^2}{n} = \frac{\sum (x_i - \bar{x})^2 \cdot f_i}{\sum f_i}$$

- Simpangan baku/standar deviasi** adalah akar kuadrat dari ragam yang menunjukkan homogenitas kelompok.

$$S = \sqrt{R} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} = \sqrt{\frac{\sum (x_i - \bar{x})^2 \cdot f_i}{\sum f_i}}$$

Makin kecil nilai simpangan baku maka datanya makin homogen.

- Pada pengolahan data tunggal**, jika setiap data dikali/dibagi a dan/atau ditambah/dikurang b:

- 1) **Ukuran pemusatan data** berubah **sesuai urutan perubahan data** yang terjadi.

Contoh:

Jika setiap data berikut: 2, 2, 4, 4, 6, 7, 8, 10 ditambah satu, kemudian dikali dua, maka rata-ratanya menjadi?

Pembuktian:

Rata-rata awal:

$$\bar{x} = \frac{2+2+4+4+6+7+8+10}{8} = 5,375$$

Perubahan data menjadi:

2, 2, 4, 4, 6, 7, 8, 10

3, 3, 5, 5, 7, 8, 9, 11 ditambah 1

6, 6, 10, 10, 14, 16, 18, 22 dikali 2

Rata-rata setelah perubahan:

$$\bar{x}' = \frac{6+6+10+10+14+16+18+22}{8} = 12,75$$

Nilai rata-rata 12,75 didapat dari:

$$\bar{x}' = (\bar{x} + 1) \times 2 = (5,375 + 1) \times 2$$

$$\bar{x}' = 12,75$$

- 2) **Ukuran penyebaran data** selain ragam hanya berubah sesuai perubahan **dikali/dibagi**.

Contoh:

Jika setiap data berikut: 2, 2, 4, 4, 6, 7, 8, 10,

a. Jika dikali 2

b. Jika dikali 2 kemudian ditambah 2

c. Jika ditambah 1 kemudian dikali 4

maka jangkauan masing-masingnya adalah?

Pembuktian:

Range awal:

$$J = 10 - 2 = 8$$

a. Perubahan: 4, 4, 8, 8, 12, 14, 16, 20,

$$J' = 20 - 4 = 16$$

(didapat dari  $J' = 2J$ )

b. Perubahan: 6, 6, 10, 10, 14, 16, 18, 22,

$$J' = 22 - 6 = 16$$

(didapat dari  $J' = 2J$ )

c. Perubahan: 12, 12, 20, 20, 28, 32, 36, 44,

$$J' = 44 - 12 = 32$$

(didapat dari  $J' = 4J$ )

- 3) Untuk ragam, hanya berubah sesuai perubahan **dikali/dibagi**, namun **faktornya dikuadratkan** terlebih dahulu sebelum dikali/dibagi.

Contoh:

Jika setiap data berikut: 5, 5, 8, 9, 14, 16, 20, dikali dua, maka ragamnya menjadi?

Pembuktian:

Rata-rata awal:

$$\bar{x} = \frac{5+5+8+9+14+16+20}{7} = 11$$

Ragam awal:

$$R = \frac{(5-11)^2 + (5-11)^2 + (8-11)^2 + \dots + (20-11)^2}{7}$$

$$R = \frac{6^2 + 6^2 + 3^2 + 2^2 + 3^2 + 5^2 + 9^2}{7} = \frac{200}{7}$$

Perubahan data menjadi:

5, 5, 8, 9, 14, 16, 20

10, 10, 16, 28, 32, 40      dikali 2

Rata-rata setelah perubahan:

$$\bar{x}' = 2\bar{x} = 22$$

Ragam setelah perubahan:

$$R' = \frac{(10-22)^2 + (10-22)^2 + (16-22)^2 + \dots + (40-22)^2}{7}$$

$$R' = \frac{12^2 + 12^2 + 6^2 + 4^2 + 6^2 + 10^2 + 18^2}{7} = \frac{800}{7}$$

(didapat dari  $R' = (2)^2 R$ )

## F. PENGOLAHAN DATA MAJEMUK

- Pengolahan data majemuk** pada dasarnya sama dengan data tunggal namun memiliki cara yang berbeda untuk menghitungnya.

## G. PEMUSATAN DATA MAJEMUK

- Mean** dapat dihitung dengan tiga cara:

- 1) **Metode biasa**

$$\bar{x} = \frac{\sum x_i \cdot f_i}{\sum f_i} \quad x_i = \text{nilai tengah tiap kelas}$$

- 2) **Metode simpangan**

$$\bar{x} = \bar{x}_s + \frac{\sum d_i \cdot f_i}{\sum f_i}$$

$\bar{x}_s$  = rata-rata sementara, diambil dari salah satu nilai tengah kelas

$d_i$  = selisih nilai tengah tiap kelas dengan rata-rata sementara ( $\bar{x}_i - \bar{x}_s$ )

- 3) **Metode coding**

$$\mu_i = \frac{d_i}{c} \quad \bar{x} = \bar{x}_s + \frac{\sum \mu_i \cdot f_i}{\sum f_i} \cdot c \quad \begin{matrix} u_i = \text{kode kelas } i \\ c = \text{panjang kelas} \end{matrix}$$

- Modus** terletak pada kelas/interval dengan frekuensi terbanyak.

- Modus** dapat dicari:

$$Mo = T_B + \left( \frac{S_1}{S_1 + S_2} \right) \cdot c$$

$T_B$  = tepi bawah kelas modus

$S_1$  = selisih frekuensi dengan kelas sebelum kelas modus

$S_2$  = selisih frekuensi dengan kelas sesudah kelas modus

$c$  = panjang kelas

- Median, kuartil, desil, persentil** terletak pada kelas yang merupakan batas dari kuartil, desil atau persentil tersebut.

Cara menentukan batas kuartil, desil dan persentil sama dengan caradata tunggal.

- Median** dapat dihitung dengan rumus:

$$Q_2 = T_B + \frac{\frac{1}{2}n - f_{kq_2}}{f_{q_2}} \cdot c$$

$T_B$  = tepi bawah kelas median

$f_{kq}$  = frekuensi kumulatif kelas sebelum kelas median

$f_{q_i}$  = frekuensi kelas median

- Kuartil** dapat dihitung dengan rumus:

$$Q_i = T_B + \frac{\frac{i}{4}n - f_{kq_i}}{f_{q_i}} \cdot c$$

$T_B$  = tepi bawah kelas  $Q_i$

$f_{kq}$  = frekuensi kumulatif kelas sebelum kelas  $Q_i$

$f_{q_i}$  = frekuensi kelas  $Q_i$

- Desil** dapat dihitung dengan rumus:

$$D_i = T_B + \frac{\frac{i}{10}n - f_{kd_i}}{f_{d_i}} \cdot c$$

$T_B$  = tepi bawah kelas  $D_i$

$f_{kd}$  = frekuensi kumulatif kelas sebelum kelas  $D_i$

$f_{d_i}$  = frekuensi kelas  $D_i$


 **Persentil** dapat dihitung dengan rumus:

$$P_i = T_B + \frac{\frac{i}{100} n - f_{kp_i}}{f_{p_i}} \cdot c$$

$T_B$  = tepi bawah kelas  $P_i$

$f_{kp}$  = frekuensi kumulatif kelas sebelum kelas  $P_i$

$f_{p_i}$  = frekuensi kelas  $P_i$

 **Daerah batasan** selain kuartil, desil dan persentil dapat ditentukan melalui persamaan:

$$N = T_B + \frac{x - f_{ks}}{f_k} \cdot c$$

$N$  = nilai tertinggi dari  $x$  data yang pertama

$T_B$  = tepi bawah kelas batasan

$x$  = banyak data daerah sebelum  $N$

$f_{ks}$  = frekuensi kumulatif kelas sebelum kelas batasan

$f_k$  = frekuensi kelas batasan

Contoh:

Diketahui nilai ulangan Matematika suatu kelas:

Nilai	Jumlah murid
60-64	3
65-69	4
70-74	6
75-79	2
80-84	20
85-89	5

Ternyata, guru Matematika kelas tersebut menyatakan 45% murid di kelas tersebut lulus ulangan. Tentukan KKM untuk lulus!

Jawab:

Sementara, kita anggap batas nilai terendah untuk lulus adalah nilai tertinggi dari murid yang tidak lulus.

Jumlah murid tidak lulus =  $55\% \times 40 = 22$  murid

Berarti, batasan terletak pada nilai 80-84.

$$N = 79,5 + \frac{22-15}{20} \times 5$$

$$N = 79,5 + 1,75 = 81,25$$

## H. PENYEBARAN DATA MAJEMUK

 **Range** dapat dirumuskan:

$$J = x_{maks} - x_{min}$$

 **Hamparan** dapat dirumuskan:

$$H = Q_3 - Q_1$$

 **Simpangan kuartil** dapat dirumuskan:

$$Q_d = \frac{1}{2} H$$

 **Simpangan rata-rata** dapat dirumuskan:

$$S_R = \frac{\sum |x_i - \bar{x}| \cdot f_i}{\sum f_i} \quad x_i = \text{nilai tengah tiap kelas}$$

 **Ragam dan simpangan baku** dapat dihitung dengan cara:

### 1) Metode biasa

Ragam

$$R = S^2 = \frac{\sum (x_i - \bar{x})^2 \cdot f_i}{\sum f_i}$$

Simpangan baku

$$S = \sqrt{R} = \sqrt{\frac{\sum (x_i - \bar{x})^2 \cdot f_i}{\sum f_i}}$$

### 2) Metode simpangan

Ragam

$$R = S^2 = \frac{\sum d_i^2 \cdot f_i}{\sum f_i} - \left( \frac{\sum d_i \cdot f_i}{\sum f_i} \right)^2$$

Simpangan baku

$$S = \sqrt{R} = \sqrt{\frac{\sum d_i^2 \cdot f_i}{\sum f_i} - \left( \frac{\sum d_i \cdot f_i}{\sum f_i} \right)^2}$$

### 3) Metode coding

Ragam

$$R = S^2 = \left[ \frac{\sum \mu_i^2 \cdot f_i}{\sum f_i} - \left( \frac{\sum \mu_i \cdot f_i}{\sum f_i} \right)^2 \right] \cdot c$$

Simpangan baku

$$S = \sqrt{R} = \sqrt{\left[ \frac{\sum \mu_i^2 \cdot f_i}{\sum f_i} - \left( \frac{\sum \mu_i \cdot f_i}{\sum f_i} \right)^2 \right] \cdot c}$$