



# Statistika Deskriptif

## *Materi 2 - STK511 Analisis Statistika*

Sep, 2017

- Merupakan teknik penyajian dan peringkasan data sehingga menjadi informasi yang mudah dipahami
- Apa yang disajikan dan diringkas? --> **PEUBAH**

**Univariate** vs **Bivariate** vs **Multivariate**

Type: Kategorik vs Numerik

Numerik: **Diskret** vs **Kontinu**

Skala Pengukuran:

Kategorik : **Nominal** vs **Ordinal**

Numerik: **Interval** vs **Rasio**

# Statistika Deskriptif



- Bagaimana cara menyajikan data?

Tabel

Grafik

- Bagaimana cara meringkas data?

Ukuran Pemusatan

Ukuran Penyebaran



# Penyajian Data

- Tabel disajikan untuk menyajikan **statistik** berdasarkan **observasi** atau **kategori**
- Peubah Kategori: Tabel Frekuensi, Tabel kontingensi
- Peubah Numerik: Dikategorikan

No	JK	Tinggi	Berat	Agama
1	1	167	63	Islam
2	1	172	74	Islam
3	2	161	53	Kristen
4	2	157	47	Hindu
5	1	165	58	Islam
6	2	167	60	Islam
7	1	162	52	Budha
8	2	151	45	Katholik
9	2	158	54	Kristen
10	1	162	63	Islam
11	1	176	82	Islam
12	1	167	69	Islam
13	2	163	57	Kristen
14	2	158	60	Islam
15	1	164	58	Katholik
16	2	161	50	Islam
17	1	159	61	Kristen
18	1	163	65	Islam
19	1	165	62	Islam
20	2	169	59	Islam
21	1	173	70	Islam

# Peubah Kategorik



Frekuensi		Count of Tinggi	JK		
Jenis Kelamin	Total	Agama	1	2	Grand Total
1	12	Budha	1		1
2	9	Hindu		1	1
Grand Total	21	Islam	9	4	13
		Katholik	1	1	2
		Kristen	1	3	4
		Grand Total	12	9	21

Frekuensi			
Agama	Total		
Budha	1		
Hindu	1		
Islam	13		
Katholik	2		
Kristen	4		
Grand Total	21		

Tabel Kontingensi

Tabel Frekuensi

- Menyusun Kategori:

Tentukan jumlah kelas (Sturges' rule ):  $k = 3.3 \log (n)+1$

Tentukan lebar kelas :  $l = (X_{\max} - X_{\min})/k$

Tetapkan nilai awal ( $\leq X_{\min}$ )

Tentukan batas masing-masing kelas

Hitung frekuensi pengamatan pada masing-masing kelas

Frekuensi Relatif : cari proporsi dari masing-masing kelas

# Peubah Numerik



- Peubah: Tinggi

$$n \text{ kelas: } k = 3.3 \log(21) + 1 = 5.36 \approx 6$$

$$\text{lebar kelas: } l = (176 - 151) / 6 = 4.2 \approx 5$$

Selang Kelas	Tengah Kelas	Turus	Frekuensi	Frekuensi Relatif	%
151 - 155	153	I	1	0.05	4.76
156 - 160	158	IIII	4	0.19	19.05
161 - 165	163	IIII	9	0.43	42.86
166 - 170	168	IIII	4	0.19	19.05
171 - 175	173	II	2	0.1	9.52
176 - 180	178	I	1	0.05	4.76
Total			21	1	100

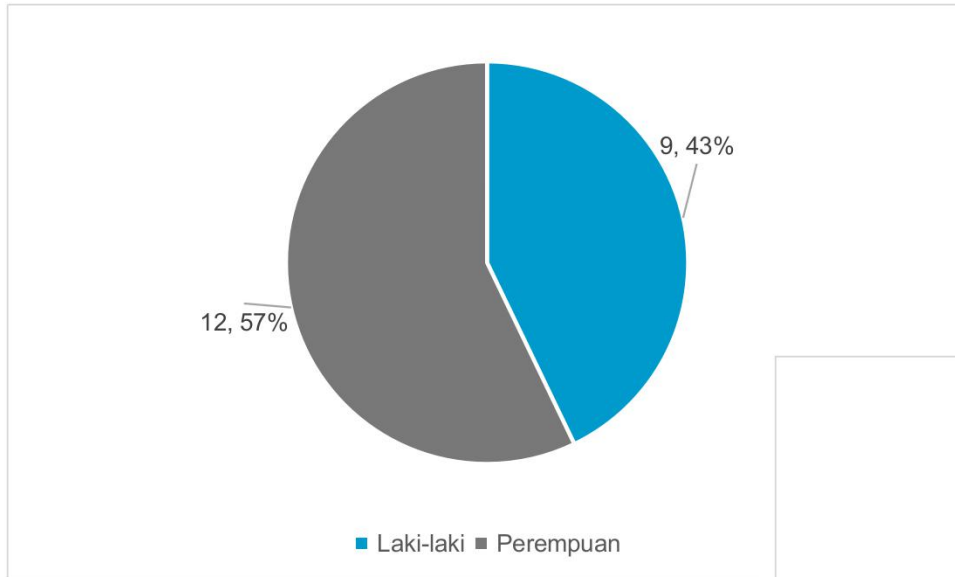


- Tabel Ringkasan Statistik

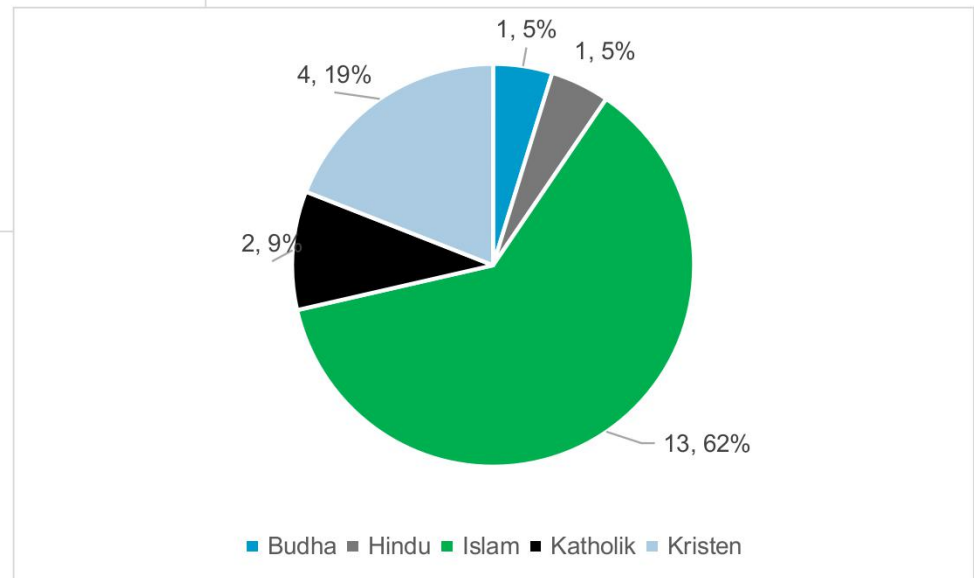
Peubah	Jenis Kelamin	N	Mean	StDev	Minimum	Median	Maximum
Tinggi	Perempuan	9	160.56	5.43	151	161	169
	Laki-laki	12	166.25	5.07	159	165	176
Berat	Perempuan	9	53.89	5.62	45	54	60
	Laki-laki	12	64.75	8.04	52	63	82

# Penyajian Grafik

- Pie Chart

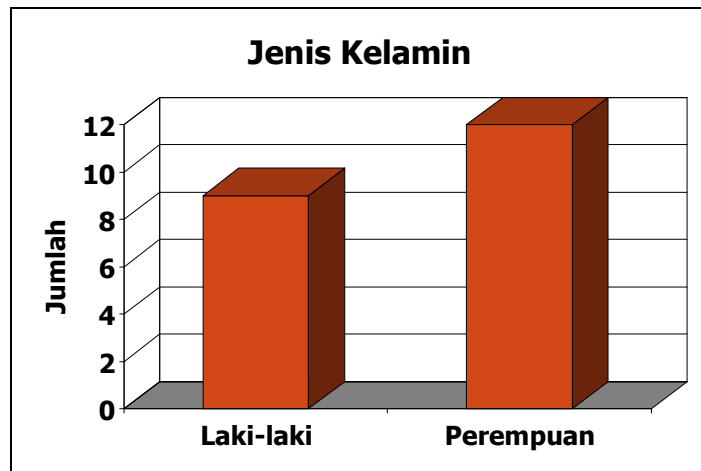


Umumnya untuk Peubah Kategorik



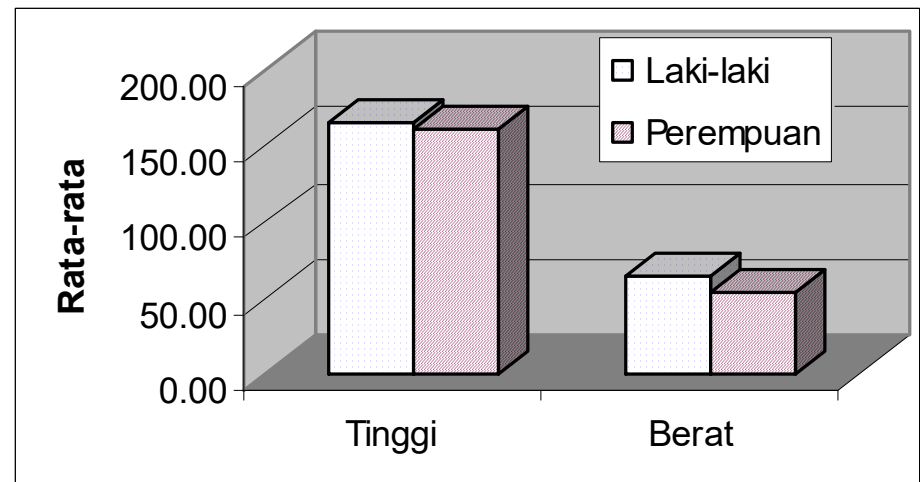
# Penyajian Grafik

- Bar Chart



Peubah Kategorik

Peubah Numerik



# Peubah Numerik

*Penyajian Data dengan Grafik*

# Histogram

Sebuah grafik dari suatu sebaran frekuensi

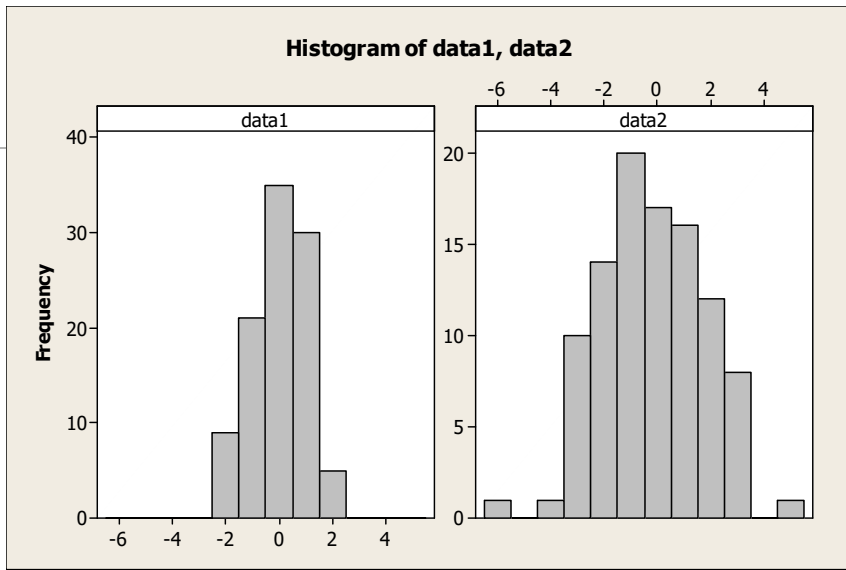
Bisa distribusi dari frekuensi-nya atau frekuensi relatif-nya

Digunakan untuk melihat distribusi dari data:

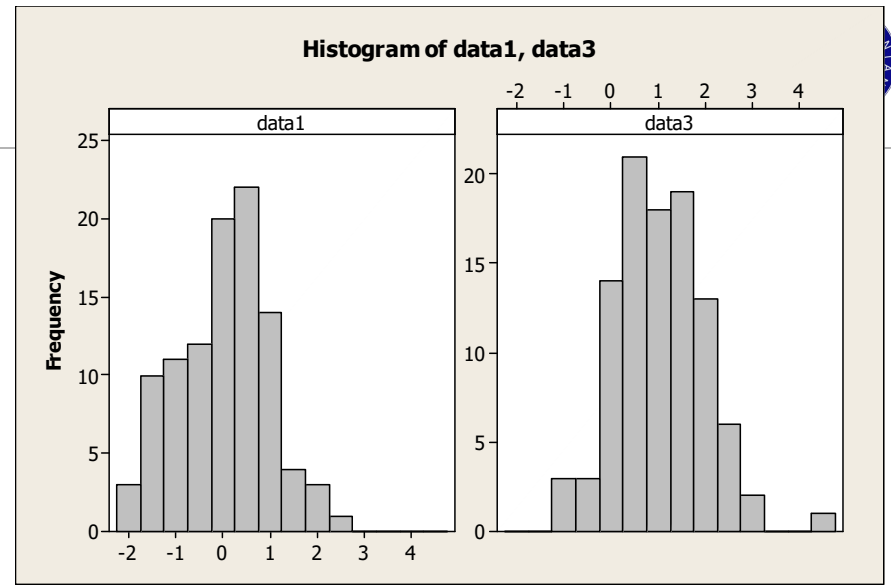
Melihat ukuran penyebaran dan ukuran pemusatan data

Melihat adanya data outlier

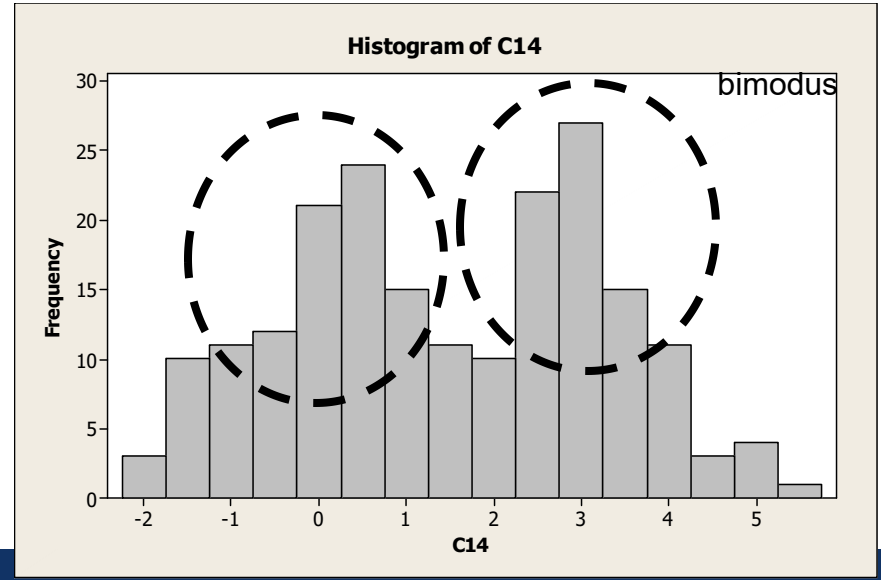
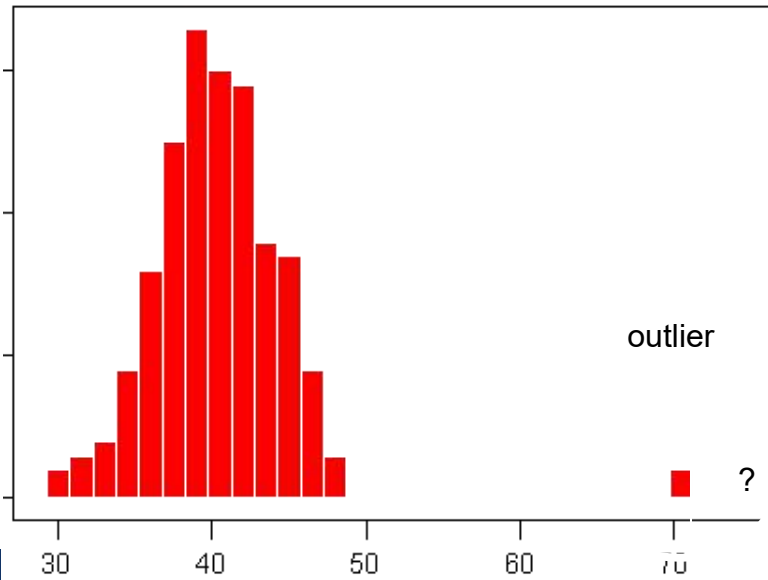
Mendeteksi ada bimodus/tidak



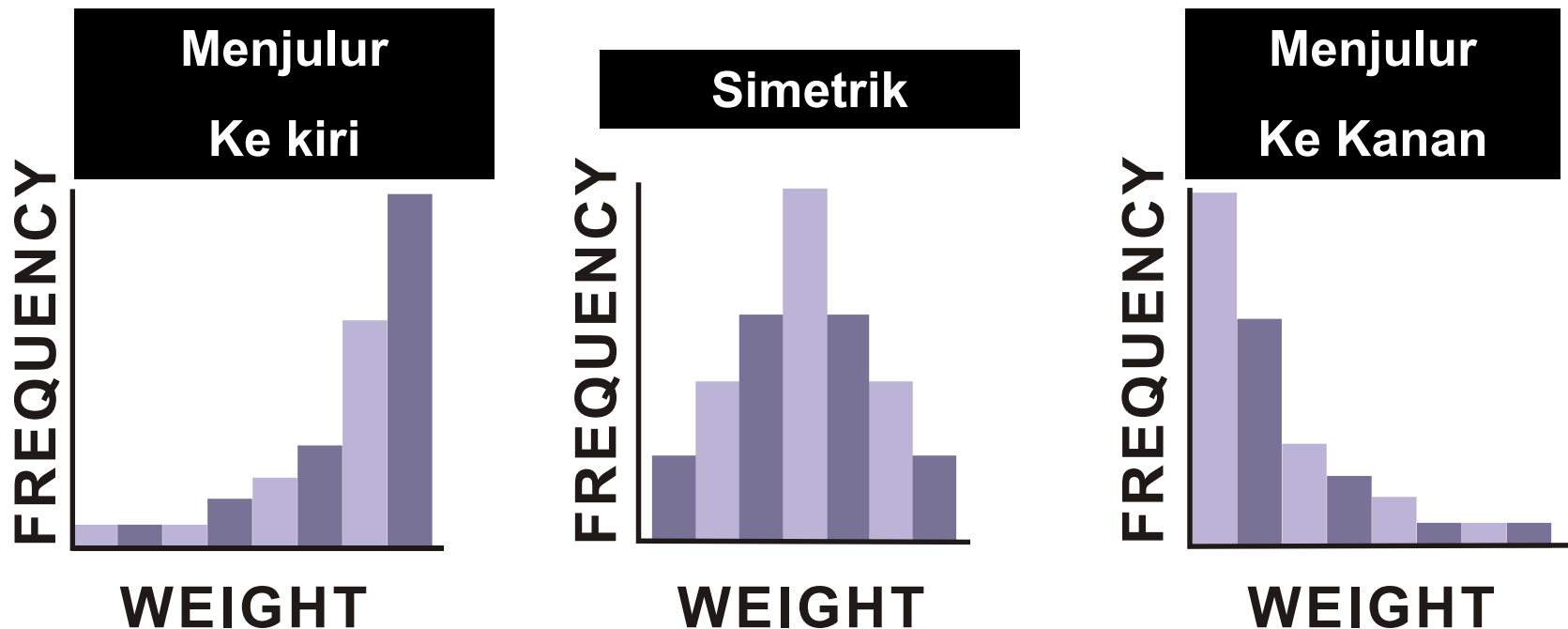
Ukuran Pemusatan relatif sama namun ukuran penyebaran relatif berbeda



Ukuran Pemusatan relatif berbeda namun ukuran penyebaran relatif sama



# Histogram – Mengukur bentuk sebaran



# Ilustrasi

Data 2

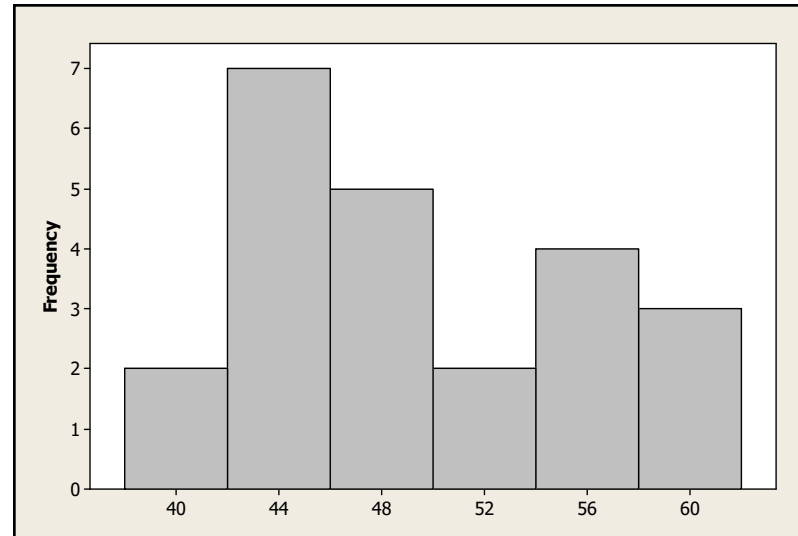
58      57      50      56      44      59      43      52      55      49  
 43      43      49      55      58      48      46      42      44      48  
 40      40      42

Selang kelas	Tengah Kelas	Tepi Batas kelas	Turus	Frekuensi	Frekuensi Relatif	%
38-41	39.5	37.5 - 41.5		2	0.09	8.70%
42-45	43.5	41.5 - 45.5		7	0.30	30.43%
46-49	47.5	45.5 - 49.5		5	0.22	21.74%
50-53	51.5	51.5 - 53.5		2	0.09	8.70%
54-57	55.5	53.5 - 57.5		4	0.17	17.39%
58-61	59.5	57.5 - 61.5		3	0.13	13.04%
Total				23	1	100.00%



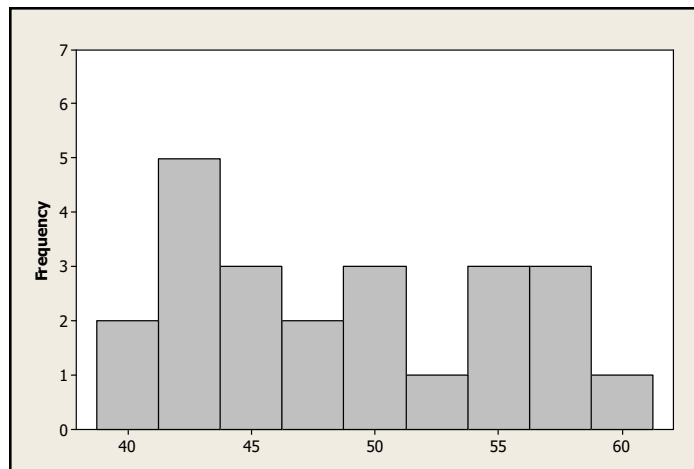
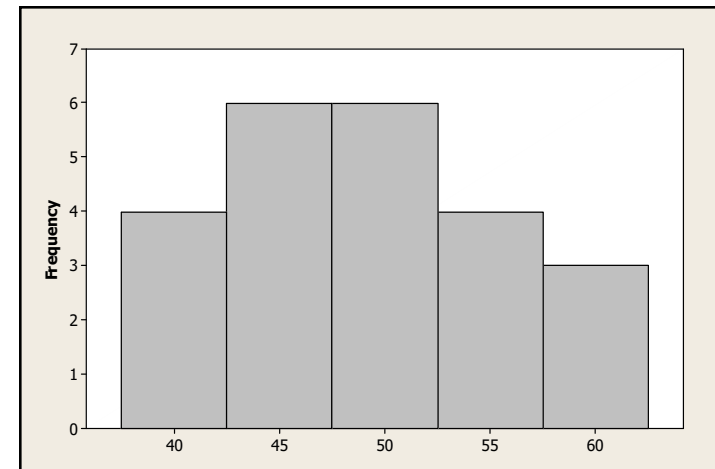
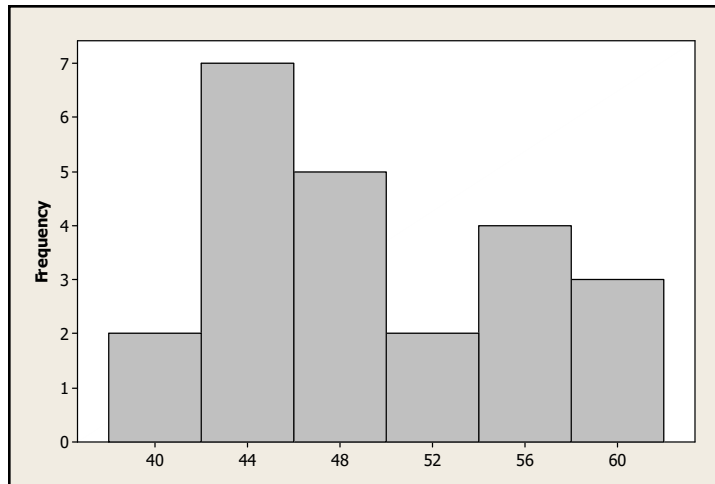
## Ilustrasi—Data 2

***Berdasarkan tabel sebaran frekuensi tersebut maka tampilan histogramnya sebagai berikut:***



Sebagian besar berusia kurang dari 50 tahun, sedangkan frekuensi paling banyak berada pada usia 44 tahun. Bentuk sebaran tidak simetrik, terdapat dua kelompok usia (kurang dari 50 tahun dan lebih dari 50 tahun) → bimodus

# Keragaman berbagai bentuk histogram



Bentuk histogram tidak unik → tergantung nilai awal dan lebar batang (*bandwidth*)

# Diagram Dahan Daun



- Sebuah diagram yang menampilkan distribusi dari data numerik yang sudah terurut dari terkecil dan terbesar
- Sesuai dengan namanya diagram dahan daun terdiri dari bagian dahan dan bagian daun. Bagian daun selalu terdiri dari satu digit. Bagian dahan terletak di sebelah kiri dan bersesuaian dengan bagian daun (jika ada) di sebelah kanan
- Secara visual, diagram dahan daun hampir sama dengan bar chart dimana kategori-kategorinya didefinisikan dengan struktur desimal dari bilangan yang ada

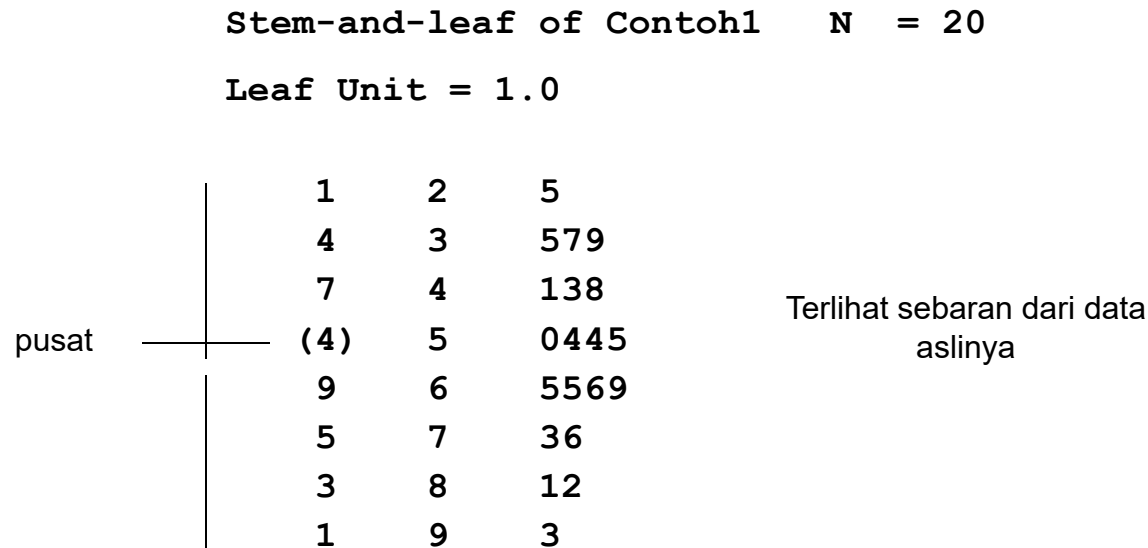
# Manfaat diagram dahan daun

- Mendapatkan sebaran dari data

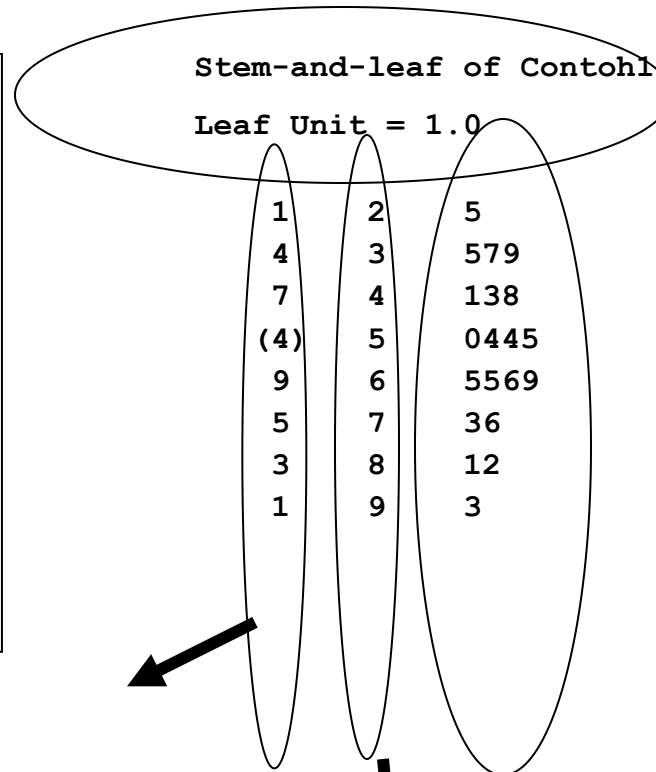
Mendapatkan ukuran penyebaran dan ukuran pemusatan data

Mendeteksi adanya data outlier (jika ada)

Mendeteksi ada bimodus/tidak



Contoh1	
25	65
65	93
82	66
37	50
54	43
41	69
48	73
76	81
54	35
39	55



N = 20

Informasi satuan dari daun  
→ satuan

Bagian daun

Bagian dahan

Frekuensi kumulatif dari jumlah daun pada masing-masing dahan. Dihitung dari atas dan bawah sampai ketemu di posisi median

# Cara membuat diagram dahan daun

- Pisahkan bagian dahan dan daun. Untuk contoh diatas misalkan dahan berupa puluhan dan daunnya berupa satuan
- Bagian dahan urutkan dari terkecil sampai terbesar

2  
3  
4  
5  
6  
7  
8  
9

- Plot daun sesuai dengan dahan yang tersedia. Sebagai langkah awal untuk memudahkan pekerjaan identifikasi secara berurutan dari data yang ada

2 5  
 3 795  
 4 183  
 5 4405  
 6 5569  
 7 63  
 8 21  
 9 3

Contoh1	
25	65
65	93
82	66
37	50
54	43
41	69
48	73
76	81
54	35
39	55

2 5  
 3 579  
 4 138  
 5 0445  
 6 5569  
 7 36  
 8 12  
 9 3

- Urutkan bagian daun dari terkecil sampai yang terbesar

- Perhatikan data berikut:
- Nilai minimum: 8 dan maks : 38
- Diagram Dahan Daun:

0 899  
1 02235666779  
2 01344689  
3 18

Contoh2	
9	26
16	29
38	23
15	17
24	10
12	17
20	16
9	13
31	8
28	12
24	21
19	16



# Dahan terbagi dalam 2 dahan

- Aturan main: dahan 1 untuk digit 0-4 dan dahan 2 untuk digit 5-9
- Perhatikan data berikut:

Contoh2	
9	26
16	29
38	23
15	17
24	10
12	17
20	16
9	13
31	8
28	12
24	21
19	16

```

0 899
1 02235666779
2 01344689
3 18
  
```

Stem-and-leaf of Contoh2      N = 24

Leaf Unit = 1.0

```

3      0 899
7      1 0223
(7)    1 5666779
10     2 01344
5      2 689
2      3 1
1      3 8
  
```

# Quintuple stem

- Bagi dahan ke dalam 5 dahan per 10 nilai bilangan. Aturan main sebagai berikut: \* untuk daun 0 dan 1, t untuk 2 dan 3, f untuk 4 dan 5, s untuk 6 dan 7, dan "." untuk 8 dan 9

- Perhatikan data berikut:

<b>Contoh3</b>	
<b>3</b>	<b>8</b>
<b>11</b>	<b>15</b>
<b>27</b>	<b>12</b>
<b>14</b>	<b>7</b>
<b>9</b>	<b>9</b>
<b>4</b>	<b>10</b>
<b>10</b>	<b>14</b>
<b>16</b>	<b>5</b>
<b>7</b>	<b>15</b>
<b>18</b>	<b>13</b>
<b>17</b>	<b>12</b>
<b>11</b>	

**Stem-and-leaf of Contoh3      N = 23**

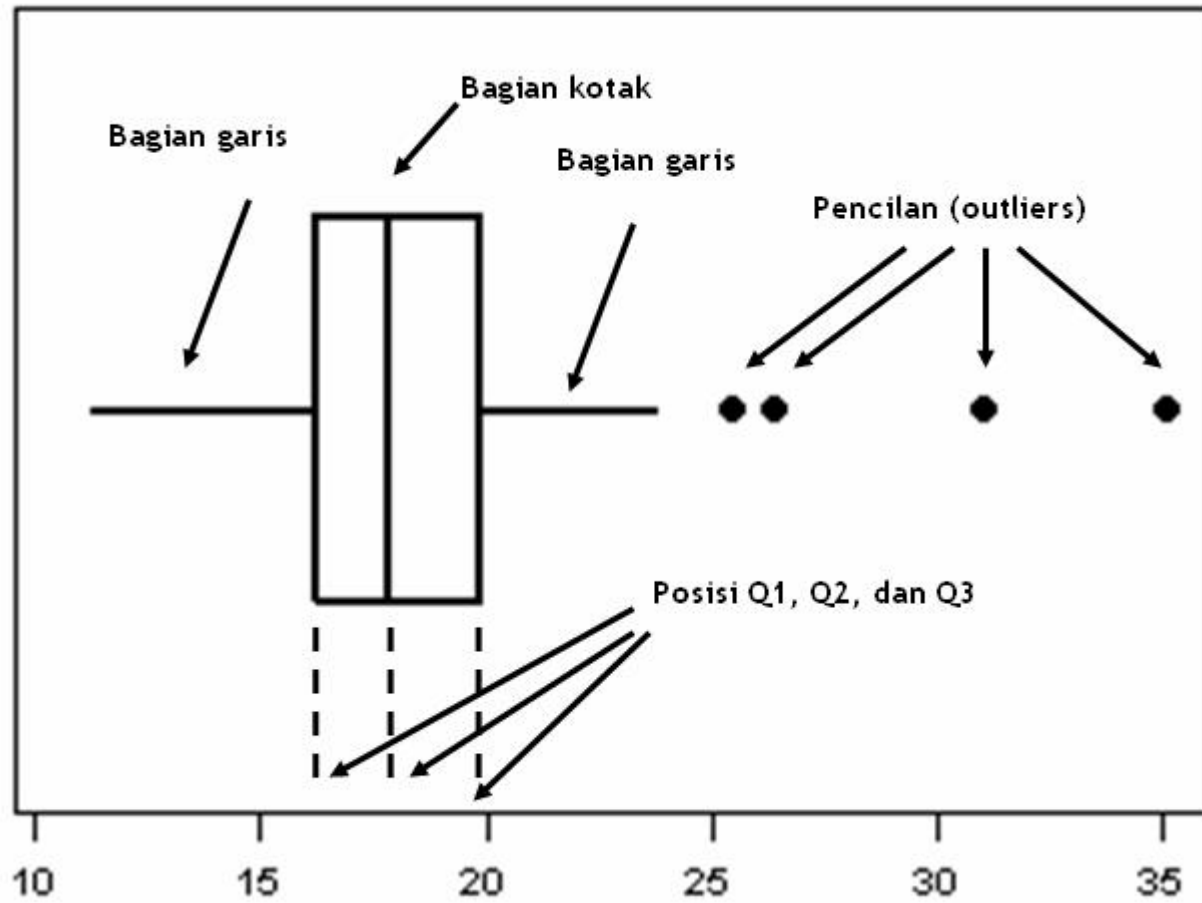
Leaf Unit = 1.0

0 t	3	1	0 3
f	45	3	0 45
s	77	5	0 77
.	899	8	0 899
1 *	0011	(4)	1 0011
t	223	11	1 223
f	4455	8	1 4455
s	67	4	1 67
.	8	2	1 8
2 *		1	2
t		1	2
f		1	2
s	7	1	2
		1	2 7

Aturan banyaknya dahan yang digunakan :  
 antara 4-12 dahan  
 Sesuaikan dengan informasi yang diperoleh berkaitan dengan bentuk sebaran, ukuran pemusatan dan penyebaran data

# BOXPLOT

- informasi ukuran pemusatan dan penyebaran (berupa kuartil)
- informasi bentuk sebaran
- informasi data ekstrim



# Tahapan

- hitung statistik lima serangkai (Min, Q1, Q2, Q3, Max)
- hitung pagar dalam atas
$$PDA = Q3 + 3/2 (Q3 - Q1)$$
- hitung pagar dalam bawah
$$PDB = Q1 - 3/2 (Q3 - Q1)$$
- deteksi keberadaan pencilan, yaitu data yang nilainya kurang dari PDB atau data yang lebih besar dari PDA
- gambar kotak, dengan batas Q1 sampai Q3, dan letakkan tanda garis di tengah kotak pada posisi Q2

- Tarik garis ke kanan, mulai dari Q3 sampai data terbesar di dalam batas atas
- Tarik garis ke kiri, mulai dari Q1 sampai data terkecil di dalam batas bawah
- tandai pencilan dengan bulatan kecil

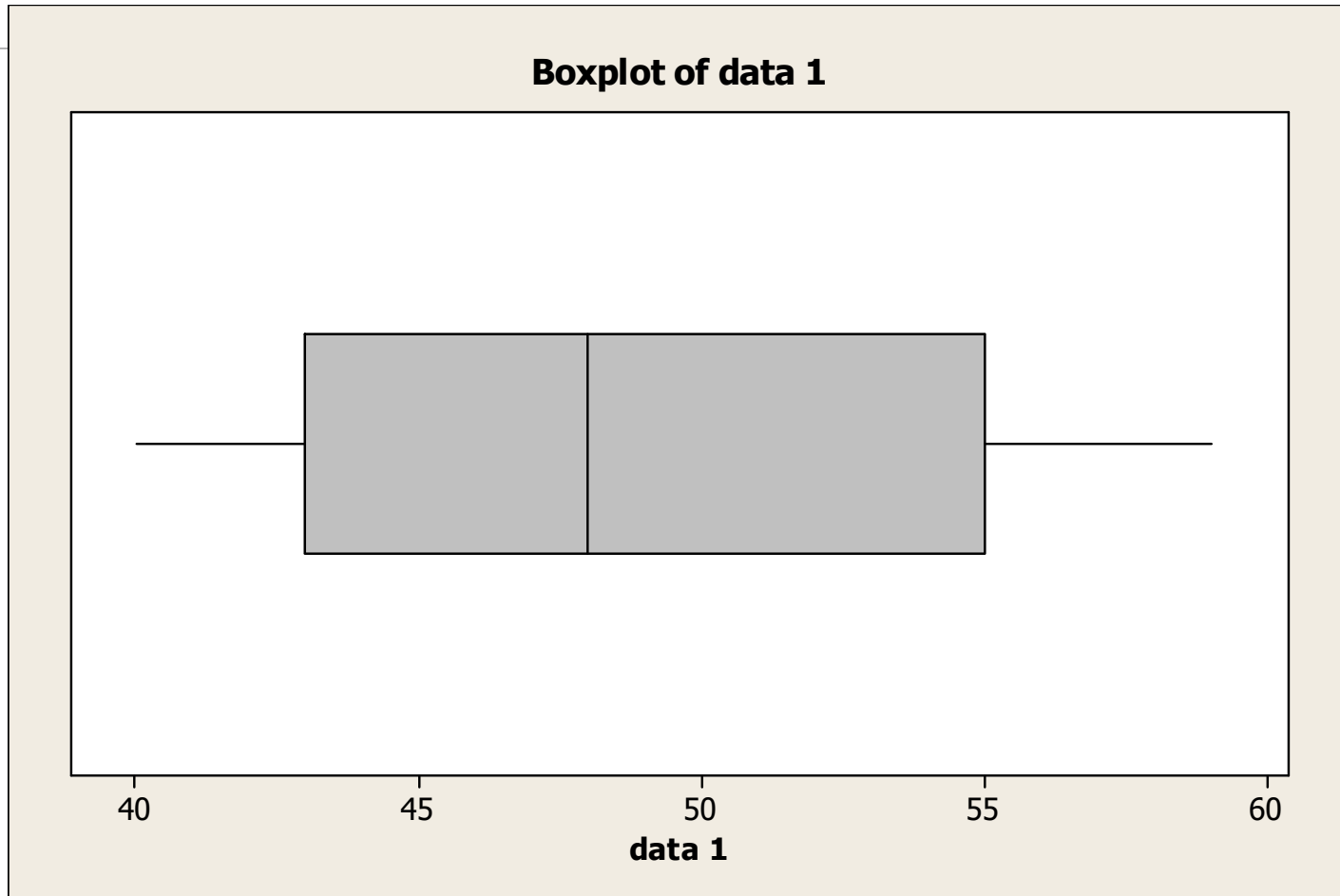
## Ilustrasi (1)

- Statistik 5 serangkai dari data sbb:

	Me		48
Q1	Q3	43	55
Min	Max	40	59

- $PDA = 55 + 1.5 (55 - 43) = 73$
- $PDB = 43 - 1.5 (55 - 43) = 25$
- Tidak ada pencilan





Sebaran data tidak simetrik, karena nilai median lebih dekat ke Q1 → menjulur ke kanan

Tidak ada pencilan

# Ilustrasi (2)

Stem-and-leaf of data 1 N = 23

Leaf Unit = 1.0

```

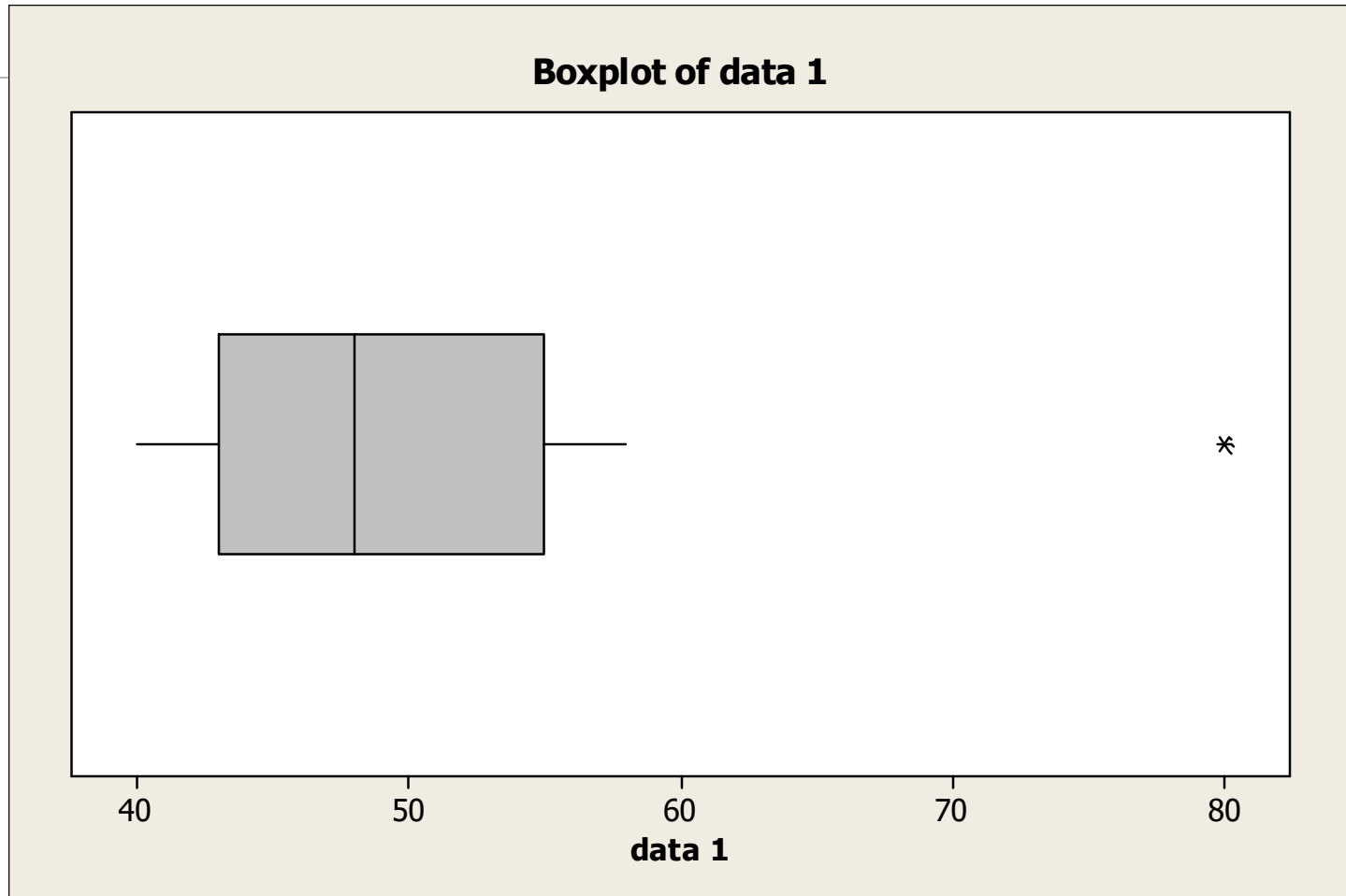
9 4 002233344
(5) 4 68899
9 5 02
7 5 556788
1 6
1 6
1 7
1 7
1 8 0
  
```

	Me		48
Q1	Q3	43	55
Min	Max	40	80

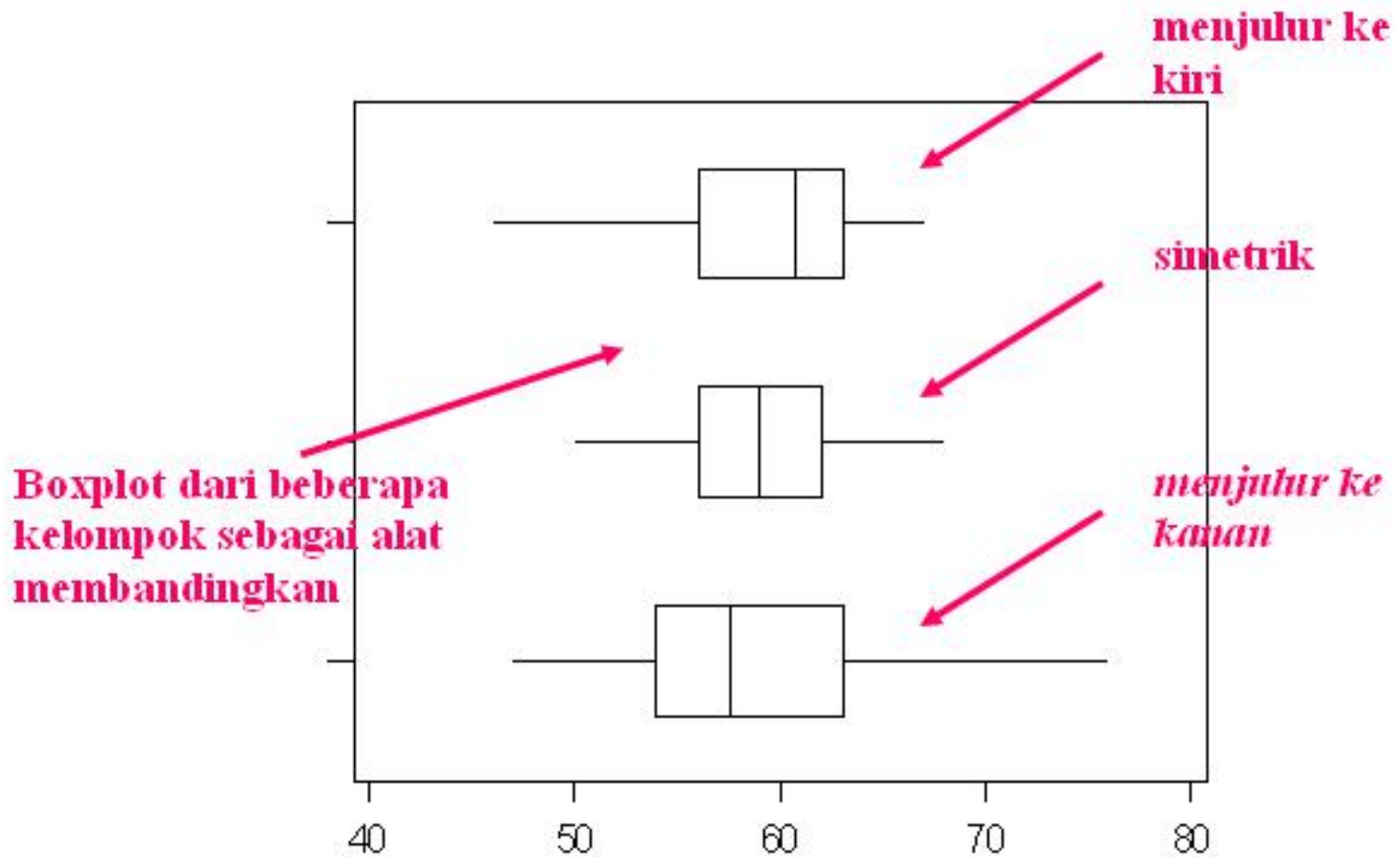
$$PDA = 55 + 1.5 (55 - 43) = 73$$

$$PDB = 43 - 1.5 (55 - 43) = 25$$

Pencilan : 80



Sebaran data tidak simetrik, karena nilai median lebih dekat ke  
Q1 → menjulur ke kanan  
Terdapat nilai pencilan (80)



# PERINGKASAN DATA

# Deskripsi Data Numerik

- Tujuan Mendeskripsikan data → Mengetahui karakteristik data sesederhana mungkin tetapi memiliki pengertian yang dapat menjelaskan data secara keseluruhan
- Data Numerik memiliki pusat dan keragaman:
  - Ukuran pemusatan
  - Ukuran penyebaran

# UKURAN PEMUSATAN

# Ukuran Pemusatan

- Definisi:

merupakan suatu gambaran (informasi) yang memberikan penjelasan bahwa data memiliki satu (mungkin lebih) titik nilai dimana dia memusat atau terkumpul

- Beberapa Ukuran:

**Median**

**Modus**

**Nilai tengah (rata-rata/rerata)**



# Median

- Definisi : suatu nilai data yang membagi dua sama banyak kumpulan data yang telah diurutkan.
- Langkah Teknis:

Urutkan data dari kecil ke besar

Cari posisi median ( $n_{\text{med}} = (n+1)/2$ )

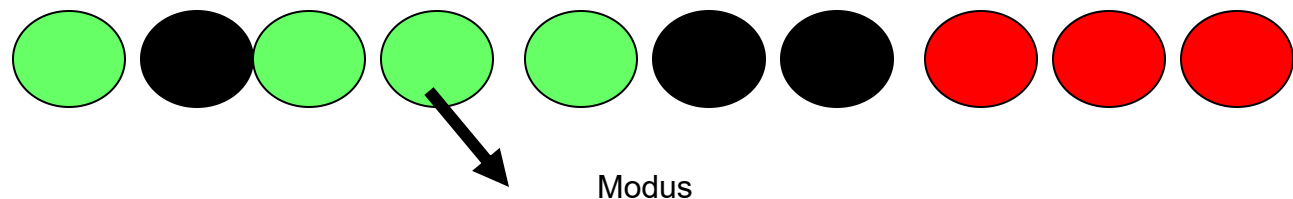
Nilai median

Jika  $n_{\text{med}}$  bulat, maka Median =  $X_{(n+1)/2}$

Jika  $n_{\text{med}}$  pecahan, maka Median =  $(X_{(n)/2} + X_{(n)/2+1})/2$  (rata-rata dua pengamatan yang berada sebelum dan setelah posisi median)

# Modus (Mode)

- Merupakan nilai pengamatan yang paling sering muncul
- Dalam satu gugus data dapat mengandung lebih dari satu modus
- Dapat digunakan untuk semua jenis data, tapi paling banyak digunakan untuk data kategorik atau data diskret dengan hanya sedikit nilai yang mungkin muncul



## Nilai tengah (rata-rata)

- Definisi: merupakan ukuran yang menimbang data menjadi dua kelompok data yang memiliki massa yang sama
- Apabila  $x_1, x_2, \dots, x_N$  adalah anggota suatu populasi terhingga berukuran  $N$ , maka nilai tengah populasinya adalah:

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i$$

## Nilai tengah (rata-rata/rata-rata)

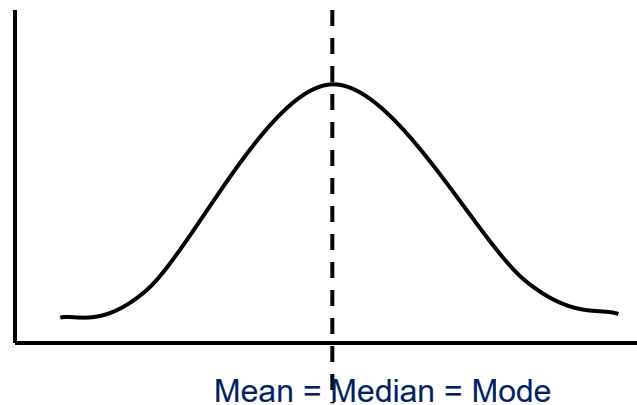
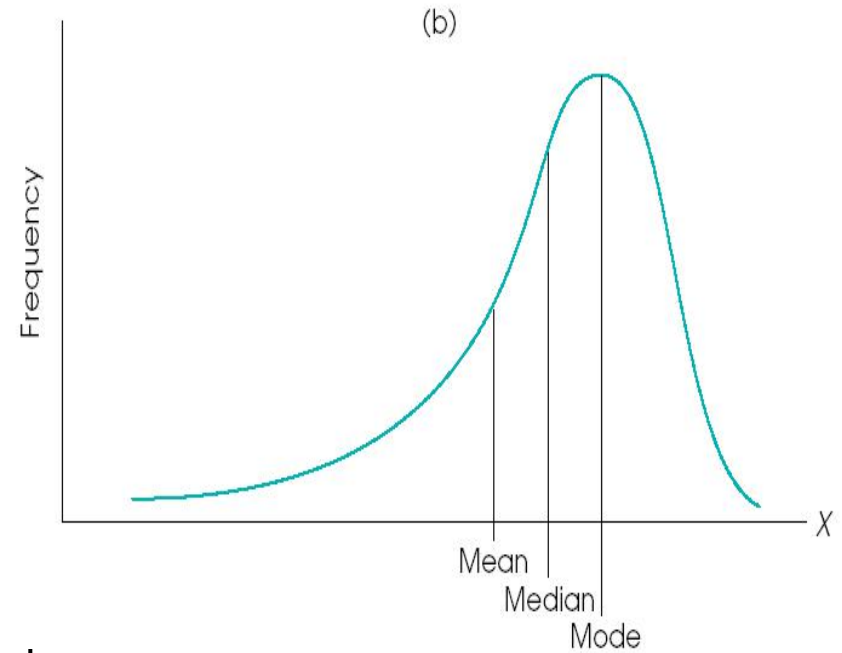
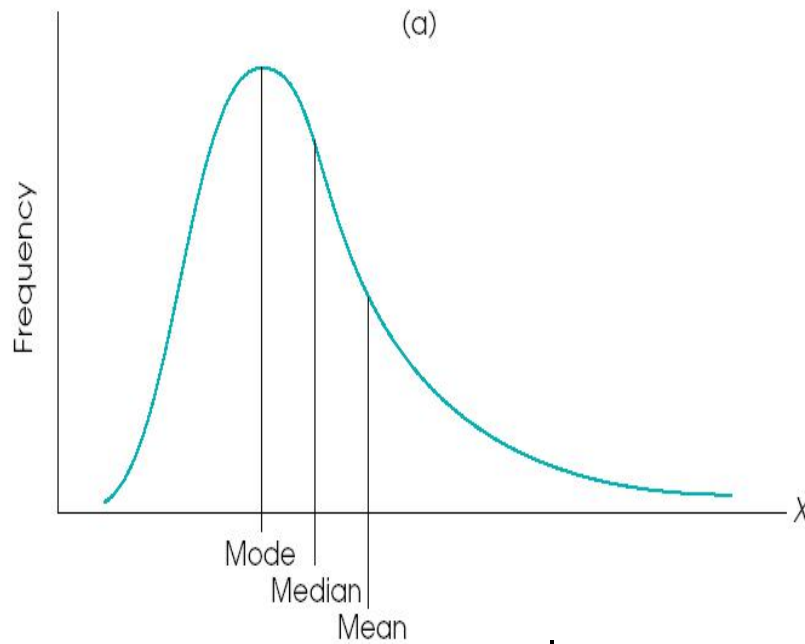
- sedangkan jika  $x_1, x_2, \dots, x_n$  adalah anggota suatu contoh berukuran  $n$ , maka rata-rata contoh tersebut adalah:

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

dalam Bahasa Inggris, rata-rata populasi disebut dengan **mean** dan rata-rata contoh disebut **average**

# KAITAN ANTAR UKURAN PEMUSATAN

# Kaitan antar bentuk sebaran dengan ukuran pemusatan



# Ilustrasi Ukuran Pemusatan (Mean vs Median)-1

- Perhatikan data berikut:

1, 3, 6, 7, 8, 9, 10, 12

Data tersebut memiliki rata-rata = 7 dan median = 7.5

- Selanjutnya pada data berikut

1, 3, 6, 7, 8, 9, 10, 120

memiliki rata-rata = 20.5 dan median = 7.5

# Ilustrasi Ukuran Pemusatan (Mean vs Median)-2

- Kedua data di atas hanya memiliki satu data yang berbeda yaitu yang terakhir. Terlihat bahwa nilai rata-rata berbeda jauh ketika ada data yang ekstrim.  
→ Rata-rata memiliki sifat **tidak kekar (*robust*)**, artinya terpengaruh oleh nilai ekstrim.



# Ilustrasi Ukuran Pemusatan (Mean vs Median)-3



- Jika ada nilai ekstrim besar, maka rata-rata akan bergeser ke kanan (ke nilai besar).
- Sebaliknya jika ada data yang ekstrim kecil, rata-rata akan bergeser ke kiri.
  - diperlukan kehati-hatian ketika menggunakan rata-rata.
- Untuk mengatasi keberadaan data ekstrim sering disarankan menggunakan *5% trimmed mean* (rata-rata terpangkas 5%), yaitu menghitung rata-rata dengan membuang 2.5% data terkecil dan 2.5% data terbesar.

# Ilustrasi Ukuran Pemusatan (Mean vs Median)-4

- Deskripsi data pendapatan per kapita di Indonesia seringkali menimbulkan salah pengertian karena disajikan dalam bentuk rata-rata.
- Karena ada *segelintir* orang yang memiliki pendapatan (sangat) tinggi maka rata-rata pendapatan masyarakat Indonesia akan cenderung lebih tinggi dibandingkan kenyataannya.

# UKURAN PENYEBARAN

# Ukuran Penyebaran

- Definisi : suatu ukuran untuk memberikan gambaran seberapa besar data menyebar dalam kumpulannya.

- Beberapa Ukuran:

Wilayah (*Range*)

Jarak Antar Kuartil (*Interquartile Range*)

Ragam (*Variance*)

Simpangan Baku (*Standard Deviation*)

dll

## Wilayah (*Range*)

- Definisi : suatu ukuran yang dihitung dari selisih pengamatan terkecil dengan pengamatan terbesar

$$W = X_{[N]} - X_{[1]}$$

- Ukuran ini cukup baik digunakan untuk mengukur penyebaran data yang simetrik dan nilai pengamatannya menyebar merata.
- Tetapi ukuran ini akan menjadi tidak relevan jika nilai pengamatan maksimum dan minimum merupakan data-data ekstrem

# Jarak antar kuartil (*Interquartile Range*)

- Definisi : Jarak antar kuartil mengukur penyebaran 50% data ditengah-tengah setelah data diurut.
- Ukuran penyebaran ini merupakan ukuran penyebaran data yang terpangkas 25% yaitu dengan membuang 25% data yang terbesar dan 25% data terkecil.

# Kuartil (Quartile)

- **Definisi** : suatu nilai data yang membagi empat sama banyak kumpulan data yang telah diurutkan

- **Langkah Teknis**

Metode Belah dua

Metode Interpolasi

# Metode Belah dua

- Urutkan data dari kecil ke besar
- Cari posisi kuartil

$$n_{q2} = (n+1)/2$$

$n_{q1} = (n_{q2}^* + 1)/2 = n_{q3}$ ,  $n_{q2}^*$  posisi kuartil dua terpangkas  
(pecahan dibuang)

- Nilai kuartil 2 ditentukan sama seperti mencari nilai median. Kuartil 1 dan 3 prinsipnya sama seperti median tapi kuartil 1 dihitung dari kiri, sedangkan kuartil 3 dihitung dari kanan.



# Metode Interpolasi

- Urutkan data dari kecil ke besar
- Cari posisi kuartil

$$n_{q1} = (1/4)(n+1)$$

$$n_{q2} = (2/4)(n+1)$$

$$n_{q3} = (3/4)(n+1)$$

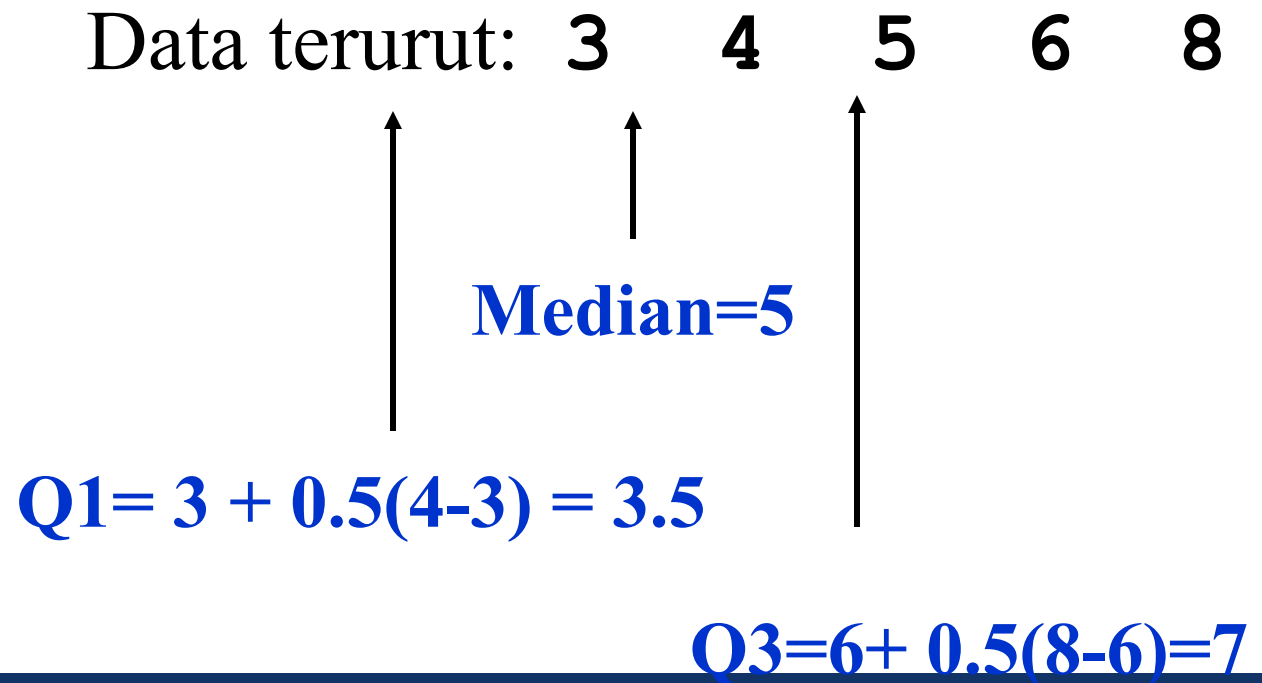
- Nilai kuartil dihitung sebagai berikut:

$$X_{qi} = X_{a,i} + h_i (X_{b,i} - X_{a,i})$$

$X_{a,i}$  = pengamatan sebelum posisi kuartil ke- $i$ ,  $X_{b,i}$  = pengamatan setelah posisi kuartil ke- $i$  dan  $h_i$  adalah nilai pecahan dari posisi kuartil

# Perhatikan ilustrasi data I

- Posisi  $Q_2 = n_{Q_2} = (5+1) / 2 = 3$
- Posisi  $Q_1 = \frac{1}{4}(5+1) = 1.5$
- Posisi  $Q_3 = \frac{3}{4}(5+1) = 4.5$



## Perhatikan ilustrasi data II

- Posisi  $Q_2 = n_{Q_2} = (6+1) / 2 = 3.5$
- Posisi  $Q_1 = \frac{1}{4}(6+1) = 1.75$
- Posisi  $Q_3 = \frac{3}{4}(6+1) = 5.25$

Data terurut: 3 4 5 6 8 8

Median=5.5

$$Q_1 = 3 + 0.75(4-3) = 3.75$$

$$Q_3 = 8 + 0.25(8-8) = 8$$

# Jarak antar kuartil (*Interquartile Range*)



- Jarak antar kuartil dihitung dari selisih antara kuartil 3 (Q3) dengan kuartil 1 (Q1):

$$\text{JAK atau IQR} = Q3 - Q1$$

- Ukuran ini sangat baik digunakan jika data yang dikumpulkan banyak mengandung data pencilan

# Ragam (*Variance*)

- Definisi : Ragam merupakan ukuran penyebaran data yang mengukur rata-rata jarak kuadrat semua titik pengamatan terhadap titik pusat (rataaan).
- Apabila  $x_1, x_2, \dots, x_N$  adalah anggota suatu populasi terhingga berukuran  $N$ , maka ragam populasinya adalah

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2$$

## Ragam (*Variance*)

- apabila  $x_1, x_2, \dots, x_n$  adalah anggota suatu contoh berukuran  $n$ , maka ragam contoh tersebut adalah:

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

# Simpangan Baku (*Standard Deviation*)

- Definisi : Merupakan akar dari ragam, yaitu  $\sigma$  simpangan baku populasi dan  $s$  simpangan baku sampel.  
→ diperoleh satuan yang sama dengan data aslinya

# Teladan

- Perhatikan hasil ringkasan terhadap data pendapatan masyarakat (juta rupiah per bulan) dari dua kabupaten berikut ini:

Kabupaten	$\bar{x}$	s
Kabupaten A	0.85	0.56
Kabupaten B	0.82	0.23



# Teladan

- Jika kita hanya menyajikan nilai rata-rata saja dari kedua kabupaten, maka dinyatakan bahwa masyarakat di kedua kabupaten memiliki pendapatan yang relatif sama.
- Penjelasan yang lebih banyak akan diperoleh jika kita melihat nilai-nilai simpangan bakunya.
- Kabupaten A memiliki simpangan baku yang lebih besar daripada Kabupaten B. Artinya, pendapatan masyarakat di Kabupaten A lebih heterogen dibandingkan di Kabupaten B. Implikasi dari informasi ini terhadap kesimpulan bisa signifikan.

***Selesai...***