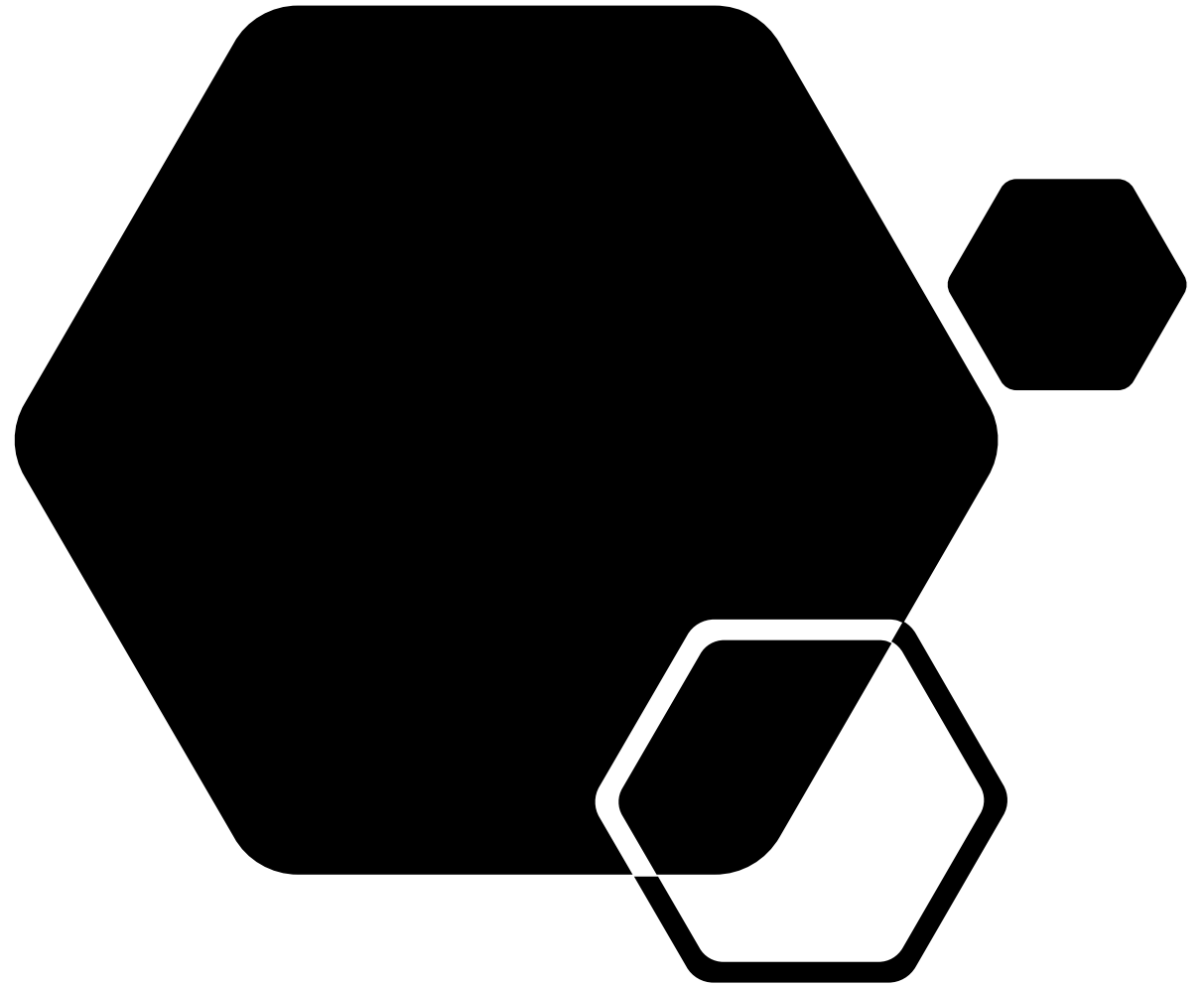


# *Analisis Data Eksploratif*

## *Pertemuan 5*

### *Standarisasi Data*

*Standarisasi*



# Standarisasi Angkatan

- Ringkasan numerik yang biasa digunakan adalah ukuran pemusatan dan penyebaran data.
- Ada 1 hal penting dalam analisis data eksploratif yang jarang digunakan yaitu bentuk data.
- Dalam topik statistika inferensia, memerlukan asumsi bahwa Angkatan berdistribusi normal.

- Bentuk data digunakan untuk mengetahui apakah suatu Angkatan berdistribusi normal atau tidak.
- Biasanya bentuk tertutupi oleh ukuran pemusatan dan sebaran data.
- Untuk memperjelas bentuk data maka pusat dan sebaran data harus dikeluarkan.
- Cara Mengeluarkan pusat data:
  1. Menentukan pusat yang akan digunakan
  2. Mengurangkan pusat tersebut dari setiap observasi di dalam Angkatan.

- Hasil dari pengeluaran pusat akan diperoleh Angkatan data yang baru dengan pusat 0.
  - Misal data observasi awal adalah  $x_1, x_2, x_3, \dots, x_n$  dan data observasi yang baru adalah  $x'_1, x'_2, x'_3, \dots, x'_n$ .
- Jika pusat yang digunakan adalah rata-rata, maka untuk  $x' = x - a$  akan diperoleh  $\bar{x}' = \bar{x} - a$ . Dengan demikian, apabila  $a = \bar{x}$  maka rata-rata  $\bar{x}' = 0$ .

□ Jika pusat yang digunakan adalah median, maka untuk  $x' = x - Med$   
maka

➤  $n$  genap

Karena  $Med = \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2}$  sehingga

$$Med' = \frac{X_{\frac{n}{2}} - Med + X_{\frac{n}{2}+1} - Med}{2} = \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1} - 2Med}{2} = 0$$

➤  $n$  ganjil

Karena  $Med = X_{\frac{n+1}{2}}$  sehingga

$$Med' = X_{\frac{n+1}{2}} - Med = 0$$

□ Begitu pula untuk ukuran pemusatan yang lain juga menunjukkan nilai pemusatan yang baru menjadi 0.

# Contoh

Diberikan data 10, 20, 21, 25, 43, 49, 42, 86. Jika diambil pusatnya adalah Median yaitu 33,5 maka diperoleh data observasi baru dengan  $x' = x - Med = x - 33,5$  sehingga diperoleh  $Med' = 0$ , ditunjukkan dalam table berikut.

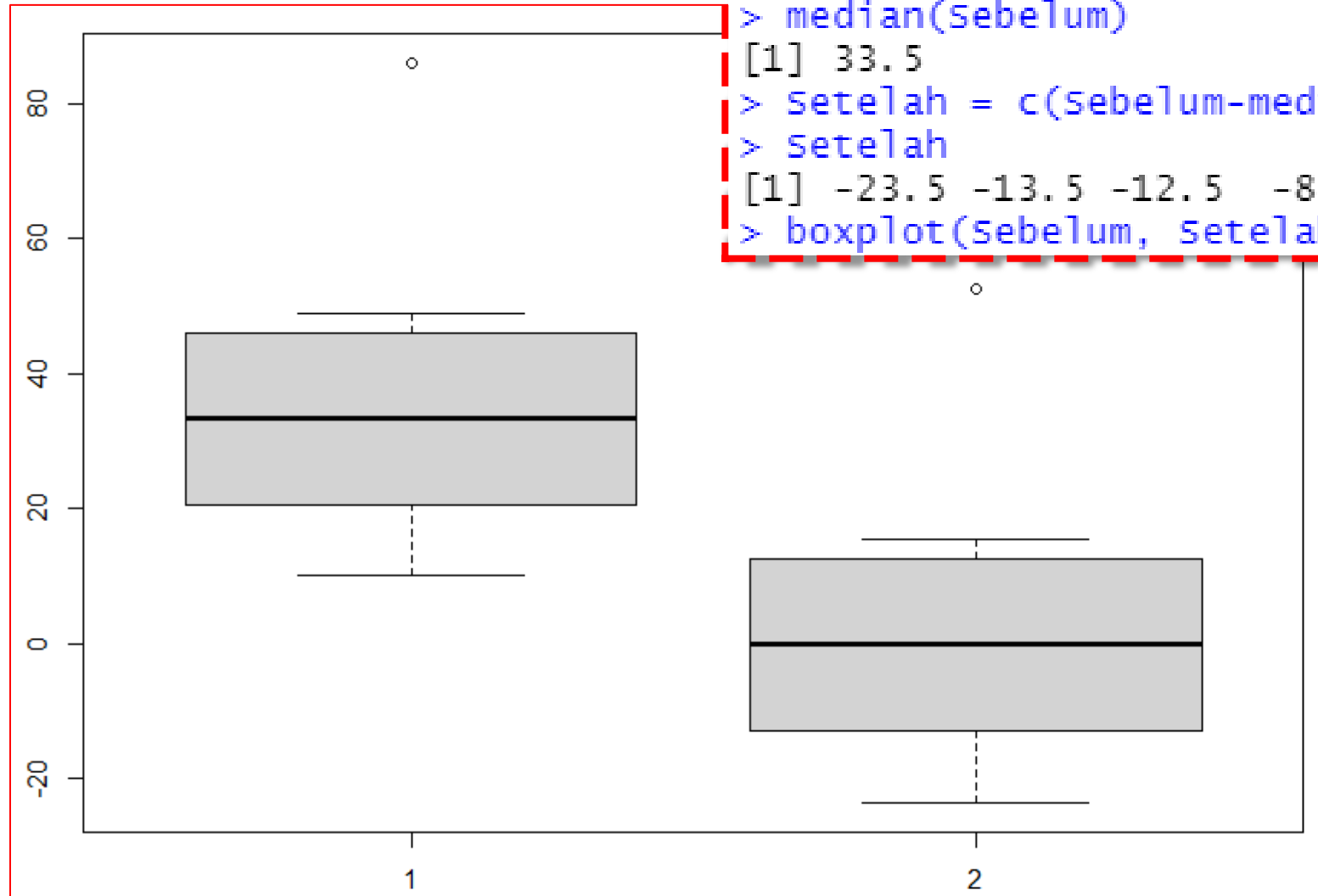
Data awal ( $x$ )	10	20	21	25	42	43	49	86
Data baru ( $x'$ )	-23,5	-13,5	-12,5	-8,5	8,5	9,5	15,5	52,5



**Med'=0**



## Visualisasi dalam bentuk boxplot dengan Rstudio diperoleh



```
> Sebelum = c(10, 20, 21, 25, 43, 49, 42, 86)
> median(Sebelum)
[1] 33.5
> Setelah = c(Sebelum-median(Sebelum))
> Setelah
[1] -23.5 -13.5 -12.5 -8.5 9.5 15.5 8.5 52.5
> boxplot(Sebelum, Setelah)
```

- Dari visualisasi tersebut, tampak bahwa pusat Angkatan untuk observasi original adalah dengan Median 33,5. Sedangkan pusat observasi baru adalah Median 0.
- Sebaran data dari Angkatan tersebut juga tidak mengalami perubahan.

# Standarisasi

- Pengeluaran atau penghilangan sebaran tidak akan mengubah nilai sebaran yang dimiliki oleh Angkatan tersebut.
- Hal ini sesuai dengan sifat variansi bahwa untuk  $x' = x - a$  maka  $Var(x') = Var(x)$ .
- Perlu diambil langkah dengan membagi Angkatan baru dengan sebarannya.

- Standarisasi adalah mengeluarkan pusat dan sebaran observasi  $x$  untuk mendapatkan nilai observasi baru  $x'$ , dengan rumus:

$$x' = \frac{x - \text{pusat}}{\text{sebaran}}$$

- Hal ini menyebabkan  $x'$  akan memiliki pusat 0 dan sebaran 1.

- Misal digunakan pusat adalah rata-rata  $\bar{x}$  dan sebaran adalah standar deviasi  $s$ .

Diambil standarisasi  $x'_i = \frac{x_i - \bar{x}}{s}$  maka

$$\bar{x}' = \frac{\sum \frac{x_i - \bar{x}}{s}}{n} = \frac{\sum x_i - \bar{x}}{ns} = \frac{\frac{\sum x_i}{n} - \frac{n\bar{x}}{n}}{s} = \frac{\bar{x} - \bar{x}}{s} = 0$$

Dan

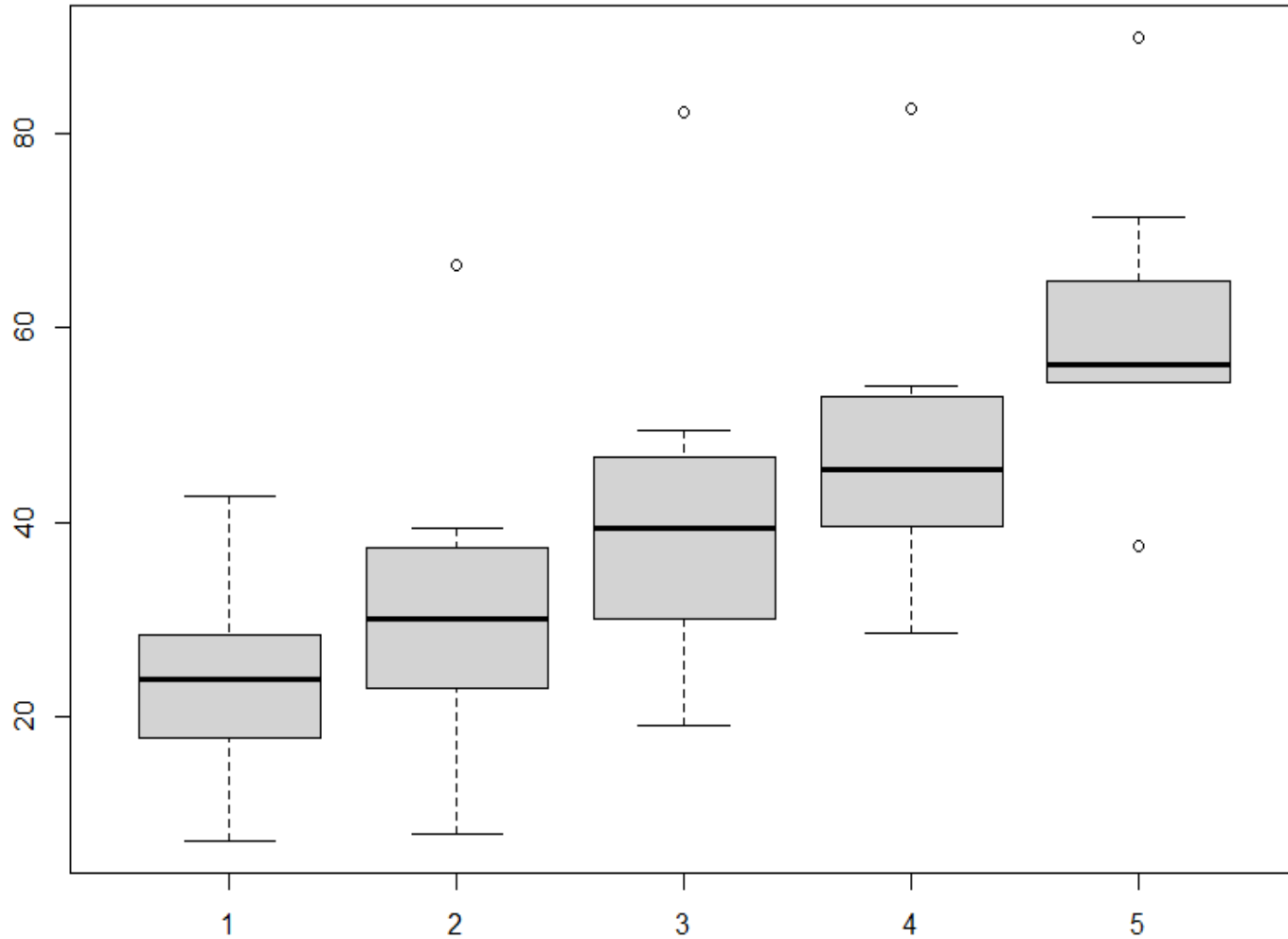
$$var(x') = \frac{\sum (x'_i - \bar{x}')^2}{n - 1} = \frac{\sum \left( \frac{x_i - \bar{x}}{s} - 0 \right)^2}{n - 1} = \frac{\sum (x_i - \bar{x})^2}{s^2(n - 1)} = \frac{s^2}{s^2} = 1$$

# Contoh

Diberikan data angka kematian akibat penyakit jantung di beberapa kota di DIY dan Jateng sbb

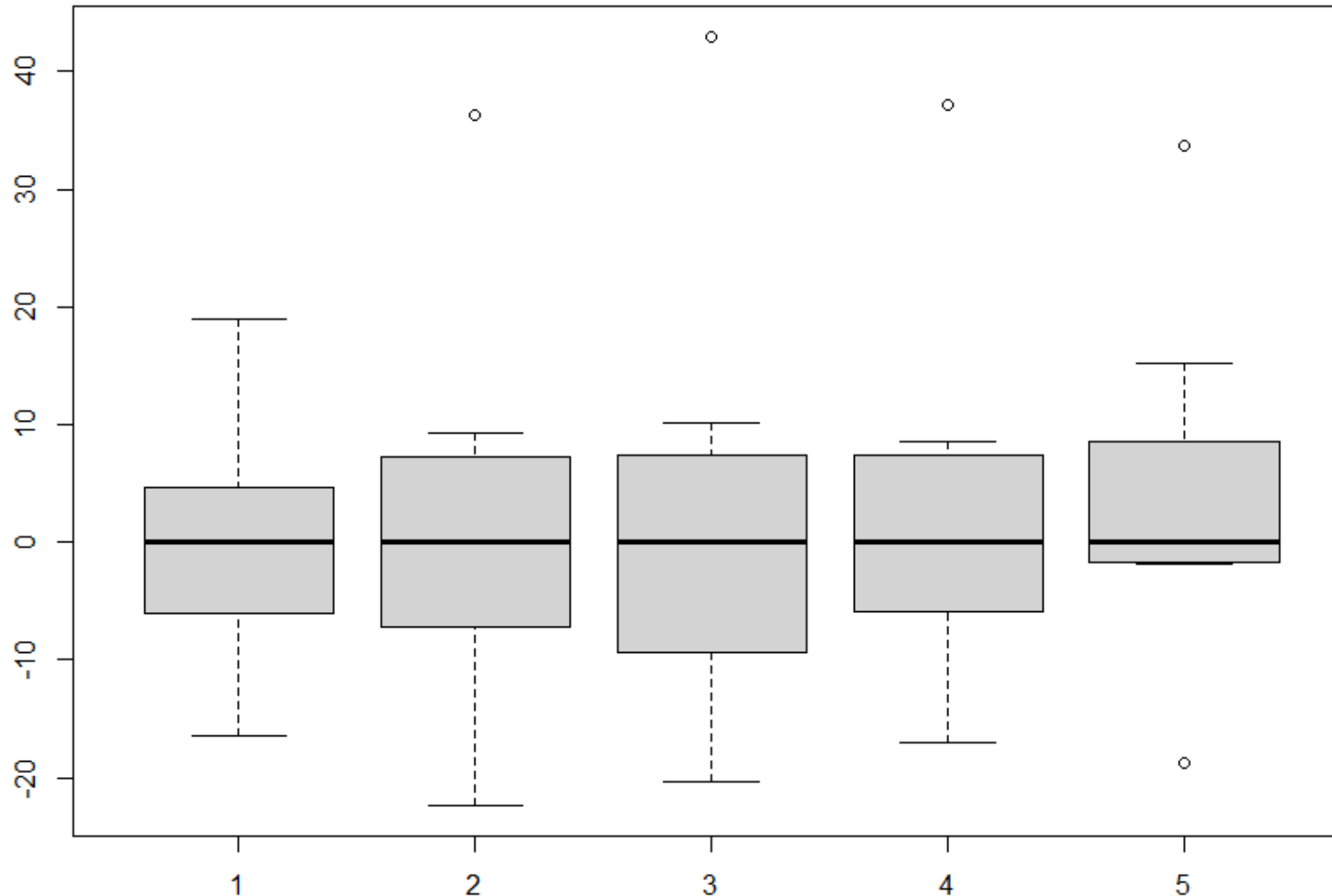
Kabupaten	JK	Usia				
		20-29	30-39	40-49	50-59	60-69
Bantul	L	23,7	30,1	39,4	45,5	56,2
Sleman	L	7,2	7,8	19,0	28,5	37,5
Brebes	L	20,3	19,2	26,3	38,8	71,4
Tegal	L	29,2	39,4	44,0	51,8	54,4
Kendal	L	15,1	26,7	33,8	40,3	54,5
Batang	L	27,4	35,3	49,5	54,1	58,2
Cilacap	L	42,7	66,4	82,3	82,6	89,9

Secara visual data tersebut dapat digambarkan dalam diagram boxplot sbb:



- Dalam visualisasi tersebut tampak bahwa nilai ekstrim setiap Angkatan dapat diketahui dengan baik.
- Selain itu, nilai median setiap Angkatan meningkat seiring meningkatnya angka kematian pada setiap Angkatan.
- Secara keseluruhan, dengan penambahan angka kematian pada setiap Angkatan, data setiap Angkatan juga semakin menyebar.
- Bentuk sebaran data belum terlihat dengan jelas

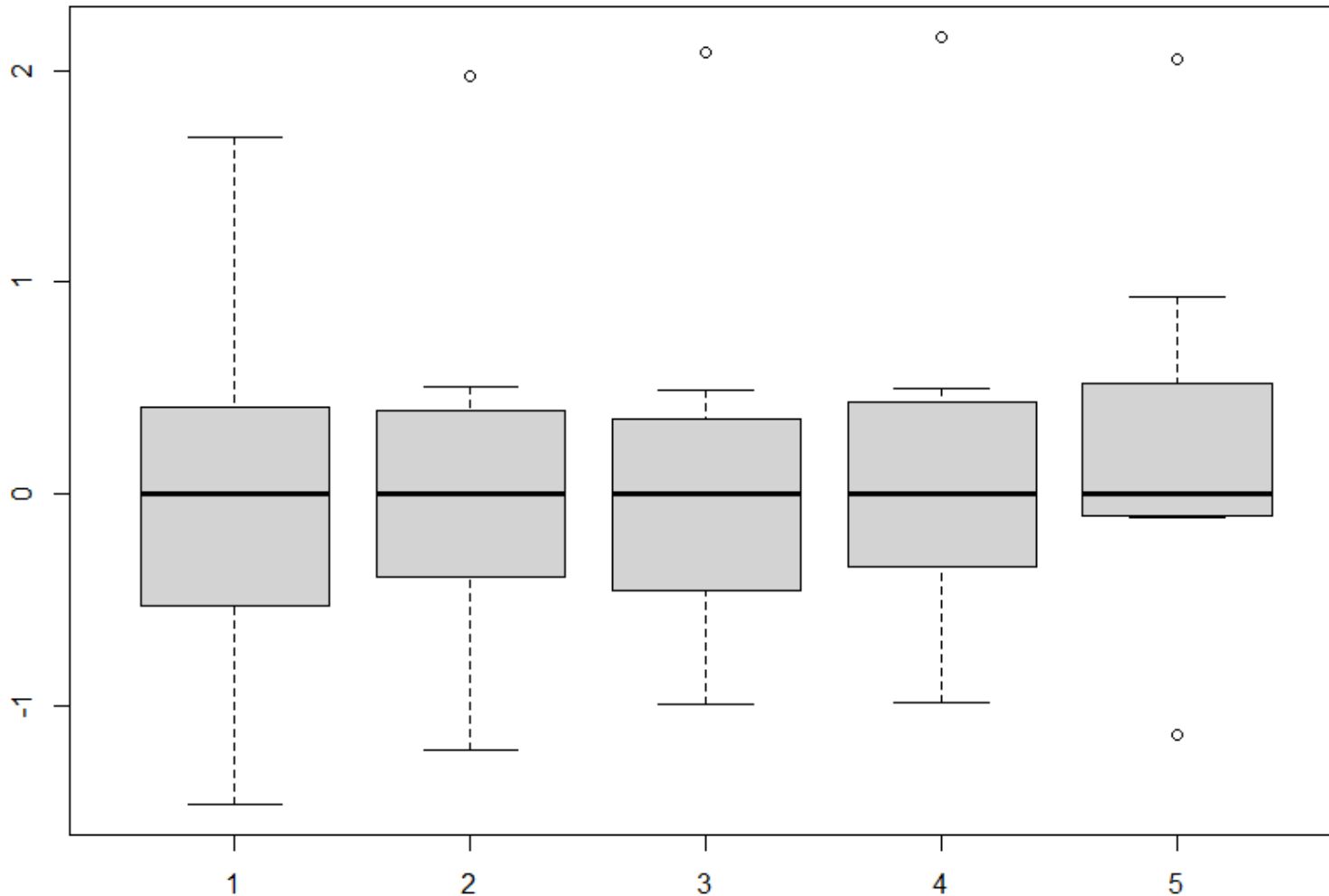
Dengan mengeluarkan pusat Median dari setiap Angkatan, diperoleh Angkatan baru yang Secara visual data dapat digambarkan dalam diagram boxplot sbb:



- Dari Boxplot disamping, tampak bahwa semua Angkatan baru memiliki pusat 0 dengan sebaran yang berbeda-beda
- Bentuk sebaran data masih belum jelas



Dengan melakukan standarisasi yaitu mengeluarkan pusat Median dan Standar Deviasi dari setiap Angkatan, diperoleh Angkatan baru yang Secara visual data dapat digambarkan dalam diagram boxplot sbb:



- Dari Boxplot disamping, tampak bahwa semua Angkatan baru memiliki pusat 0 dengan sebaran 1
- Bentuk sebaran data jauh lebih jelas

# Syntax R studio

```
> x1= c(23.7, 7.2, 20.3, 29.2, 15.1, 27.4, 42.7)
> x2= c(30.1, 7.8, 19.2, 39.4, 26.7, 35.3, 66.4)
> x3= c(39.4, 19, 26.3, 44, 33.8, 49.5, 82.3)
> x4= c(45.5, 28.5, 38.8, 51.8, 40.3, 54.1, 82.6)
> x5= c(56.2, 37.5, 71.4, 54.4, 54.5, 58.2, 89.9)
> boxplot(x1,x2,x3,x4,x5)
```

```
> z1= c(x1-median(x1))
> z2= c(x2-median(x2))
> z3= c(x3-median(x3))
> z4= c(x4-median(x4))
> z5= c(x5-median(x5))
> boxplot(z1,z2,z3,z4,z5)
```

```
> y1= c((x1-median(x1))/sqrt(var(x1)))
> y2= c((x2-median(x2))/sqrt(var(x2)))
> y3= c((x3-median(x3))/sqrt(var(x3)))
> y4= c((x4-median(x4))/sqrt(var(x4)))
> y5= c((x5-median(x5))/sqrt(var(x5)))

> boxplot(y1,y2,y3,y4,y5)
```

# Latihan

Diberikan data mobil yang lewat selama interval waktu lima menit di 3 tempat masing-masing melalui jalan yang berbeda.

Jalan 1 : 30, 45, 26, 44, 18, 24, 38, 42, 29

Jalan 2 : 24, 33, 31, 16, 31, 13, 12, 25, 27

Jalan 3 : 35, 47, 31, 43, 46, 27, 31, 21, 24

- a. Bandingkan masing-masing Angkatan melalui diagram Boxplot
- b. Analisis boxplot yang terbentuk, jika pusat rata-ratanya dikeluarkan
- c. Analisis boxplot yang terbentuk, jika dilakukan standarisasi dengan mengeluarkan rata-rata dan simpangan baku dari masing-masing angkatan