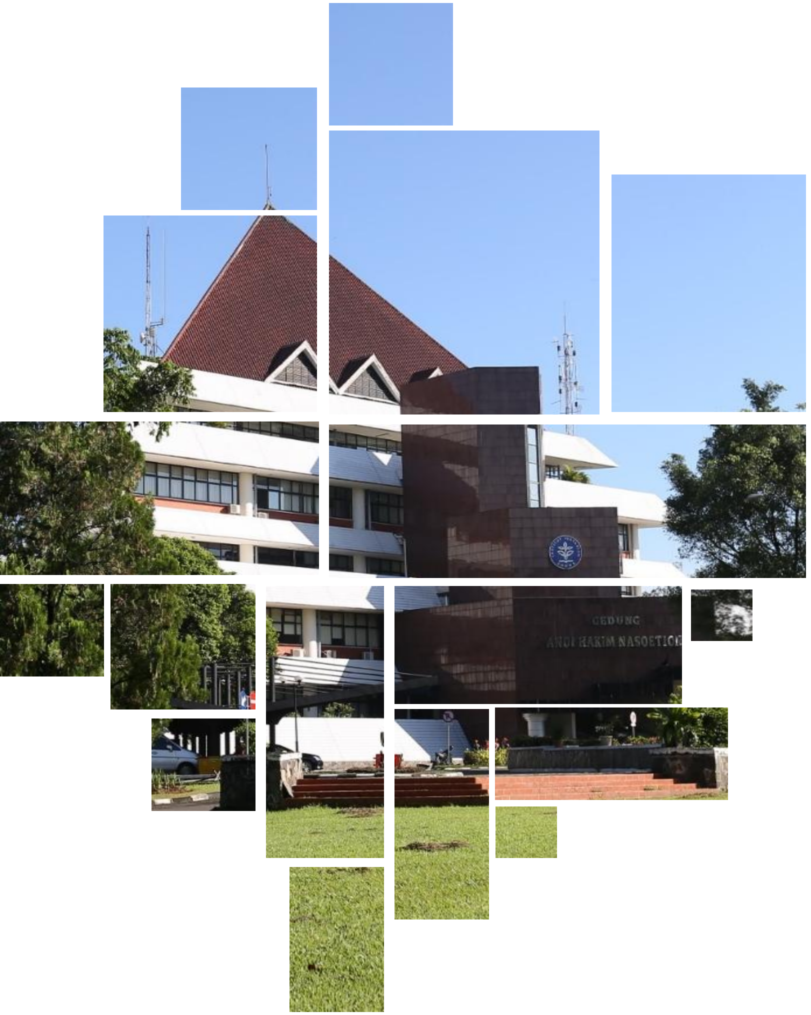


Pertemuan 2:

STATISTIKA DESKRIPTIF

IPB University
Bogor Indonesia

Menyajikan Data dengan Ringkasan Angka

- Metode grafik sering kali tidak cukup untuk mendeskripsikan data, begitu juga sebaliknya.
- **Ringkasan angka** dapat digunakan baik untuk data **populasi** dan data **contoh**
  - Suatu **parameter** is deskripsi numerik (angka) yang dihitung untuk suatu **populasi**
  - Suatu **statistik** adalah deskripsi numerik (angka) untuk suatu **contoh**

Menyajikan Data dengan Ringkasan Angka

Ukuran Pemusatan dan Ukuran Penyebaran (Keragaman)

- **Ukuran Pemusatan (Central tendency)**: adalah suatu ukuran yang menggambarkan pusat dari kumpulan data yang bisa mewakilinya.

- **Ukuran Penyebaran (Dispersion)**:

  adalah suatu ukuran untuk mengetahui seberapa besar penyimpangan data dari ukuran pemusatannya

Menyajikan Data dengan Ringkasan Angka

Pemusatan

- Rataan (Mean)
- Modus (Mode)
- Median
- Persentil (Percentile)

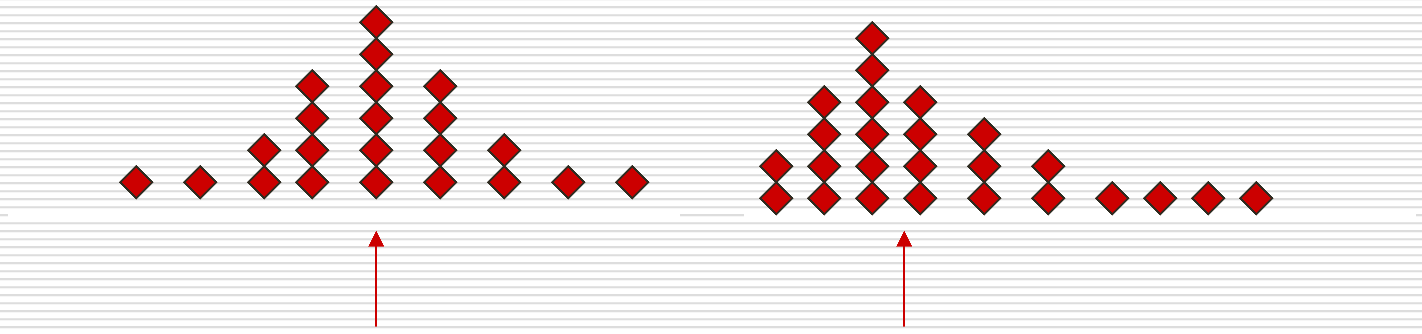
Penyebaran

- Jangkauan
- Jangkauan Antar Kuartil
- Ragam
- Simpangan Baku

Bentuk Sebaran

- Skewness
- Kurtosis

1. Ukuran Pemusatan



Rataan (Mean or Average)

- **Rataan** dari gugus data adalah jumlah dari data tersebut dibagi dengan banyaknya data tersebut.

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

dengan n = banyaknya data

$\sum x_i$ = jumlah semua pengamatan

Jika data kita adalah data populasi, maka **rataan populasi** dinamakan **μ** (Huruf Yunani, “mu”).

Rataan Terboboti (Weighted Mean)

Jika nilai *n* buah data adalah $\mathbf{x_1, x_2, x_3, \ldots x_n}$, dan masing-masing frekuensinya adalah $\mathbf{f_1, f_2, f_3, \ldots f_n}$, nilai rata-rata hitung sekumpulan data tersebut didefinisikan sebagai berikut.

$$\bar{x} = \frac{f_1.x_1 + f_2.x_2 + f_3.x_3 + \ldots. + f_n.x_n}{n}$$

Or

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

atau

$$\bar{x} = \frac{\Sigma f_i.x_i}{\Sigma f_i}$$

$\Sigma f_i.x_i$ = Jumlah hasil perkalian setiap data dan frekuensinya

f_i = Frekuensi data ke-i

x_i = Data ke-i

$\Sigma f_i = n$ = banyak data

Modus (Mode)

- **Modus** adalah pengamatan (data) yang paling sering muncul.
- Gugus data: 2, 4, 9, 8, 8, 5, 3
  - Modusnya ialah **8**.
- Gugus data: 2, 2, 9, 8, 8, 5, 3
  - Terdapat dua modus—**8** dan **2** (bimodal)
- Gugus data: 2, 4, 9, 8, 5, 3
  - Tidak ada modus

Contoh

Banyaknya sarden (dalam kaleng) yang dibeli oleh 25 rumah tangga:

0 0 1 1 1 1 1 2 2 2 2 2 2 2 2 2 3 3
3 3 3 4 4 4 5

- **Rataan?**

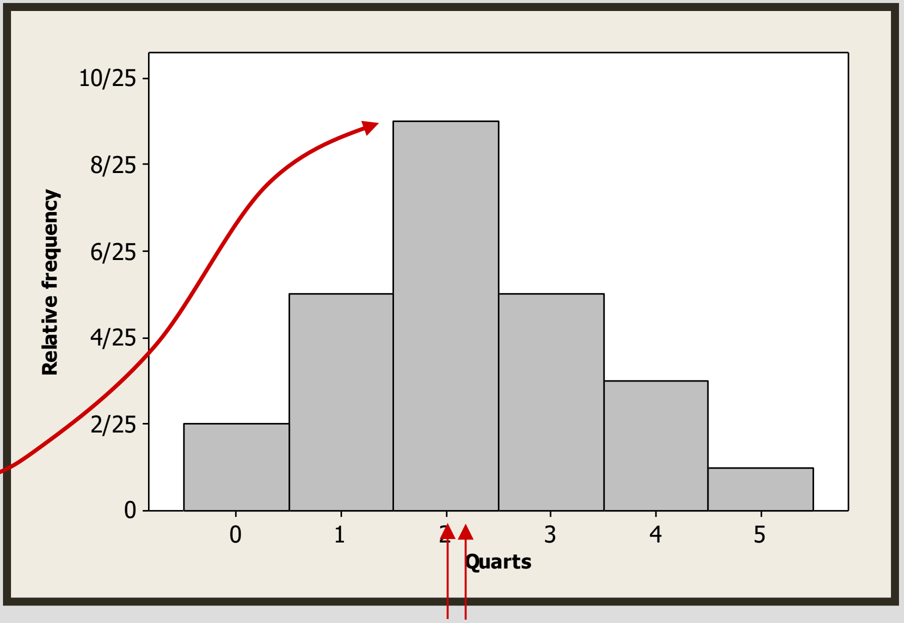
$$\bar{x} = \frac{\sum x_i}{n} = \frac{55}{25} = 2.2$$

- **Median?**

$$m = 2$$

- **Modus? (Puncak Tertinggi)**

$$\text{modus} = 2$$



Median

- **Median** adalah pengamatan pada posisi tengah dari gugus data setelah gugus data tersebut diurutkan dari kecil ke besar
- **Posisi dari median** adalah

$$\boxed{0.5(n + 1)}$$

Setelah pengamatan diurutkan.

Contoh

- Gugus data : 2, 4, 9, 8, 6, 5, 3 $n = 7$
- Diurutkan : 2, 3, 4, 5, 6, 8, 9
- Posisi : $0.5(n + 1) = 0.5(7 + 1) = 4^{th}$

Median = 5, yaitu pengamatan ke-4

- Gugus data : 2, 4, 9, 8, 6, 5 $n = 6$
- Diurutkan : 2, 4, 5, 6, 8, 9
- Position: $0.5(n + 1) = 0.5(6 + 1) = 3.5^{th}$

Median = (5 + 6)/2 = 5.5 — rataan pengamatan ke-3 dan ke-4

Persentil (Percentile)

Persentil ke-p dari suatu gugus data adalah suatu nilai yang membagi gugus data menjadi dua bagian sehingga ada p persen data berada di sebelah kiri atau sama dengan nilai tersebut.

Langkah-langkah:

- Urutkan data dari kecil ke besar,
- **Posisi** dari persentil ke-p adalah:

$$0.p(n+1)$$

- Jika hasilnya bukan merupakan bilangan bulat, tentukan persentil dengan interpolasi

Contoh

Tentukan persentil ke-25, ke-75, dan ke-90 dari gugus data berikut:

53, 55, 70, 58, 64, 57, 53, 78, 57, 68, 53

Diurutkan 53, 53, 53, 55, 57, 57, 58, 64, 68, 70, 78

Persentil ke-25=P_{25}

Posisinya ialah $0.25(11+1)=3$ → $P_{25} = 53$

Persentil ke-75=P_{75}

Posisinya ialah $0.75(11+1)=9$ → $P_{75} = 68$

Persentil ke-90=P_{90}

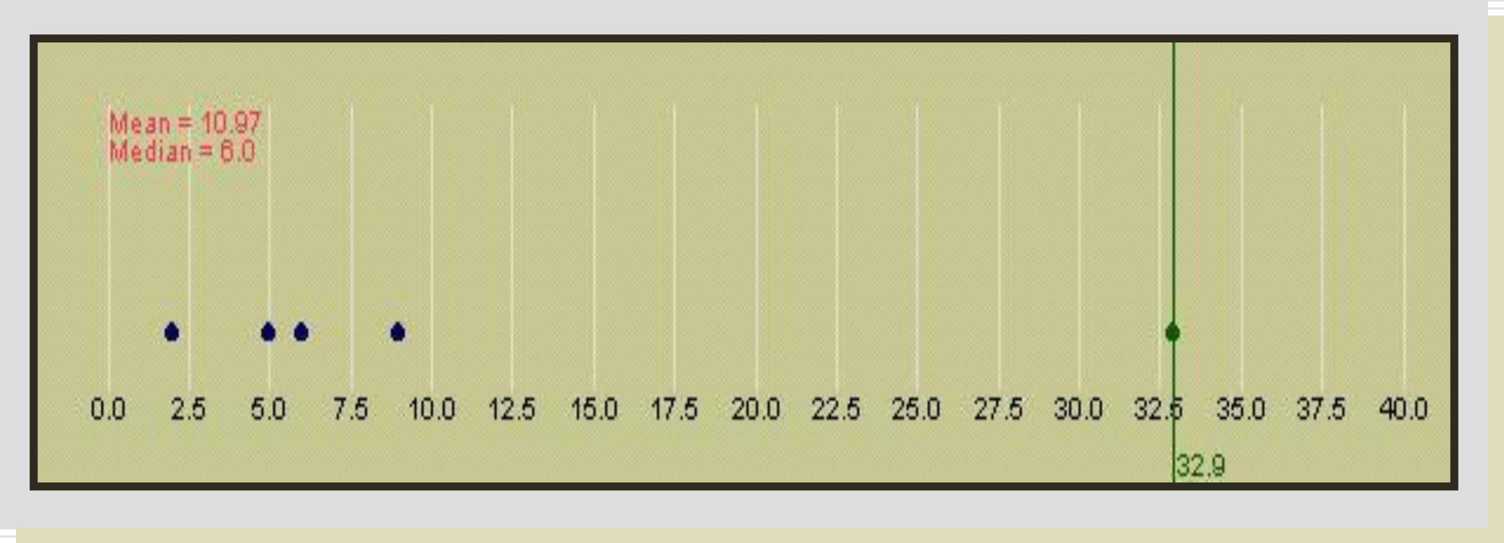
Posisinya ialah $0.90(11+1)=10.8$

P_{90} = 70+0.8(78-70) = 76.4

Catatan Tambahan untuk Ukuran Pemusatan

Nilai-nilai ekstrim

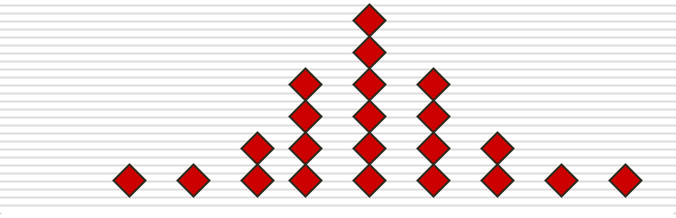
- Rataan sangat mudah dipengaruhi oleh nilai data yang ekstrim daripada median.



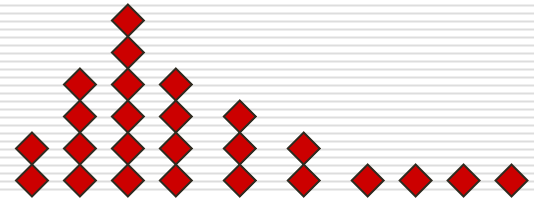
- Median sering kali digunakan sebagai ukuran pemusatan jika datanya tidak simetris.

Catatan Tambahan untuk Ukuran Pemusatan

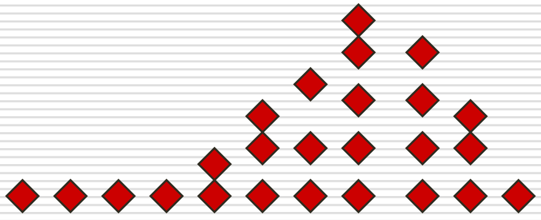
Nilai-nilai ekstrim



Symmetric: Mean = Median



Skewed right: Mean > Median



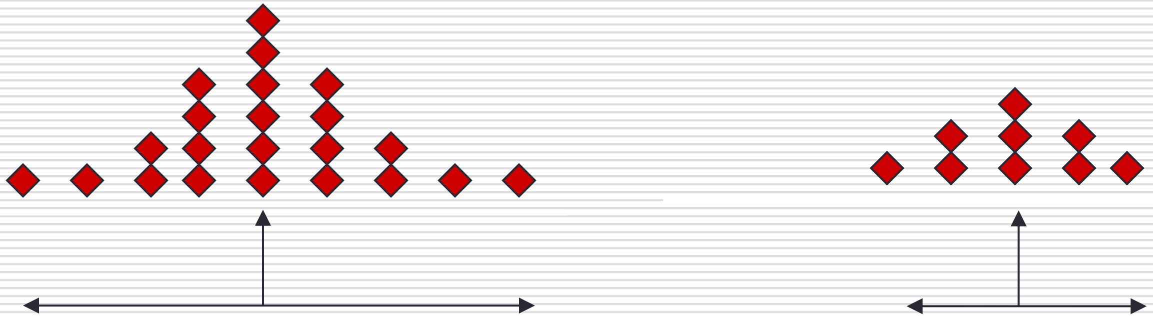
Skewed left: Mean < Median

Catatan Tambahan untuk Ukuran Pemusatan

- **Use Mean** when distribution is reasonably symmetrical, with few extreme scores and has one mode.
- **Use Median** with nonsymmetrical distributions because it is not sensitive to skewness.
- **Use Mode** when dealing with frequency distribution for nominal data

2. Ukuran Penyebaran

- Suatu ukuran sepanjang garis horisontal dari sebaran data yang menggambarkan **penyebaran** sebaran data terhadap ukuran pemusatannya.



- **Jangkauan (Range)**

✓Selisih antara maksimum dan minimum

- **Jangkauan Antar Kuartil (JAK) atau Interquartile Range (IQR)**

✓Selisih antar Kuartil ke-3 dan Kuartil ke-1 **(Q_3 - Q_1)**

- **Ragam (Variance)**

✓Rataan dari selisih kuadrat antara pengamatan dan rataannya

- **Simpangan Baku (Standard Deviation)**

✓Akar dari ragam

Jangkauan (Range)

- Jangkauan atau **range, R,** dari *n* pengamatan adalah perbedaan antara pengamatan terbesar dan terkecil.
- **Contoh:** Catatan banyaknya barang yang terjual dalam 5 hari ialah:

5, 12, 6, 8, 14

- Jangkauannya adalah

$$R = 14 - 5 = 9.$$

Cepat, mudah, tetapi hanya menggunakan 2 nilai pengamatan.

Jangkauan Antar Kuartil (JAK)
Inter-Quartile Range (IQR)

- **Kuartil bawah dan Kuartil atas (Q_1 and Q_3),** dapat dihitung sebagai berikut:

Posisi $Q_1 = P_{25}$ adalah $0.25(n+1)$

Posisi $Q_3 = P_{75}$ adalah $0.75(n+1)$

Persentil ke-50 = P_{50} = Median (Q_2)

Persentil ke-25 = P_{25} = Kuartil bawah (Q_1)

Persentil ke-75 = P_{75} = Kuartil atas (Q_3)

Jangkauan Antar Kuartil (JAK) $= Q_3 - Q_1$

Contoh

Harga 18 merk sepatu yang sudah diurutkan adalah:

40 60 65 65 65 68 68 70 70 70 70 70 70 74 75 75 90 95

Posisi $Q_1 = 0.25(18 + 1) = 4.75$

Posisi $Q_3 = 0.75(18 + 1) = 14.25$

Interpolasi:

Q_1 berada di antara data ke-4 dan ke-5 setelah diurutkan

$Q_1 = 65 + 0.75(65 - 65) = 65.$

Q_3 berada di antar data ke-14 dan ke-15 setelah diurutkan

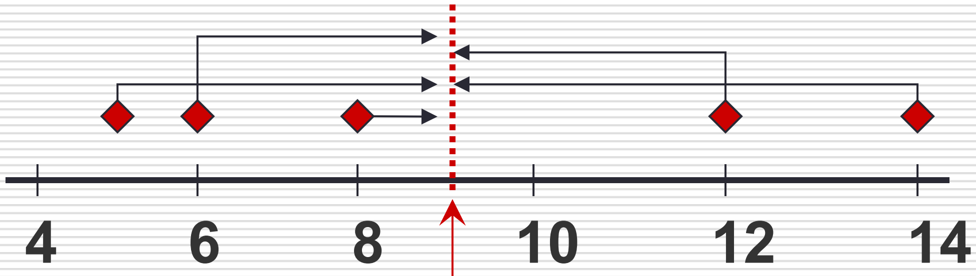
$Q_3 = 74 + .25(75 - 74) = 74.25$

JAK = $Q_3 - Q_1 = 74.25 - 65 = 9.25$

Ragam (Variance)

- **Ragam** adalah ukuran keberagaman yang menggunakan semua pengamatan. Ragam mengukur perbedaan antara setiap pengamatan terhadap rataannya.
- **Data Penjualan:** **5, 12, 6, 8, 14**

$$\bar{x} = \frac{45}{5} = 9$$



Ragam (Variance)

- **Ragam Populasi** dari N pengamatan adalah rataan dari kuadrat defiasi antara pengamatan-pengamatan terhadap rataannya, μ.

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$$

- **Ragam Contoh** dari n pengamatan adalah jumlah kuadrat dari defiasi setiap pengamatan terhadap rataannya, dibagi dengan $(n - 1)$.

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

Standard Deviation (Simpangan Baku)

- Dalam penghitungan ragam, kita mengkuadratkan semua defiasi, akibatnya satuan pengukuran juga berubah.

- Untuk mengembalikannya ke satuan pengukuran aslinya, kita menghitung **simpangan baku**, yaitu akar kuadrat positif dari ragam.

$$\text{Simpangan baku populasi} : \sigma = \sqrt{\sigma^2}$$

$$\text{Simpangan baku contoh} : s = \sqrt{s^2}$$

Keragaman dan Simpangan Baku (Contoh)

	x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
	5	-4	16
	12	3	9
	6	-3	9
	8	-1	1
	14	5	25
Jumlah	**45**	**0**	**60**

Menggunakan rumus:

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

$$= \frac{60}{4} = 15$$

$$s = \sqrt{s^2} = \sqrt{15} = 3.87$$

Bentuk Sebaran Data

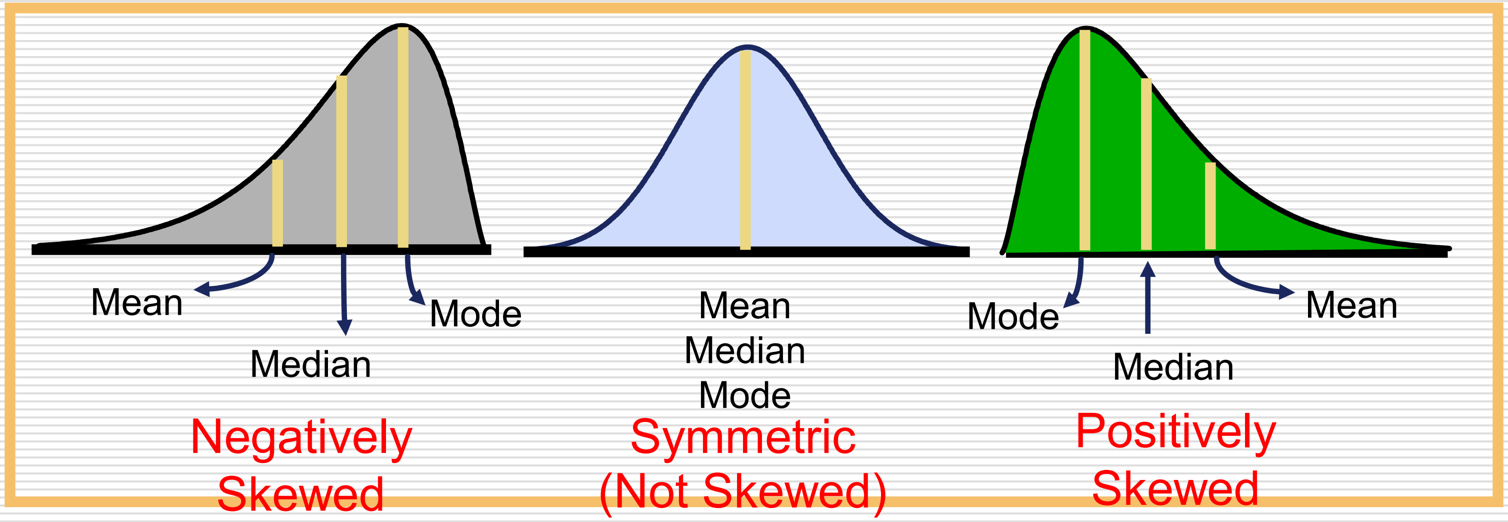
- Ukuran Pemusatan dan Ukuran Penyebaran memberikan informasi tentang lokasi dan keragaman suatu distribusi.
- Selain ukuran-ukuran tersebut, kita juga perlu memiliki gambaran tentang bentuk distribusi.
- Ukuran kemenjuluran (Skewness) memberikan arah dan besarnya ketidaksimetrisan sedangkan Ukuran Keruncingan (Kurtosis) memberikan ide tentang Keruncingan suatu distribusi.

Kemenjuluran (Skewness)

- ✓ Kurangnya kesimetrisan disebut skewness untuk suatu distribusi frekuensi.
- ✓ Jika distribusinya tidak simetris, frekuensi tidak akan terdistribusi merata di sekitar pusat distribusi
- ✓ Dalam Matematika, suatu bangun disebut simetris jika terdapat suatu titik di dalamnya yang melaluinya jika garis tegak lurus ditarik pada sumbu X, ia membagi bangun tersebut menjadi dua bagian yang kongruen yaitu identik dalam segala hal atau satu bagian dapat ditumpangkan pada yang lain. yaitu seperti cermin satu sama lain.

Kemenjuluran (Skewness)

- ✓ Dalam Statistika, suatu distribusi disebut simetris jika nilai rata-rata, median dan modus kurang lebih sama. Jika tidak, distribusi menjadi asimetris.
- ✓ Jika ekor kanan lebih panjang, kita mendapatkan distribusi menjulur positif di mana nilai Rataan > median > mode sedangkan jika ekor kiri lebih panjang, kita mendapatkan distribusi yang menjulur negative, di mana Rataan < Median < Modus.



Ukuran Kemenjuluran

- ✓ Ukuran kemenjuluran membantu kita untuk mengetahui sejauh mana dan ke arah mana (positif atau negatif) distribusi frekuensi menyimpang dari simetri.
- ✓ Meskipun skewness positif atau negatif dapat dideteksi secara grafis tergantung pada apakah ekor kanan atau kiri lebih panjang, tetapi kita tidak mengetahui besarnya.
- ✓ Batas antara simetri dan asimetri mungkin sulit dideteksi secara grafis.

Ukuran Kemenjuluran

- ✓ Beberapa ukuran statistik diperlukan untuk menentukan besarnya ketidaksimetrisan.
- ✓ Ukuran kemenjuluran yang baik harus memiliki tiga kriteria:

1. Harus berupa angka tanpa satuan sehingga bentuk distribusi yang berbeda dapat dibandingkan meskipun satuan dari variabel yang mendasarinya berbeda;

2. Jika distribusinya simetris, nilai ukurannya harus nol. Demikian pula, ukuran tersebut harus memberikan nilai positif atau negatif sesuai dengan distribusi masing-masing yang memiliki kemiringan positif atau negatif; dan

3. Saat kita bergerak dari kemenjuluran negatif yang ekstrim ke kemenjuluran positif yang ekstrim, nilai ukuran harus bervariasi menyesuaikan.

Ukuran Kemenjuluran

- ✓ Ukuran kemenjuluran (skewness) dapat bersifat absolut dan relatif.
- ✓ Karena dalam distribusi simetris mean, median dan modus identik, semakin besar mean menjauh dari modus, semakin besar asimetri atau skewnessnya.
- ✓ Ukuran skewness absolut tidak dapat digunakan untuk keperluan perbandingan karena besarnya skewness yang sama mempunyai arti yang berbeda pada distribusi dengan variasi kecil dan pada distribusi dengan variasi besar.

Ukuran Kemenjuluran Absolut

1. $S_k = Mean - M_{\boldsymbol{e}}$
2. $S_k = Mean - M_{\boldsymbol{o}}$
3. $S_k = (Q_3 - Q_2) - (Q_2 - Q_1)$

Ukuran Kemenjuluran Relatif

Untuk membuat perbandingan yang valid antara kemenjuluran dua distribusi atau lebih, kita harus menghilangkan pengaruh variasi distribusi.

Penghapusan tersebut dapat dilakukan dengan membagi kecondongan absolut dengan simpangan baku.

Berikut ini adalah beberapa metode untuk mengukur kemenjuluran relatif:

- Metode Karl Pearson
- Metode Bowley
- Metode Kelly

Ukuran Kemenjuluran Pearson

- Metode ini paling sering digunakan untuk mengukur kemenjuluran.

$$S_k = \frac{Mean - M_0}{\sigma}$$

  - Jika $S_k = 0$ → data memiliki distribusi simetris
  - Jika $S_k > 0$ → data memiliki distribusi menjulur ke kanan
  - Jika $S_k < 0$ → data memiliki distribusi menjulur ke kiri

- Nilai koefisien kemiringan Karl Pearson biasanya terletak antara ±1 untuk distribusi menjulur sedang.

Ukuran Kemenjuluran Pearson

- Jika modus tidak didefinisikan dengan baik atau tidak diketahui,maka koefisien kemenjuluran bisa dihitung dengan rumus:

$$S_k = \frac{3(Mean - M_e)}{\sigma}$$

- Jika menggunakan hubungan:

$$Mode = 3\,Me\; - 2\,Mean$$

- Maka

$$-3 \;\leq S_k \leq 3$$

Ukuran Kemenjuluran Bowley

- Koefisien dihitung berdasarkan nilai-nilai kuartil

$$S_k = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_1)}$$

S_k = 0 → symmetric (Q_2-Q_1 = Q_3-Q_2)
S_k = + → positively skewed (Q_2-Q_1 < Q_3-Q_2)
S_k = - → negatively skewed (Q_2-Q_1 > Q_3-Q_2)

$S_k = \pm 0{,}1$ (not significantly skewed)
$S_k > \pm 0{,}3$ (significantly skewed)

Ukuran Kemenjuluran Kelly

- Koefisien dihitung berdasarkan nilai-nilai persentil dan desil

Berdasarkan Persentil

$$S_k = \frac{(P_{90} - P_{50}) - (P_{50} - P_{10})}{(P_{90} - P_{10})}$$

Berdasarkan Desil

$$S_k = \frac{(D_9 - 2D_5 + D_1)}{(D_9 - D_1)}$$

Skewness

$$G_1 = \frac{k_3}{k_2^{3/2}} = \frac{n^2}{(n-1)(n-2)} \; \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^3}{\left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2\right]^{3/2}}$$

Keruncingan (Kurtosis)

- Jika kita memiliki pengetahuan tentang Ukuran Pemusatan, Ukuran Penyebaran, dan Kemenjuluran, kita belum memperoleh gagasan lengkap tentang distribusi data tersebut.
- Kita perlu mengetahui ukuran lain untuk mendapatkan gambaran lengkap tentang bentuk distribusi yang dapat dipelajari dengan bantuan Kurtosis.

Keruncingan (Kurtosis)

Berdasarkan keruncingan, kurva distribusi dapat dibedakan menjadi 3 macam, yaitu:

Leptokurtik

Distribusi yang memiliki puncak relatif tinggi

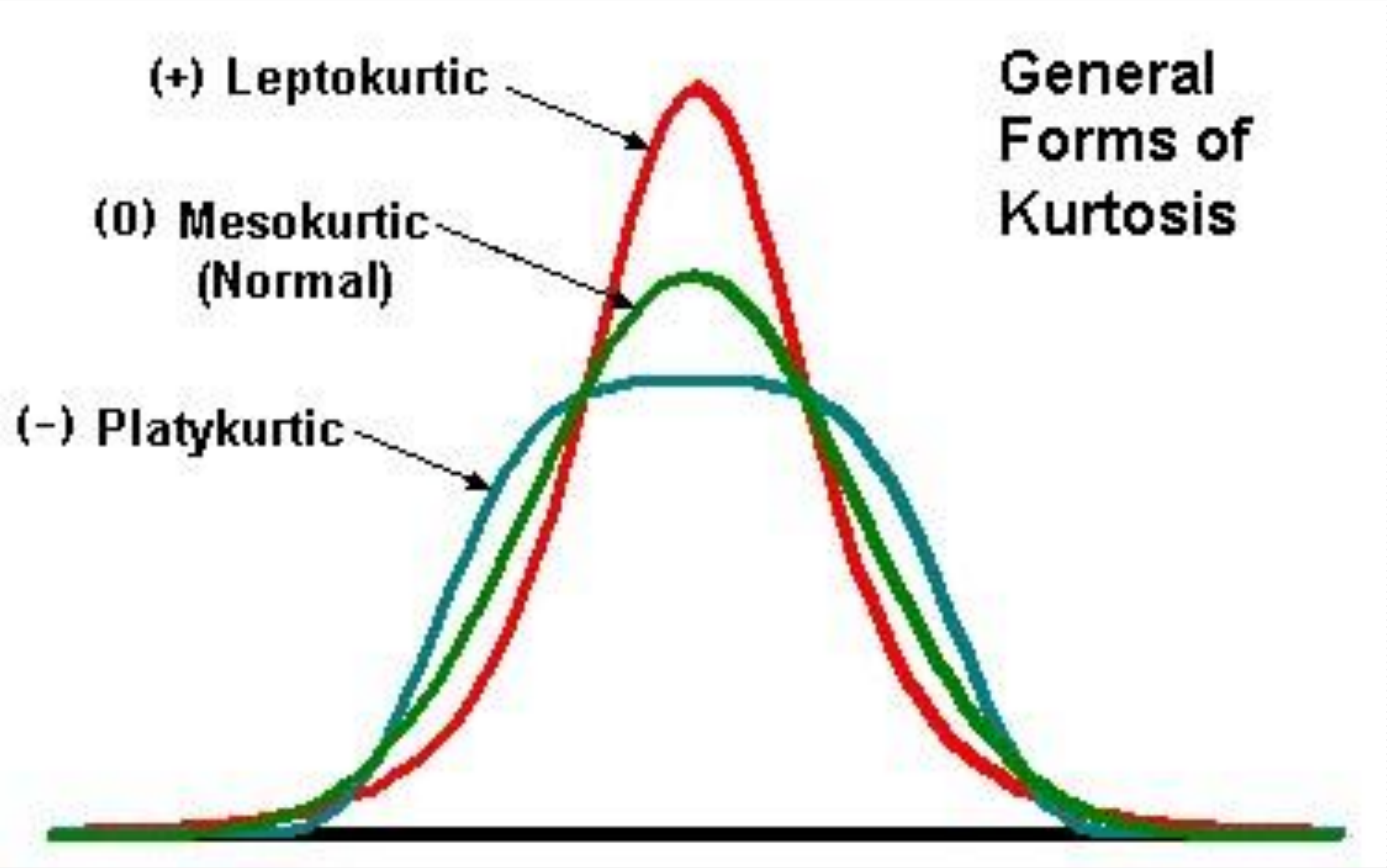
Platikurtik

Distribusi yang memiliki puncak hampir mendatar

Mesokurtik

Distribusi yang memiliki puncak yang tinggi dan tidak mendatar

Keruncingan (Kurtosis)



Ukuran Keruncingan (Kurtosis)

- Beberapa ukuran keruncingan, antara lain:

Metode Karl Pearson

Untuk menghitung kurtosis digunakan momen sentral kedua (μ_2) dan keempat (μ_4), sebagai berikut:

$$\beta_2 = \frac{\mu_4}{\mu_2^2} \quad \text{atau} \quad \gamma_2 = \beta_2 - 3$$

- Jika $\beta_2 = 3$ atau $\gamma_2 = 0$ → Mesokurtik
- Jika $\beta_2 < 3$ atau $\gamma_2 < 0$ → Platikurtik
- Jika $\beta_2 > 3$ atau $\gamma_2 > 0$ → Leptokurtik

Metode Bowley

Untuk menghitung kurtosis digunakan rumus sebagai berikut:

$$\beta_2 = \frac{P_{75} - P_{25}}{P_{90} - P_{10}}$$

- Jika $\beta_2 > 0.263$ → Platikurtik
- Jika $\beta_2 < 0.263$ → Leptokurtik
- Jika $\beta_2 = 0.263$ → Mesokurtik

Tugas

- Hitung rataan, simpangan baku, skewness dan kurtosis dari
  - total pengeluaran rumah tangga sebulan terakhir di salah satu kabupaten/kota
  - proporsi pengeluaran makanan rumah tangga di salah satu kabupaten/kota
  - total pengeluaran rumah tangga sebulan terakhir di salah satu kabupaten/kota, masing-masing untuk rumah tangga yang memiliki mobil dan yang tidak memiliki mobil

- Catatan: kabupaten/kota yang dipilih, berbeda-beda antar kelompok