

# STATISTIKA DAN PROBABILITAS

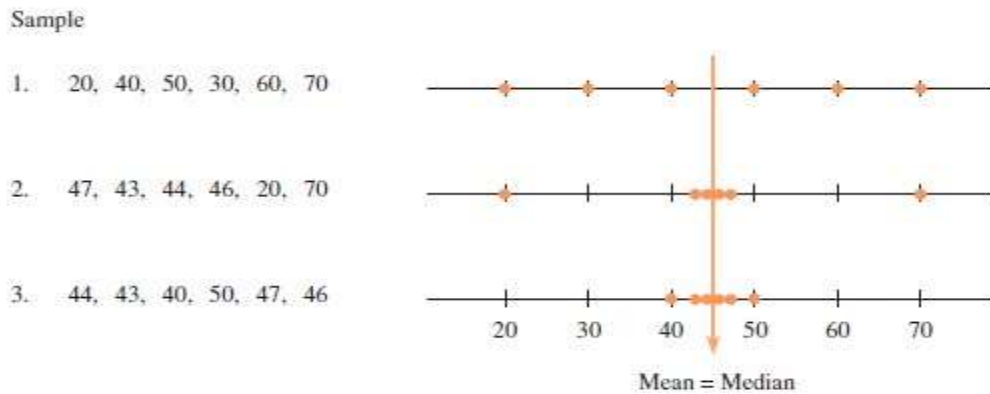
## UKURAN PENYEBARAN DATA, SKEWNESS DAN KURTOSIS

### UKURAN PENYEBARAN DATA

**Rata-rata hitung (mean), median, modus adalah ukuran pemusatan data** yang memberikan informasi tentang bagaimana data-data ini mengumpul atau memusat. Selain memusat, **data juga tersebar disekitar ukuran pemusataanya**. Sebagaimana ukuran pada pemusatan data, penyebaran data juga memiliki ukuran. Ukuran ini digunakan untuk mengetahui variasi atau dispersi data, yaitu derajat penyebaran data terhadap nilai rata-rata. Ukuran penyebaran data yang sering digunakan adalah range, rata-rata deviasi, innterquartile, dan standar deviasi.

#### ***A. UKURAN PENYEBARAN DATA***

Selain ukuran pemusatan data, statistika masih memiliki ukuran lain yaitu ukuran penyimpangan atau ukuran variasi atau ukuran penyebaran(disversi) data. Ukuran dispersi adalah ukuran yang menyatakan seberapa banyak nilai-nilai data yang berbeda dengan nilai pusatnya atau seberapa jauh penyimpangan nilai-nilai data tersebut dari nilai pusat. Seperti yang ditunjukkan pada gambar berikut :



Ada banyak variabilitas dalam sampel pertama dibandingkan dengan sampel ketiga. Sampel kedua menunjukkan variabilitas kurang dari variabilitas pertama dan lebih dari yang ketiga, sebagian besar variabilitas dalam sampel kedua ini disebabkan oleh dua nilai ekstrim yang begitu jauh dari pusat. Ukuran ini juga menggambarkan derajat berpencarnya data kuantitatif.

Ukuran variasi pada dasarnya merupakan pelengkap dari ukuran nilai pusat dalam rangka penggambaran sekumpulan data, karena ukuran nilai pusat secara terpisah tidaklah dapat menggambarkan keadaan keseluruhan data dengan baik. Ukuran nilai pusat tersebut hanya memberikan informasi tentang sebuah nilai dimana nilai-nilai data yang lain berpencar atau dengan kata lain setengah dari keseluruhan data berada sebelum nilai pusat dan setengahnya lagi berada setelah nilai pusat. Berikut adalah beberapa fungsi atau kegunaan ukuran dispersi:

- Ukuran penyebaran dapat digunakan untuk menentukan apakah nilai rata-ratanya benar-benar representatif atau tidak. Apabila suatu kelompok data mempunyai penyebaran yang tidak sama terhadap nilai rata-ratanya, maka dikatakan bahwa nilai rata-rata tersebut tidak representatif.
- Ukuran penyebaran dapat digunakan untuk mengadakan perbandingan terhadap variabilitas data.

- c. Ukuran penyebaran dapat membantu penggunaan ukuran statistika misal dalam pengujian hipotesis untuk menentukan apakah dua sampel berasal dari populasi yang sama atau tidak.

Jenis ukuran dispersi meliputi jangkauan (range) yang terdiri dari jangkauan antar kuartil dan interkuartil, simpangan rata-rata, simpangan baku, varians, koefisien variasi, dan angka baku.

## 1 UKURAN JANGKAUAN (RANGE)

Range merupakan ukuran variasi yang paling sederhana dan paling mudah ditentukan nilainya. Range R merupakan selisih nilai tertinggi  $X_{\max}$  data observasi dengan data terendahnya  $X_{\min}$  dan dirumuskan sebagai berikut.

$$R = X_{\max} - X_{\min}$$

Contoh :

Jumlah Penumpang Kereta Tahun 2011( dalam ribu)

9273    9678    9692    9777    9852    10147  
10152    10188    10354    10513    10733    10749

*Sumber : PT Kereta Api Indonesia*

$$\text{Range (R)} = 10749 - 9273 = 1476$$

Jika nilai-nilai observasi telah dikelompokkan ke dalam distribusi frekuensi, maka jangkauan distribusi dirumuskan sebagai beda antara pengukuran nilai titik tengah kelas pertama dan nilai titik tengah kelas terakhir.

Contoh :

Interval	Frekuensi
9273-9597	1
9597-9921	4
9922-10246	3
10247-10571	2
10572-10896	2

Nilai tengah kelas pertama =  $(9273+9597)/2 = 9435$

Nilai tengah kelas terakhir =  $(10572+10896)/2 = 10734$

$$\text{Range } R = 10734 - 9435 = 1299$$

## 2 RATA-RATA SAMPANGAN

Ukuran variabilitas yang juga banyak digunakan untuk mendeskripsikan sejauh mana sampel pengamatan menyimpang dari rata-rata sampel  $\bar{x}$  adalah rata-rata penyimpangan dari mean atau rata-rata simpangan. Rata-rata

Simpangan untuk data tunggal dirumuskan sebagai

$$S_x = \frac{\sum_1^n |x_i - \bar{x}|}{n}$$

Untuk data kelompok dirumuskan sebagai

$$S_x = \frac{\sum_1^n f_i |x_i - \bar{x}|}{n}$$

Contoh :

Tentukan rata-rata simpangan data berikut :

6092	5249	5851	5843	6505	6659	6883	4814	6661	5910	5913	6556
------	------	------	------	------	------	------	------	------	------	------	------

Rata-rata ( $\bar{x}$ ) = 5635

$x_i$	$ x_i - \bar{x} $
6092	457
5249	-386
5851	216
5843	208
6505	870
6659	1024
6883	1248

4814	-821
6661	1026
5910	275
5913	278
6556	921
$\Sigma = 5316$	

$$S_x = \frac{\sum_1^n |x_i - \bar{x}|}{n} = \frac{\sum_1^{12} |x_i - \bar{x}|}{12}$$

$$= \frac{5316}{12} = 443$$

Contoh :

$x_i$	Frekuensi	$f_i x_i$	$ x_i - \bar{x} $	$f_i  x_i - \bar{x} $
55	1	55	-20,56	-20,56
60	4	240	-15,56	-62,22
65	4	260	-10,56	-42,22
70	6	420	-5,56	-33,33
75	5	375	-0,56	-2,78
80	3	240	4,44	13,33
85	3	255	9,44	28,33
90	2	180	14,44	28,89
100	1	100	24,44	24,44
	$\Sigma = 29$			$\Sigma = -66,12$

Rata-rata = 75,56

$$S_x = \frac{\sum_1^n f_i |x_i - \bar{x}|}{n} = \frac{-66,12}{29}$$

$$= -2,28$$

### 3 SIMPANGAN BAKU ( DEVIASI STANDAR )

Untuk populasi yang berjumlah besar, sangat tidak mungkin untuk mendapatkan nilai rata-rata populasi  $\mu$  serta deviasi standarnya  $\sigma$ . Untuk mengestimasi

(menaksir) nilai  $\mu$  dan  $\sigma$ , diambil sampel data. Nilai  $\mu$  diestimasi oleh  $\bar{x}$  dan  $\sigma$  diestimasi oleh  $s$ .

**Rumus Deviasi Populasi dengan :**

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

$N$  = Jumlah observasi dalam populasi

$\mu$  = Rata-rata populasi.

**Standar Deviasi Sampel**

**Simpangan baku** atau deviasi standar (*Standard Deviation*) merupakan ukuran penyebaran yang paling baik, karena menggambarkan besarnya penyebaran tiap-tiap unit observasi. Karl Pearson menamakannya deviasi standar dan dirumuskan sebagai :

$$S = \sqrt{\frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N - 1}}$$

$N$  = Jumlah sampel

$\bar{X}$  = Rata-rata sampel.

Kuadrat dari deviasi standar dinamakan **variansi** :

Contoh :

Diberikan sample dengan data 6, 7, 8, 9, 10, Hitunglah standar deviasinya (simpangan baku)

Hitung nilai rata-rata sampel :

$$\bar{X} = \frac{6 + 7 + 8 + 9 + 10}{5} = 8$$

$X_i$	$X_i - \bar{X}$	$(X_i - \bar{X})^2$
6	-2	4
7	-1	1
8	0	0
9	1	1
10	2	4
Jumlah ( $\Sigma$ )		10

$$S = \sqrt{\frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N-1}} = \sqrt{\frac{10}{5-1}} = \sqrt{2,5}$$

Standar Deviasi dari data kelompok distribusi frekuensi yang berasal dari sampel didefinisikan :

$$S = \sqrt{\frac{\sum_{i=1}^N f_i (X_i - \bar{X})^2}{N-1}}$$

$N$  = Jumlah sampel

$\bar{X}$  = Rata-rata sampel.

$f_i$  = Frekuensi kelas ke-  $i$

$X_i$  = nilai tengah kelas ke -  $i$

Contoh:

Kelas	Frekuensi ( $f_i$ )	Nilai tengah ( $X_i$ )	$f_i X_i$	$X_i - \bar{X}$	$(X_i - \bar{X})^2$	$f_i (X_i - \bar{X})^2$
50-54	1	52	52	-23,375	546,3906	546,3906
55-59	2	57	114	-18,375	337,6406	675,2813
60-64	11	62	682	-13,375	178,8906	1967,797
65-69	10	67	670	-8,375	70,14063	701,4063
70-74	12	72	864	-3,375	11,39063	136,6875
75-79	21	77	1617	1,625	2,640625	55,45313
80-84	6	82	492	6,625	43,89063	263,3438
85-89	9	87	783	11,625	135,1406	1216,266
90-94	4	92	368	16,625	276,3906	1105,563
95-99	4	97	388	21,625	467,6406	1870,563
	80		6030			8538,75

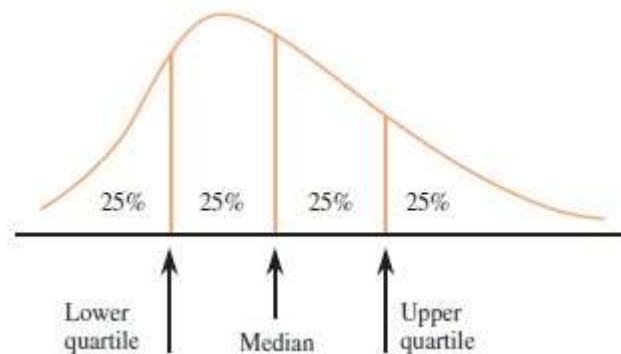
Nilai rata-rata,  $\bar{X} = \frac{6030}{80} = 75,375$

Standar deviasi data berkelompok:

$$s = \sqrt{\frac{\sum_{i=1}^N f_i (X_i - \bar{X})^2}{N - 1}} = \sqrt{\frac{8538,75}{80 - 1}} = \sqrt{108,09} = 10,396$$

#### 4 RANGE INTERQUARTIL

Median didefinisikan sebagai nilai yang membagi seluruh rentang nilai menjadi dua bagian yang sama dan kuartil didefinisikan sebagai nilai yang membagi seluruh rentang nilai menjadi empat bagian yang sama. Rangeinterkuartil adalah ukuran variabilitas berdasarkan kuartil. Perhatikan gambar berikut.



Pada gambar di atas terlihat kuartil bawah memisahkan 25% kumpulan data kebawah dan kuartil atas memisahkan 25% dari kumpulan data ke atas. Kuartil tengah adalah median dan memisahkan 50% kumpulan data. Jika jumlah data n ganjil maka median bisa digantikan dengan nilai kuartil 2 ( $Q_2$ )  
atau

$$Iqr = \text{kuartil atas} - \text{kuartil bawah}$$



$$Iqr = Q_3 - Q_1$$

Pengukuran dispersi atas dasar jangkauan inter-kuartil dinamakan deviasi kuartil atau simpangan kuartil (*quartile deviation*) dan dirumuskan sebagai.

$$D_{iqr} = \frac{\text{Kuartil atas} - \text{kuartil bawah}}{2}$$

atau

$$D_{iqr} = \frac{Q_3 - Q_1}{2}$$

Contoh :

Berikut adalah tabel produksi pulsa telpon di Indonesia

Tahun	Lokal	SLJJ
	(000 pulsa)	(Menit)
1998	16236246427	29668416066
1999	16236724396	31021632143
2000	18516778571	34342636
2001	20227877123	38161484336
2002	19730308403	41397291119
2003	23887950222	42447349726
2004	19936304184	45215914717
2005	22920220767	57745329624
2006	23646924115	61443360381
2007	29018054840	53129188172
2008	22233240642	40706864477

Sumber: Kantor Pusat PT. TELKOM Indonesia

Median = 23887950222

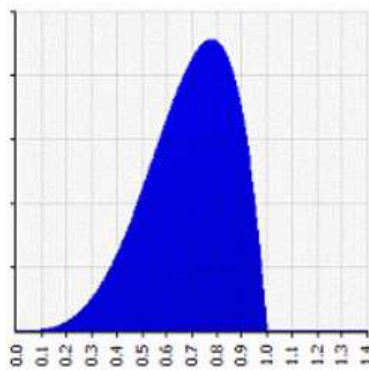
Kuartil bawah = 18516778571

Kuartil atas = 23646924115

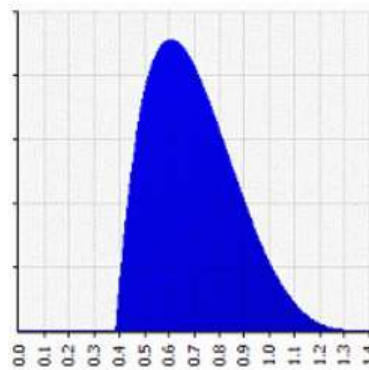
$D_{iqr} = (23646924115 - 18516778571) / 2 = 2565072772$

## B. Skewnes dan Kutosis

Skewnes (Kemiringan) adalah derajat ketidaksimetrisan suatu distribusi. Jika kurva frekuensi suatu distribusi memiliki ekor yang lebih memanjang ke kanan (dilihat dari meannya) maka dikatakan menceng kanan (positif) dan jika sebaliknya maka menceng kiri (negatif). Secara perhitungan, skewness adalah momen ketiga terhadap mean. Distribusi normal (dan distribusi simetris lainnya, misalnya distribusi t atau Cauchy) memiliki skewness 0 (nol). Perhatikan gambar berikut. Kedua gambar memiliki  $\mu = 0.6923$  and  $\sigma = 0.1685$  yang sama tetapi keduanya memiliki kemencengan yang berbeda.



Beta( $\alpha=4.5$ ,  $\beta=2$ )  
skewness = -0.5370



1.3846 - Beta( $\alpha=4.5$ ,  $\beta=2$ )  
skewness = +0.5370

### Rumus koefisien skewnes

$$g_1 = \frac{m_3}{m_2^{3/2}}$$

dimana :

$$m_2 = \frac{\sum_i^n (x_i - \bar{x})^2}{n} \text{ disebut varian}$$

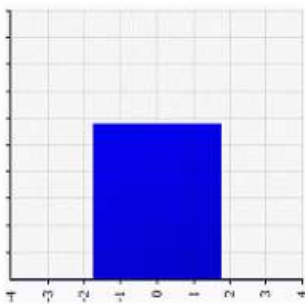
$$m_3 = \frac{\sum_i^n (x_i - \bar{x})^3}{n} \text{ Momen ketiga}$$

### Skewnes sampel data

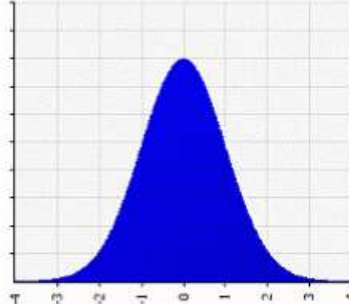
$$G_1 = \frac{\sqrt{n(n-1)}}{n-1} g_1$$

**Kurtosis (keruncingan)** adalah derajat keruncingan suatu distribusi (biasanya diukur relatif terhadap distribusi normal). Kurva yang lebih lebih runcing dari distribusi normal dinamakan leptokurtik, yang lebih datar platikurtik dan distribusi normal disebut

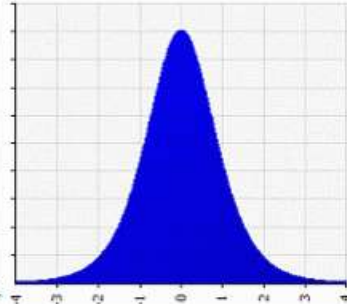
mesokurtik. Kurtosis dihitung dari momen keempat terhadap mean. Distribusi normal memiliki kurtosis = 3, sementara distribusi yang leptokurtik biasanya kurtosisnya >3. Visualisasi kurtosis dapat dilihat pada gambar berikut.



Uniform(min=-√3, max=√3)  
kurtosis = 1.8, excess = -1.2



Normal(μ=0, σ=1)  
kurtosis = 3, excess = 0



Logistic(α=0, β=0.55153)  
kurtosis = 4.2, excess = 1.2

### Rumus Kurtois

$$\text{Kurtosis} \quad a_4 = \frac{m_4}{m_2^2}$$

$$\text{Excess Kurtosis} \quad g_2 = a_4 - 3$$

dimana :

$$m_2 = \frac{\sum_i^n (x_i - \bar{x})^2}{n} \text{ disebut varian}$$

$$m_4 = \frac{\sum_i^n (x_i - \bar{x})^4}{n} \text{ Momen ke empat}$$

### Skewnes sampel data

$$G_2 = \frac{n-1}{(n-2)(n-3)} [(n+1)g_2 + 6]$$

Contoh : Diketahui data tinggi mahasiswa dalam berikut

Tinggi (inches)	Nilai Tengah x	frekuensi f
59.5–62.5	61	5
62.5–65.5	64	18
65.5–68.5	67	42
68.5–71.5	70	27
71.5–74.5	73	8

Tentukan kurtois dan excess kurtoisnya.

Nilai tengah	frekuensi f	$x-\bar{x}$	$(x-\bar{x})^4f$
61	5	-6.45	8653.84
64	18	-3.45	2550.05
67	42	-0.45	1.72
70	27	2.55	1141.63
73	8	5.55	7590.35
		$\Sigma$	19937.60
		$m_4$	1.993.760

Kurtois  $a_4 = m_4 / m_2^2 = 199.3760/8.5275^2 = 2.7418$

Excess kurtosis  $g_2 = 2.7418-3 = -0.2582$

Karena data yang adalah sampel maka

excess kurtosis  $G_2 = [99/(98 \times 97)] [101 \times (-0.2582) + 6] = -0.2091$