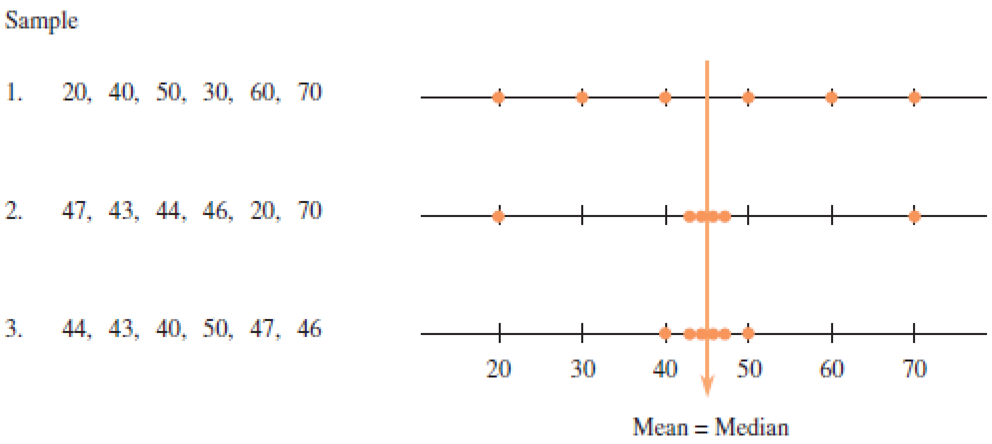
Ukuran Penyebaran Data

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 rata-rata Ukuran penyebaran data yang sering digunakan adalah range, rata-rata deviasi, innterquartile, dan standar deviasi penyebaran data.

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 ditunjukan 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 :

a. 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.
b. 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.

- Ukuran Jangkauan (Range)

  Range merupakan ukuran variasi yang paling sederhana dan paling mudah ditentukan nilainya. Range R merupakan selisih nilai tertinggi Xmax data observasi dengan data terendahnyaXmin dan dirumuskan sebagai berikut.

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

  Contoh :

Jumlah Penumpang Kereta Tahun 2011(dalam 1000)					
9273	9678	9692	9777	9852	10147
10152	10188	10354	10513	10733	10749

Sumber : PT Kereta Api Indonesia

Range R = 10749000 -9273000 = 1476000

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

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

Titik tengah kelas terahir = (10572+10896) =10734

Range R = 10734 – 9435 = 1299

- Rata-rata simpangan

Ukuran variabelitas yang juga banyak digunakan untuk mendeskripsikan sejauh mana sampel pengamatan menyimpang dari rata-rata sampel 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 :

609 2	524 9	585 1	584 3	650 5	665 9	688 3	481 4	666 1	59 1	591 3	655 6

Rata-rata = 5635

x_i	$\lvert x_i\text{-}x\rvert$
6092	457
5249	-386
5851	216
5843	208
6505	870
6659	1024
6883	1248
4814	-821
6661	1026
5910	275
5913	278
6556	921
Σ	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	$f_i x_i$	$\lvert x_i - x\rvert$	$f_i\lvert x_i\text{-}x\rvert$
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
	∑=29			∑= -66,12

Rata-rata = 75,56

$$S_x = \frac{\sum_1^n f_i |x_i - \bar{x}|}{n} = \frac{-66,12}{29} = -2,28$$

- Simpangan baku (deviasi standar)

  Untuk populasi yang berjumlah besar, sangat tidak mungkin untuk mendapatkan nilai rata-rata populasi μ serta deviasi standartny σ. Untuk mengestimasi (

menaksir) nilai μ dan σ , diambil sampel data. Nilai μ diestimasi oleh $\overline{X}$ dan σ diestimasi oleh S .

- Rumus Deviasi Populasi

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

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

dengan :

N = Jumlah observasi dalam populasi

μ = Rata-rata populasi.

- Sampel 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 :

$$S = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{X})^2} \ .$$

Kuadrat dari deviasi standar dinamakan ***variansi*** :

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{X})^2 \ .$$

Untuk distribusi sampel dengan n < 100, Fisher, Wilks dan beberapa statistisi memberi perumusan tentang variansi dan deviasi standar sebagai berikut :

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{X})^2$$

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2}$$

Contoh

Jumlah Penumpang Kereta Api, 2011(000 (

Bulan	Non Jabotabek		
Januari	6092	14	196
Februari	5249	-829	687241
Maret	5851	-227	51529
April	5843	-235	55225
Mei	6505	427	182329
Juni	6659	581	337561
Juli	6883	805	648025
Agustus	4814	-1264	1597696
September	6661	583	339889
Oktober	5910	-168	28224
November	5913	-165	27225
Desember	6556	478	228484

Sumber : PT Kereta Api Indonesia

Rata-rata 6078

$\sum x_i - \bar{x}$ 4183624

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2$$

$$= \frac{1}{12-1}4183624 = \frac{4183624}{11} = 380329{,}5$$

$$S = \sqrt{\frac{1}{12-1}4183624} = \sqrt{\frac{4183624}{11}} = 616{,}7086$$

- Pengukuran dispersi relatif

Pengukuran jangkauan, deviasi kuartil, deviasi rata-rata dan deviasi standar merupakan pengukuran yang absolut. Pengukuran demikian itu sebetulnya hanya dapat digunakan bagi penggambaran dispersi nilai-nilai observasi sebuah distribusi secara definitive. Bila kita ingin melakukan perbandingan tingkat dispersi antara dua atau beberapa distribusi dan bila jumlah nilai-nilai observasi dari dua atau beberapa distribusi di atas tidak sama, maka pengukuran dispersi secara absolut sebagai metode guna membandingkan dispersi akan memperoleh hasil yang menyesatkan.

Dalam membandingkan tingkat variasi dua atau lebih distribusi hendaknya rata-rata distribusi digunakan sebagai dasar pengukuran variasinya secara relatif dan dinamakan **ko-efisien variasi** (*co-efficient of variation*) :

$$V = \frac{S}{\overline{X}}$$

dimana

S = deviasi standar sampel

$\overline{X}$ = rata-rata hitung sampel

Contoh :

Tentukan koefisien variansi tabel berikut

Interval Kelas	*fi*
97-103	4
104-110	8
11-117	15
118-124	35
125-131	25
132-138	6
139-145	4
146-152	3
	100

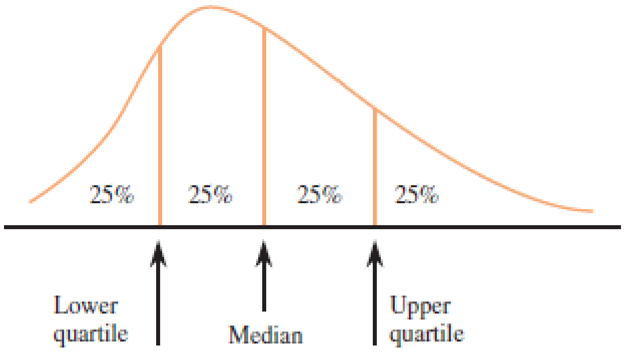
Dari hasil perhitungan diperoleh :

$\overline{X} = 122{,}26$ dan $S_S = 10{,}21$.

Jadi, $V = \frac{10{,}21}{122{,}26}$ x 100% = 8,35%.

- 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. Range interkuartil 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.

Pada distribusi kuartil, 50% dari semua nilai observasi seharusnya terletak di antara Q_1 dan Q_3. Jangkauan antara Q_1 dan Q_3 dinamakan Range Interquartil (*inter-quartile-range*) dan dirumuskan sebagai

Iqr= kuartir atas – kuartir bawah

atau

Iqr= $Q_3 - Q_1$

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

$$D_{iqr} = \frac{Kuartir\ atas - kuartir\ bawah}{2}$$

atau

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

Contoh :

Berikut adalah tabel produksi pulsa telpon di indonesia

Produksi Pulsa Telepon, Indonesia 1988 - 2008

Tahun	Lokal (000 pulsa)	SLJJ (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

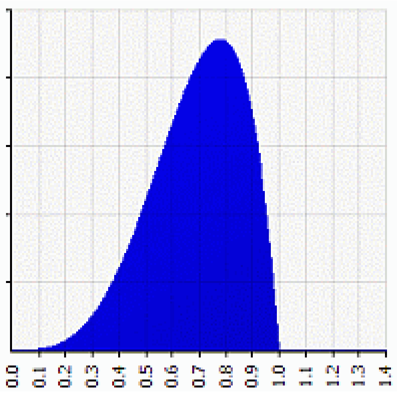
Kuartil bawah = 18516778571

Kuartil atas = 23646924115

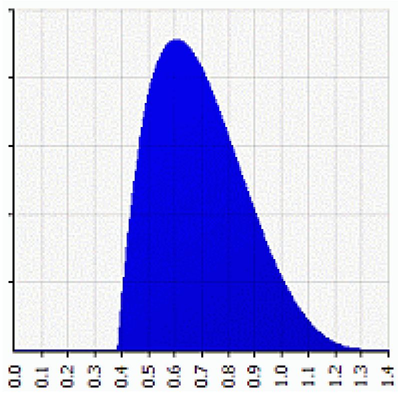
$D_{iqr} = \frac{Q_3 - Q_1}{2}$ = (23646924115- 18516778571)/2 = 2565072772

B. Skewnes dan Kutosis

Skewness 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(α=4.5, β=2)
skewness = −0.5370



1.3846 − Beta(α=4.5, β=2)
skewness = +0.5370

Rumus koefisien skewnes

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

dimana :

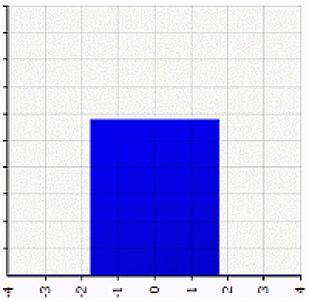
$m_2 = \sum(x-\bar{x})^2 / n$ disebut varian

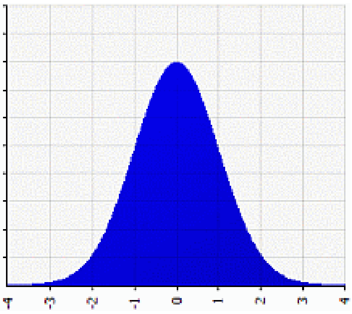
$m_3 = \sum(x-\bar{x})^3 / n$ momem ke tiga

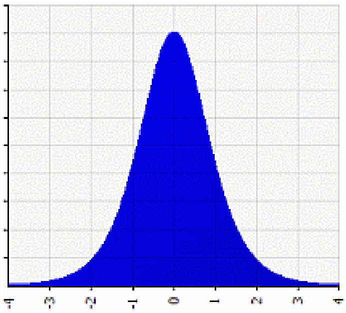
skewnes sampel data

$$G_1 = \frac{\sqrt{n(n-1)}}{n-2} g_1$$

Kurtosis 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

- kurtosis: $a_4 = m_4 / m_2^2$
- excess kurtosis: $g_2 = a_4 - 3$

dimana :

$m_2 = \sum(x-\bar{x})^2 / n$ disebut varian

$m_4 = \sum(x-\bar{x})^4 / n$ momem 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	$x-\bar{x}$	$(x-\bar{x})^4 f$
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
		$\sum$	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 $G_2 = [99/(98\times97)]\ [101\times(-0.2582)+6)] = -0.2091$

Soal

1. Hasil survey tingkat pendidikan penduduk yang lulus sekolah menengah atas diberikan sebagai berikut.

21	27	26	19	30	35	35	26	47	26	27	30
24	29	22	24	29	20	20	27	35	38	25	31
19	24	27	27	23	34	34	25	32	26	26	24
22	28	26	30	23	25	22	25	29	33	34	30
17	25	23									

Dari data diatas

Tentukan

1. Ukuran jangkauan penyebaran data
2. Rata-rata simpangan
3. deviasi standar
4. Variansi
5. Simpangan kuartil
6. Skewnes dan kutoisnya.