

## BAB 4

---

# UKURAN GEJALA PUSAT, LETAK, PENCARAN, KEMIRINGAN DAN KERUNCINGAN

### *Ukuran Gejala Pusat (Measure of Central Tendency)*

Ukuran gejala pusat merupakan suatu ukuran atau nilai yang letaknya cenderung terletak dipusat data. Berikut beberapa penjelasan mengenai ukuran gejala pusat. Smidth dan Sanders (2000:73) menyatakan sebagai berikut.

*“You know from Chapter 2 that there are several measures of central tendency. The purpose of these measures is to summarize in a single value the typical size, middle property, or central location of a set of values. The most familiar measure of central tendency is, of course, the **arithmetic mean**, which is simply the sum of the values of a group of items divided by the number of such items. But you also saw in Chapter 2 that the **median** and **mode** are other measures of central tendency that are commonly used.”*

Spiegel dan Stephens (2008:62) menyatakan sebagai berikut.

*“An **average** is a value that is typical, or **representative**, of a set of data. Since such typical values **tend to lie centrally within a set of data arranged according to magnitude**, averages are also called measures of central tendency.*

*Several types of averages can be defined, the **most common being the arithmetic mean, the median, the mode, the geometric mean, and the harmonic mean. Each has advantages and disadvantages, depending on the data and the intended purpose.”***

Berdasarkan uraian di atas, nilai rata-rata dapat diartikan sebagai nilai tipikal atau representatif atau perwakilan dari suatu set data. Beberapa contoh dari ukuran gejala pusat atau rata-rata adalah rata-rata aritmatik (*arithmetic mean*), median, modus, rata-rata geometrik, dan rata-rata harmonik. Di antara berbagai ukuran gejala pusat tersebut memiliki kelebihan dan kekurangan, bergantung pada data dan tujuan yang dimaksud.

Smidth dan Sanders (2000:73) menyatakan sebagai berikut.

*“Data often have a tendency to congregate about some central value, and this central value may then be used as a summary measure **to describe the general data pattern.**”*

Misalkan diberikan data (sampel) seperti pada Tabel 4.1.

**Tabel 4.1**

Nilai	Nilai	Nilai	Nilai	Nilai
1	5	9	12	16
2	6	10	13	17
3	7	11	14	18
4	8	11	15	

Berdasarkan data pada Tabel 4.1, berikut akan dihitung jumlah keseluruhan nilai (*sum*), rata-rata aritmatik, modus, dan median.

### Jumlah Keseluruhan Nilai (*Sum*)

Andaikan terdapat  $n$  buah nilai, yakni  $X_1, X_2, X_3, \dots, X_n$ . Jumlah dari keseluruhan nilai tersebut dihitung dengan rumus sebagai berikut.

$$\begin{aligned} \text{jumlah keseluruhan nilai} &= \sum_{i=1}^n X_i \\ \sum_{i=1}^n X_i &= X_1 + X_2 + X_3 + \dots + X_n \end{aligned}$$

Jumlah keseluruhan nilai untuk data pada Tabel 4.1 adalah  $1 + 2 + 3 + \dots + 18 = 172$ .

### Rata-Rata Aritmatik atau Rata-Rata Hitung

Rata-rata aritmatik atau sering disebut juga dengan nama rata-rata hitung, merupakan jumlah seluruh nilai dari data, dibagi dengan banyaknya data. Berikut rumus untuk menghitung nilai rata-rata aritmatik (sampel).

$$\begin{aligned} \bar{X} &= \frac{\sum_{i=1}^n X_i}{n} \\ &= \frac{X_1 + X_2 + X_3 + \dots + X_n}{n} \end{aligned}$$

Berikut akan dihitung nilai rata-rata aritmatik berdasarkan data (sampel) pada Tabel 4.1.

$$\bar{X} = \frac{1 + 2 + 3 + \dots + 18}{19}$$

$$\bar{X} = 9,578947$$

Nilai rata-rata aritmatik berdasarkan data pada Tabel 4.1 adalah **9,578947**.

1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 11, 12, 13, 14, 15, 16, 17, 18

Perhatikan bahwa rata-rata hitung 9,57 cenderung terletak di pusat data.

### Modus (*Mode*)

Modus merupakan nilai data dengan frekuensi atau jumlah kemunculan paling banyak. Berdasarkan data pada Tabel 4.1, nilai dengan frekuensi kemunculan paling banyak adalah nilai 11, yakni muncul sebanyak dua kali.

## Median

Spiegel dan Stephens (2008:64) menyatakan sebagai berikut.

*“The median of a set of numbers arranged in order of magnitude (i.e., in an array) is either the middle value or the arithmetic mean of the two middle values.*

*“Geometrically the median is the value of  $X$  (abscissa) corresponding to the vertical line which divides a histogram into two parts having equal areas. This value of  $X$  is sometimes denoted by  $\tilde{X}$ ”.*

Berdasarkan uraian tersebut, median juga disebut juga dengan **nilai tengah** (*middle value*) atau **rata-rata aritmatik** dari dua nilai tengah. Nilai dari median membagi data menjadi dua bagian yang sama. Notasi atau simbol untuk rata-rata aritmatik sampel adalah  $\bar{X}$ , sementara notasi atau simbol median sampel adalah  $\tilde{X}$ . Sebelum menghitung nilai median, terlebih dahulu data diurutkan dari yang terkecil hingga terbesar. Berikut rumus menghitung median untuk data dengan jumlah genap.

$$\text{Median} = \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2}$$

Berikut rumus menghitung median untuk data dengan jumlah ganjil.

$$\text{Median} = X_{\frac{n+1}{2}}$$

Perhatikan bahwa  $X_{\frac{n}{2}}$  merupakan nilai  $X$  yang terletak pada urutan ke- $\frac{n}{2}$ . Sebelum menghitung nilai median, data terlebih dahulu diurutkan dari yang terkecil hingga yang terbesar. Berikut disajikan kembali data pada Tabel 4.1 setelah diurutkan dari yang terkecil hingga terbesar.

1,2,3,4,5,6,7,8,9,10,11,11,12,13,14,15,16,17,18.

Diketahui banyaknya nilai  $n = 19$ , sehingga banyaknya data adalah ganjil.

$$\text{Median} = X_{\frac{n+1}{2}}$$

$$\text{Median} = X_{\frac{19+1}{2}}$$

$$\text{Median} = X_{10}$$

Perhatikan bahwa  $X_{10}$  berarti nilai median terletak pada data dengan urutan ke-10, yakni 10.

1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 11, 12, 13, 14, 15, 16, 17, 18

Nilai median 10 cenderung terletak di pusat data serta nilai median tersebut membagi data menjadi dua bagian yang sama.

Perhatikan bahwa nilai median membagi menjadi dua bagian yang sama. Bagian pertama adalah {1,2,3,4,5,6,7,8,9}, dan bagian kedua adalah {11,11,12,13,14,15,16,17,18}. Perhatikan bahwa masing-masing bagian terdiri dari 9 nilai.

Mann dan Lacke (2011:85) menyatakan sebagai berikut.

*“The median gives the center of a histogram, with half of the data values to the left of the median and half to the right of the median. **The advantage of using the median as a measure of central tendency** is that it is not influenced by outliers. Consequently, the median is preferred over the mean as a measure of central tendency for data sets that contain outliers. For example, when a data set has outliers, instead of using the mean, we can use either the **trimmed mean or median** as a measure of central tendency.”*

Berdasarkan uraian tersebut, keuntungan menggunakan median sebagai ukuran gejala pusat adalah median tidak terpengaruh oleh *outlier* (data pencilan). Oleh karena itu, median lebih disukai dibandingkan rata-rata atau *mean* (rata-rata aritmatik) sebagai ukuran gejala pusat, untuk data yang mengandung *outlier*.

### ***Ukuran Letak (Measure of Position)***

Kuartil dan desil merupakan jenis-jenis dari ukuran letak. Ukuran tersebut membagi data menjadi beberapa bagian yang sama. Sebagai contoh pada ukuran kuartil terdapat tiga buah nilai. Letak dari nilai-nilai kuartil tersebut membagi data menjadi empat bagian yang sama.

#### **Kuartil ( $K$ )**

Ukuran kuartil terdiri dari tiga buah nilai yang membagi data menjadi empat bagian yang sama.

1,2,3,4,5,6,7,8,9,10,11.

Nilai kuartil dikelompokkan atas tiga, yakni kuartil pertama ( $K_1$ ), kuartil kedua ( $K_2$ ), dan kuartil ketiga ( $K_3$ ). Angka 3, 6, dan 9 masing-masing merupakan  $K_1$ ,  $K_2$ , dan  $K_3$ . Berikut rumus untuk menghitung nilai kuartil.

$$K_i = \frac{i(n+1)}{4}; i = 1,2,3$$

Perhatikan bahwa  $K_i$  merupakan nilai dari kuartil ke- $i$  dengan  $i = 1, 2$ , dan  $3$ . Berikut disajikan kembali data pada Tabel 4.1.

1,2,3,4,5,6,7,8,9,10,11,11,12,13,14,15,16,17,18

Diketahui banyaknya nilai data  $n = 19$ . Berikut akan dihitung nilai dari  $K_1$ ,  $K_2$ , dan  $K_3$ .

$$K_1 = \frac{1(19+1)}{4}$$

$$K_1 = 5$$

$K_1 = 5$  berarti nilai  $K_1$  terletak pada data dengan urutan ke-5, yakni 5.

$$K_2 = \frac{2(19 + 1)}{4}$$

$$K_2 = 10$$

$K_2 = 10$  berarti nilai  $K_2$  terletak pada data dengan urutan ke-10, yakni 10.

$$K_3 = \frac{3(19 + 1)}{4}$$

$$K_3 = 15$$

$K_3 = 15$  berarti nilai  $K_3$  terletak pada data dengan urutan ke-15, yakni 14. Ketiga nilai kuartil tersebut membagi data menjadi empat bagian yang sama. Bagian pertama adalah {1,2,3,4}, bagian kedua adalah {6,7,8,9}, bagian ketiga adalah {11,11,12,13}, dan bagian keempat adalah {15,16,17,18}. Perhatikan bahwa banyaknya nilai untuk masing-masing bagian adalah 4.

### Desil ( $D$ )

Ukuran desil terdiri dari sembilan nilai yang membagi data menjadi sepuluh bagian yang sama.

$$1(2)3(4)5(6)7(8)9(10)11(11)12(13)14(15)16(17)18.$$

Perhatikan bahwa nilai-nilai yang dilingkar merupakan nilai-nilai desil. Nilai-nilai tersebut membagi data menjadi 10 bagian yang sama. Masing-masing bagian terdiri dari 1 nilai. Terdapat sembilan nilai desil, yakni desil pertama ( $D_1$ ), desil kedua ( $D_2$ ), dan sampai dengan desil kesembilan ( $D_9$ ). Berikut rumus untuk menghitung nilai desil.

$$D_i = \frac{i(n + 1)}{10} ; i = 1, 2, 3, \dots, 9$$

Berikut akan dihitung nilai desil pertama, kedelapan, dan kesembilan berdasarkan data pada Tabel 4.1.

$$D_1 = \frac{1(19 + 1)}{10} = 2$$

Nilai desil ke-1 terletak pada data dengan urutan ke-2, yakni 2.

$$D_8 = \frac{8(19 + 1)}{10} = 16$$

Nilai desil ke-8 terletak pada data dengan urutan ke-16, yakni 15.

$$D_9 = \frac{9(19 + 1)}{10} = 18$$

Nilai desil ke-9 terletak pada data dengan urutan ke-18, yakni 17. Sembilan nilai desil tersebut membagi data menjadi sepuluh bagian yang sama dengan banyaknya nilai untuk masing-masing bagian adalah 1.

### *Ukuran Pencaran atau Dispersi atau Sebaran*

Misalkan diberikan 4 data, beserta nilainya (Tabel 4.2).

**Tabel 4.2**

Data 1	70	70	70	70	70	$\bar{X} = 70$
Data 2	50	60	70	80	90	$\bar{X} = 70$
Data 3	20	60	70	100	100	$\bar{X} = 70$
Data 4	20	20	10	100	200	$\bar{X} = 70$

Berdasarkan Tabel 4.2, nilai rata-rata untuk data 1 adalah 70, nilai rata-rata untuk data 2 juga 70, begitu juga untuk data 3 dan data 4. Namun **nilai rata-rata untuk data manakah yang dapat mewakili data dengan baik?** Berdasarkan pengamatan, nilai rata-rata dari data 1 dapat mewakili data 1 dengan baik (secara sempurna), nilai rata-rata dari data 2 cukup baik dalam mewakili data 2, namun nilai rata-rata dari data 3 dan data 4 kurang baik dalam mewakili data 3 dan data 4.

Ukuran pencaran atau dispersi merupakan suatu nilai yang mengukur tingkat **pencaran atau sebaran nilai-nilai data terhadap nilai rata-ratanya**. Nilai **pencaran yang tinggi** menunjukkan **nilai-nilai data cenderung terletak cukup jauh terhadap nilai rata-rata** dari data tersebut. Dengan kata lain, data semakin bervariasi atau heterogen. Sebagaimana Mann dan Lacke (2011:92) menyatakan sebagai berikut.

*“Two data sets with the same mean may have different spreads. The variation among the values of observations for one data set may be larger or smaller than for the other data set. (Note that the words dispersion, and variation have the same meaning).”*

*Thus, mean, median, or mode by itself is not a sufficient measure to reveal shape of the distribution of a data set. We also need a measure that can provide some information about the variation among data values. The measures that help us learn about the spread of data set are called the measure of dispersion. The measures of central tendency and dispersion taken together give a better picture of a data set than the measures of central tendency alone. This section discusses three measures of dispersion: range, variance, and standard deviation.”*

Ukuran pencaran yang akan dipaparkan dalam tulisan ini adalah *range*, *variance*, dan standar deviasi. Misalkan diberikan data seperti pada Tabel 4.3.

**Tabel 4.3**

Nilai	Nilai	Nilai	Nilai	Nilai
10	20	30	40	50
10	30	30	40	50
10	30	30	40	50
20	30	30	50	

## Nilai Maksimum

Nilai maksimum merupakan nilai yang paling tinggi dari suatu data. Berdasarkan data pada Tabel 4.3, nilai maksimum adalah nilai 50.

## Nilai Minimum

Nilai minimum merupakan nilai yang paling rendah dari suatu data. Berdasarkan data pada Tabel 4.3, nilai minimum adalah nilai 10.

## Range

*Range* merupakan selisih antara nilai maksimum dengan nilai minimum. Diketahui nilai maksimum adalah 50 dan nilai minimum adalah 10, sehingga nilai *range* adalah  $50 - 10 = 40$ . Ukuran *range* sama seperti rata-rata aritmatik, yakni memiliki kelemahan ketika dalam suatu data mengandung *outlier*. Sebagaimana Mann dan Lacke (2011:93) menyatakan sebagai berikut.

*“The range, like the mean, has the disadvantage of being influenced by outliers. Consequently, the range is not good measure of dispersion to use for a data set that contains outliers.”*

*Another disadvantage of using the range as a measure of dispersion is that its calculation is based on two values only: the largest and smallest. All other values in a data set are ignored when calculating the range. Thus, the range is not very satisfactory measure of dispersion.”*

Sebagai contoh misalkan diberikan data dengan nilai 1, 2, 3, 4, 5, 100. Nilai *range* berdasarkan data tersebut adalah  $100 - 1 = 99$ . Seandainya data dengan nilai 100 tidak diikutsertakan dalam penghitungan nilai *range*, maka diperoleh nilai *range*  $5 - 1 = 4$ . Perhatikan bahwa nilai *range* menurun, dari 100 menjadi 4. Nilai data 100 merupakan *outlier* (data pencilan).

## Variance

*Variance* (dalam hal ini *variance* untuk sampel) dilambangkan dengan  $s^2$ . Berikut rumus untuk menghitung nilai *variance*.

$$s^2 = \frac{|X - \bar{X}|^2}{n - 1}.$$

Nilai *variance* sampel ( $s^2$ ) berdasarkan data pada Tabel 4.3 adalah

$$s^2 = \frac{3 \times |10 - 31,6|^2 + 2 \times |20 - 31,6|^2 + \dots + 4 \times |50 - 31,6|^2}{19 - 1}$$

$$s^2 = 180,7018$$

## Standar Deviasi

Standar deviasi merupakan akar kuadrat positif *variance* ( $\sqrt{s^2} = s$ ). Nilai dari standar deviasi dapat diinterpretasi sebagai nilai yang menunjukkan seberapa dekat nilai-nilai data menyebar atau berkumpul di sekitar rata-ratanya. Standar deviasi merupakan salah satu dari ukuran pencaran yang paling sering digunakan. Mann dan Lacke (2011:93) menyatakan sebagai berikut.

*“The standard deviation is the most-used measure of dispersion. The value of standard deviation tells how closely the values of a data set are clustered around the mean. In general, a lower value of the standard deviation for a data set indicates that the values of that data set are spread over a relatively smaller range around the mean. In contrast, a larger value of the standard deviation for a data set indicates that the values of that data set are spread over a relatively larger range around the mean.”*

Diketahui nilai *variance* adalah 180,7018, sehingga nilai standar deviasi adalah  $\sqrt{180,7018} = 13,4425$ . Tabel 4.4 menyajikan hasil perhitungan untuk nilai minimum, maksimum, *range*, *variance*, dan standar deviasi, berdasarkan data pada Tabel 4.2.

Berdasarkan data pada Tabel 4.4, diketahui nilai standar deviasi untuk data 1 bernilai 0, data 2 bernilai 15,811, data 3 bernilai 33,166, dan data 4 bernilai 81,240. Perhatikan bahwa pada data 1, **seluruh nilai data sama**, yakni seluruhnya 70, sehingga nilai standar deviasinya 0 (begitu juga dengan nilai *range* dan *variance*). Dapat dilihat bahwa semakin besar nilai standar deviasi dari suatu data, maka sebaran data cenderung jauh terhadap rata-ratanya (walaupun ada beberapa data yang dekat dengan rata-ratanya). Perhatikan juga Tabel 4.5.

**Tabel 4.4**

Data						Rata-Rata	Range	Variance	Standar Deviasi
Data 1	70	70	70	70	70	70	0	0	0
Data 2	50	60	70	80	90	70	40	250	15,811
Data 3	20	60	70	100	100	70	80	1100	33,166
Data 4	20	20	10	100	200	70	180	6600	81,240

**Tabel 4.5**

				Rata-Rata	Range	Variance	Standar Deviasi
Data 5	13	14	15	14	2	1	1
Data 6	12	14	16	14	4	4	2
Data 7	8	14	20	14	12	36	6
Data 8	1	14	27	14	26	169	13

Pada Tabel 4.5, nilai rata-rata untuk data 5 sampai data 8 adalah 14. Untuk data 5, jarak 13 ke 14 adalah 1, yakni  $|14 - 13| = 1$ , begitu juga jarak dari 15 ke 14, yakni  $|15 - 14| = 1$ . Nilai standar deviasinya adalah 1. Untuk data 5, jarak dari 12 ke 14 adalah 2, yakni  $|14 - 12| = 2$ , begitu juga jarak dari 16 ke 14, yakni  $|16 - 14| = 2$ . Nilai standar deviasinya adalah 2. Semakin besar nilai standar deviasi dari suatu data, maka sebaran data cenderung jauh terhadap rata-ratanya. Perhatikan juga pada Tabel 4.6.



**Tabel 4.6**

						Rata-Rata	Standar Deviasi
Data 9	14	15	16	17	18	16	1,58113883
Data 10	12	14	16	18	20	16	3,16227766
Data 11	10	13	16	19	22	16	4,74341649
Data 12	8	12	16	20	24	16	6,32455532
Data 13	6	11	16	21	26	16	7,90569415

**Koefisien Variasi (Coefficient of Variation)**

Misalkan diberikan data berat badan dan IQ dari 5 siswa (Tabel 4.7).

**Tabel 4.7**

Siswa	Berat Badan	Uang Jajan
1	54,33	20000
2	58,89	20000
3	64,33	19000
4	54,21	20000
5	53,45	19000
Rata-Rata	57,042	19600
Standar Deviasi	4,604554	547,722558
Koefisien Variasi	0,080722	0,02794503

Andaikan akan dibandingkan, data mana yang lebih bervariasi atau heterogen, apakah data berat badan atau data uang jajan? Perhatikan bahwa **satuan data untuk berat badan (puluhan) dan uang jajan (puluhan ribu) berbeda**. Berdasarkan Tabel 4.7 diketahui nilai standar deviasi dari uang jajan, yakni 547,722, lebih besar dari pada nilai standar deviasi dari berat badan, yakni 4,604. Namun belum tentu berarti bahwa data uang jajan lebih bervariasi atau heterogen dibandingkan data berat badan. Hal ini dikarenakan satuan data berbeda.

Untuk itu dapat digunakan **koefisien variasi** untuk membandingkan tingkat variasi atau heterogen di antara dua atau lebih kelompok, ketika satuan data berbeda-beda. Spiegel dan Stephens (2008:100) menyatakan sebagai berikut.

*“Note that the coefficient of variation is independent of the units used. For this reason, it is useful in comparing distributions where the units may be different. A disadvantage of the coefficient of variation is that it fails to be useful when  $\bar{X}$  is close to zero.”*

Nilai dari koefisien variasi dihitung sebagai berikut.

$$\text{Koefisien Variasi (KV)} = \frac{s}{\bar{X}}$$

Berdasarkan Tabel 4.7, diketahui koefisien variasi untuk data berat badan adalah 0,080722, sementara koefisien variasi untuk data uang jajan adalah 0,02794503. Sehingga data berat badan lebih bervariasi atau heterogen dibandingkan data uang jajan.

### Data yang Dibakukan (*Standardized Data*)

Suatu variabel yang mengukur **deviasi dari rata-rata**, dalam **unit atau satuan standar deviasi**, disebut variabel yang dibakukan (*standardized variable*). Sebagaimana Spiegel dan Stephens (2008:101) menyatakan sebagai berikut.

*“The variable that measures the deviation from the mean in units of the standard deviation is called a standardized variable, is a dimensionless quantity (i.e., is independent of the units used), and is given by*

$$z = \frac{X - \bar{X}}{s}$$

*If the deviations from the mean are given in units of the standard deviation, they are said to be expressed in standard units, or standard scores. These are of great value in the comparison of distributions.”*

Berdasarkan uraian tersebut, data dalam bentuk standar atau baku sangat berguna untuk tujuan perbandingan distribusi dari beberapa kelompok data. Suatu data dari variabel asli **X**, dapat ditransformasi dalam bentuk standar dengan rumus sebagai berikut.

$$Z = \frac{X - \bar{X}}{s}$$

**Tabel 4.8**

Siswa	Berat	Uang Jajan	Z_Baku	Z_Uang Jajan
1	54,33	20000	-0,588982091	0,730296743
2	58,89	20000	0,401341779	0,730296743
3	64,33	19000	1,582780781	-1,095445115
4	54,21	20000	-0,615043245	0,730296743
5	53,45	19000	-0,780097224	-1,095445115
Rata-Rata	57,042	19600	0	0
Standar Deviasi	4,604554	547,722558	1	1
Koefisien Variasi	0,080722	0,02794503		

Berdasarkan Tabel 4.8, nilai standar atau baku untuk uang jajan 20000 adalah 0,730296743. Nilai tersebut diperoleh sebagai berikut.

$$Z = \frac{X - \bar{X}}{s} = \frac{20000 - 19600}{547,722558} = 0,730296743$$

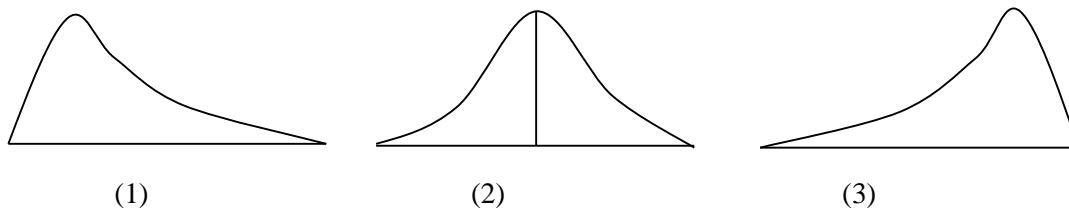
Nilai standar atau baku untuk berat badan 54,33 adalah -0,588982091. Nilai tersebut diperoleh sebagai berikut.

$$Z = \frac{X - \bar{X}}{s} = \frac{54,33 - 57,042}{4,604554} = -0,588982091$$

Data dalam bentuk standar atau baku **memiliki nilai rata-rata 0 dan standar deviasi 1**.

## Ukuran Kemiringan (Skewness)

Ukuran kemiringan atau *skewness* merupakan suatu nilai yang mengukur ketidaksimetrisan distribusi data. Suatu data dikatakan berdistribusi simetris sempurna bila nilai rata-rata, median, dan modus dalam data adalah sama.



**Gambar 4.1**

Pada Gambar 4.1 (1) kurva cenderung condong ke kanan atau disebut kurva positif, sementara Gambar 4.1 (2) kurva bersifat simetris. Pada Gambar 4.1 (3) kurva cenderung condong ke kiri atau disebut kurva negatif. Berikut rumus untuk menghitung nilai kemiringan suatu data.

$$\text{Kemiringan} = \frac{n}{(n-1)(n-2)} \left( \frac{\sum(X - \bar{X})^3}{s^3} \right)$$

Bila nilai kemiringan  $< 0$  atau negatif, maka kurva cenderung condong ke kiri (kurva negatif). Jika nilai kemiringan  $> 0$  atau positif, maka kurva cenderung condong ke kanan (kurva positif). Jika nilai kemiringan mendekati 0 atau 0, maka kurva cenderung simetris. Spiegel dan Stephens (2008:125) menyatakan sebagai berikut.

*“Skewness is the degree of asymmetry, or departure from symmetry, of a distribution. If the frequency curve (smoothed frequency polygon) of a distribution has a longer tail to the right of the central maximum than to the left, the distribution is said to be skewed to the right, or to have positive skewness. If the reverse is true, it is said to be skewed to the left, or to have negative skewness.”*

Misalkan diberikan data seperti pada Tabel 4.9. Berdasarkan data pada Tabel 4.9, berikut akan dihitung nilai kemiringan. Dari Tabel 4.10, diketahui  $\bar{X} = 3,6$  dan  $s = 1,454058$ , sehingga nilai kemiringan dapat dihitung sebagai berikut.

$$\text{Kemiringan} = \frac{n}{(n-1)(n-2)} \left( \frac{\sum(X - \bar{X})^3}{s^3} \right)$$

$$\text{Kemiringan} = \frac{15}{(15-1)(15-2)} \left( \frac{6,48}{1,454058^3} \right)$$

$$\text{Kemiringan} = 0,17372$$

**Tabel 4.9**

Nilai (X)	Nilai (X)	Nilai (X)	Nilai (X)
1	3	4	5
2	3	4	6
2	3	4	6
3	3	5	

**Tabel 4.10**

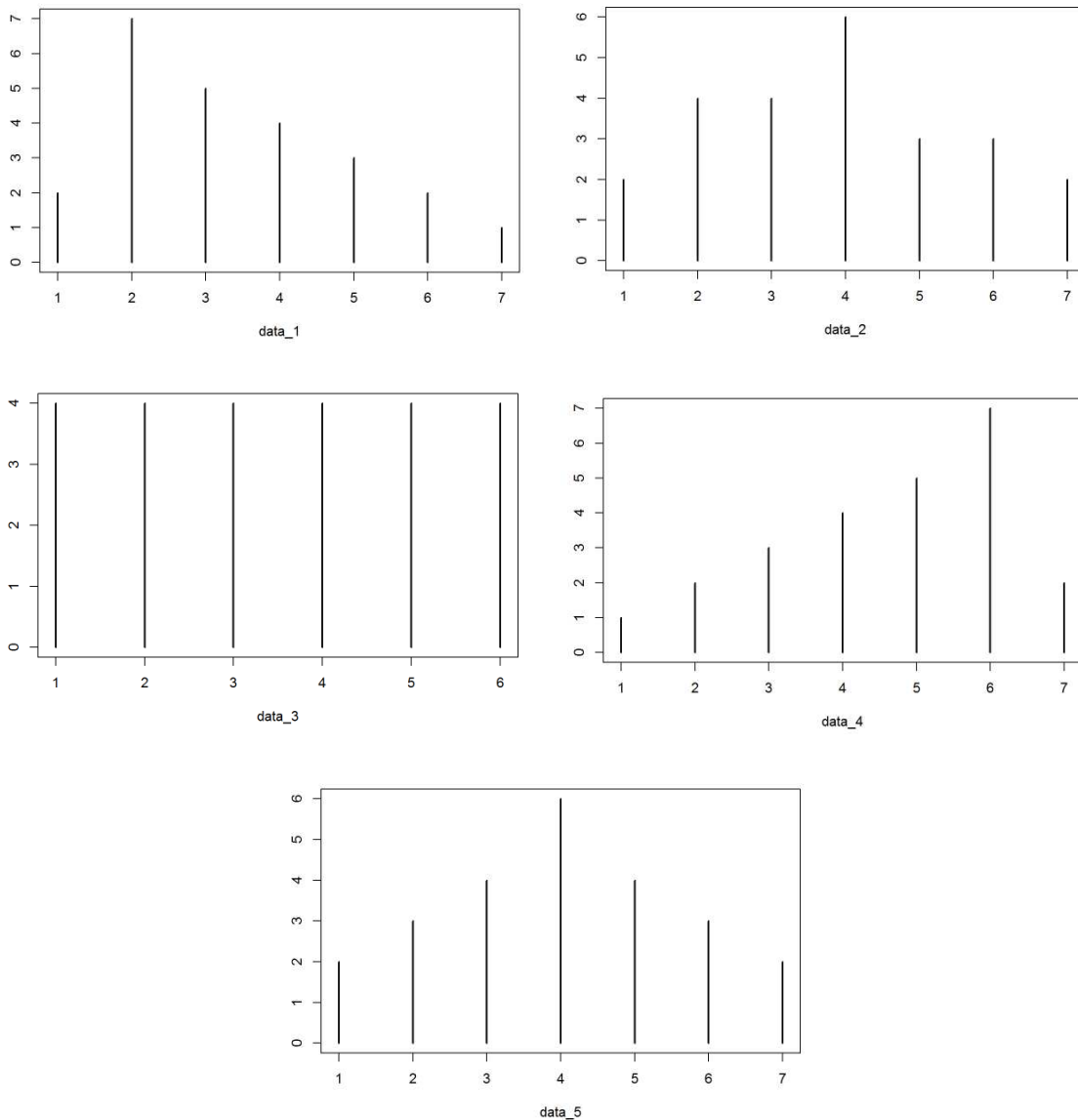
$X$	$f$	$fX$	$f \sum (X - \bar{X})^3$	
1	1	1	-17,576	
2	2	4	-8,192	
3	5	15	-1,08	
4	3	12	0,192	
5	2	10	5,488	
6	2	12	27,648	
Jumlah				
Rata-rata ( $\bar{X}$ )	3,6	15	54	6,48
Standar deviasi ( $s$ )	1,454058			

**Tabel 4.11**

No	Data 1	Data 2	Data 3	Data 4	Data 5
1	1	1	1	1	1
2	1	1	1	2	1
3	2	2	1	2	2
4	2	2	1	3	2
5	2	2	2	3	2
6	2	2	2	3	3
7	2	3	2	4	3
8	2	3	2	4	3
9	2	3	3	4	3
10	3	3	3	4	4
11	3	4	3	5	4
12	3	4	3	5	4
13	3	4	4	5	4
14	3	4	4	5	4
15	4	4	4	5	4
16	4	4	4	6	5
17	4	5	5	6	5
18	4	5	5	6	5
19	5	5	5	6	5
20	5	6	5	6	6
21	5	6	6	6	6
22	6	6	6	6	6
23	6	7	6	7	7
24	7	7	6	7	7
Kemiringan	0,5668	0,1545	0,0000	-0,5668	0,0000
Rata-Rata	3,375	3,875	3,5	4,625	4
Median	3	4	3,5	5	4
Modus	2	4	-	6	4

**Tabel 4.12**

Nilai	Frekuensi				
	Data 1	Data 2	Data 3	Data 4	Data 5
1	2	2	4	1	2
2	7	4	4	2	3
3	5	4	4	3	4
4	4	6	4	4	6
5	3	3	4	5	4
6	2	3	4	7	3
7	1	2	-	2	2



**Gambar 4.2**

Tabel 4.11 menyajikan 5 data, yakni data 1 sampai data 5. Kelima data tersebut masing-masing memiliki nilai data sebanyak 24. Nilai kemiringan untuk data 1 adalah 0,5668, data 2 bernilai 0,1545, data 3 adalah 0, data 4 adalah -0,5668, dan data 5 adalah 0. Perhatikan bahwa nilai kemiringan untuk data 3 dan data 5 bernilai 0 (**simetri terhadap rata-rata**). Pada Tabel

4.12 menyajikan tabel distribusi frekuensi untuk data 1 sampai dengan data 5, berdasarkan Tabel 4.11. Berdasarkan Tabel 4.12, untuk data 1, nilai 1 sebanyak 2, nilai 2 sebanyak 7, nilai 3 sebanyak 5, dan seterusnya. Pada data 1, diketahui rata-rata > median > modus (miring ke kanan). Sementara pada data 4, diketahui rata-rata < median < modus (miring ke kiri). Pada data 5, diketahui rata-rata = median = modus (simetri, kasus **unimodus** atau *unimodal*). Pada data 3, tidak bersifat unimodus (unimodus atau *unimodal* berarti jumlah modus dalam data sebanyak 1). Gambar 4.2 merupakan grafik untuk distribusi frekuensi pada Tabel 4.12.

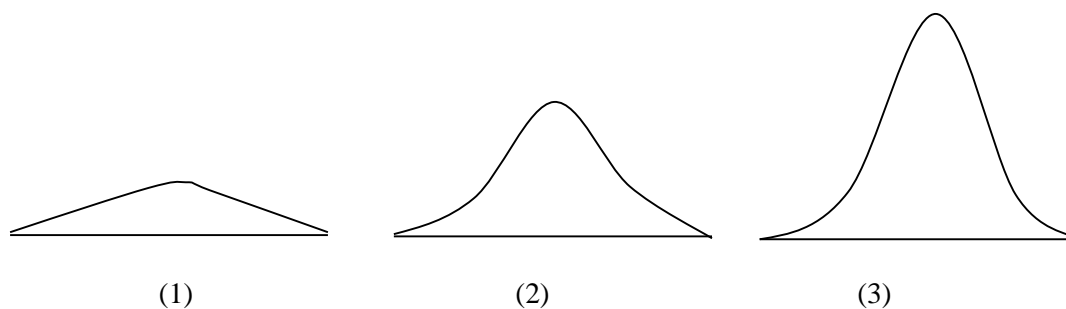
### *Ukuran Keruncingan (Kurtosis)*

Ukuran keruncingan atau *kurtosis* merupakan suatu nilai yang mengukur tingkat keruncingan atau ketinggian puncak dari distribusi data. Berikut rumus untuk menghitung kurtosis.

$$Kurtosis = \left\{ \frac{(n)(n+1) \sum (X - \bar{X})^4}{(n-1)(n-2)(n-3)s^4} \right\} - \frac{3(n-1)^2}{(n-2)(n-3)}$$

Pada Gambar 4.3 (1), (2), dan (3), ketinggian atau keruncingan puncak kurva berbeda-beda. Gambar 4.3 (2) merupakan kurva normal atau mesokurtis (kurva tidak terlalu tajam dan datar). Pada Gambar 4.3 (1), kurva cenderung datar dan puncak tidak terlalu tinggi. Kurva ini dinamakan kurva platikurtis. Pada Gambar 4.3 (3), puncak kurva terlihat lancip dan tinggi. Kurva ini dinamakan kurva leptokurtis. Spiegel dan Stephens (2008:125) menyatakan sebagai berikut.

*“Kurtosis is the degree of peakedness of a distribution, usually taken relative to a normal distribution. A distribution having a relatively high peak is called leptokurtic, while one which is flat-topped is called platykurtic. A normal distribution, which is not very peaked or very flat-topped, is called mesokurtic.”*



**Gambar 4.3**

Berikut akan dihitung nilai kurtosis berdasarkan data pada Tabel 4.9. Berdasarkan data pada Tabel 4.10, diketahui nilai  $\bar{X} = 3,6$  dan  $s = 1,454058$ , sehingga

$$Kurtosis = \left\{ \frac{(n)(n+1) \sum (X - \bar{X})^4}{(n-1)(n-2)(n-3)s^4} \right\} - \frac{3(n-1)^2}{(n-2)(n-3)}$$

$$Kurtosis = \left\{ \frac{(15)(15+1)(133,568)}{(15-1)(15-2)(15-3)(1,454^3)} \right\} - \frac{3(15-1)^2}{(15-2)(15-3)}$$

$$Kurtosis = -0,485756$$

Nilai kurtosis berdasarkan perhitungan adalah  $-0,485756$ . Tabel untuk perhitungan disajikan pada Tabel 4.13.

**Tabel 4.13**

$X$	$f$	$fX$	$f \sum (X - \bar{X})^4$
1	1	1	45,6976
2	2	4	13,1072
3	5	15	0,648
4	3	12	0,0768
5	2	10	7,6832
6	2	12	66,3552
Jumlah	15	54	133,568

Tabel 4.14 menyajikan 3 data, yakni data 1 sampai data 3. Ketiga data tersebut masing-masing memiliki nilai data sebanyak 12. Nilai kurtosis untuk data 1 adalah -1,65, data 2 bernilai -0,85556, dan data 3 adalah 0,733333. Perhatikan bahwa semakin tinggi nilai kurtosis, maka puncak kurva semakin tinggi dan lancip (lihat Gambar 4.4).

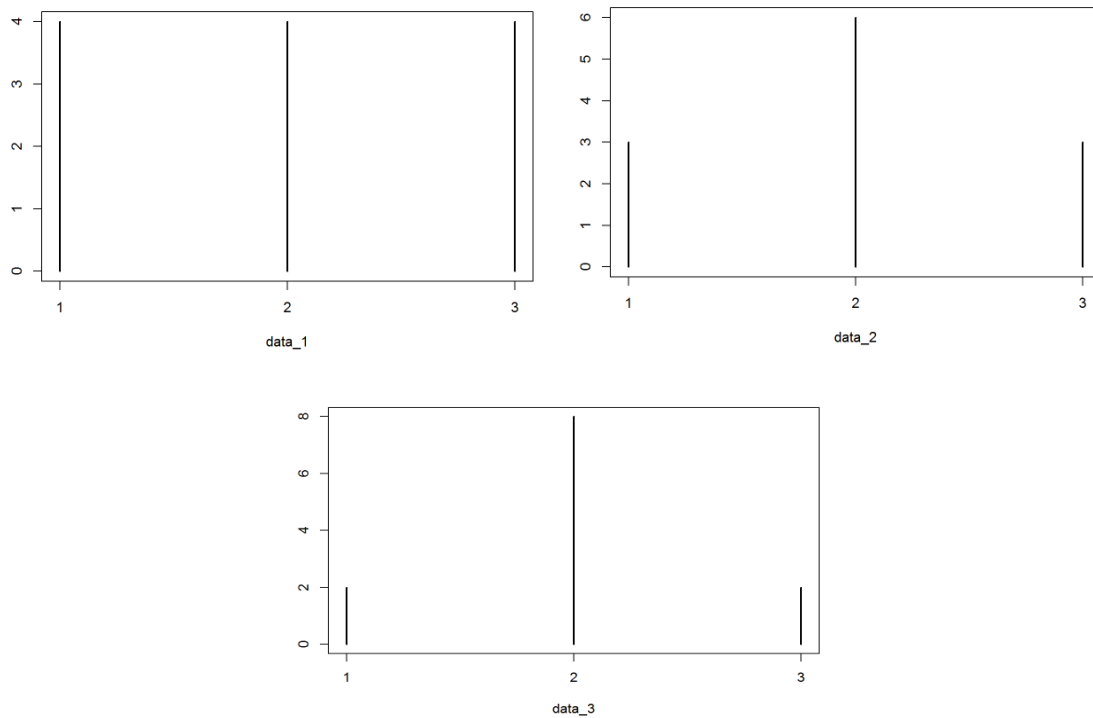
Pada Tabel 4.15 menyajikan tabel distribusi frekuensi untuk data 1 sampai dengan data 3 berdasarkan Tabel 4.14. Berdasarkan Tabel 4.15, untuk data 1, nilai 1 sebanyak 4, nilai 2 sebanyak 4, dan nilai 3 sebanyak 4. Untuk data 2, nilai 1 sebanyak 3, nilai 2 sebanyak 6, dan nilai 3 sebanyak 3. Gambar 4.4 merupakan grafik untuk distribusi frekuensi pada Tabel 4.15.

**Tabel 4.14**

No	data1	data2	data3
1	1	1	1
2	1	1	1
3	1	1	2
4	1	2	2
5	2	2	2
6	2	2	2
7	2	2	2
8	2	2	2
9	3	2	2
10	3	3	2
11	3	3	3
12	3	3	3
Kurtosis	-1,65	-0,85556	0,733333

**Tabel 4.15**

Nilai	Frekuensi		
	Data 1	Data 2	Data 3
1	4	4	4
2	3	6	3
3	2	8	2



Gambar 4.4

### Aplikasi dalam R

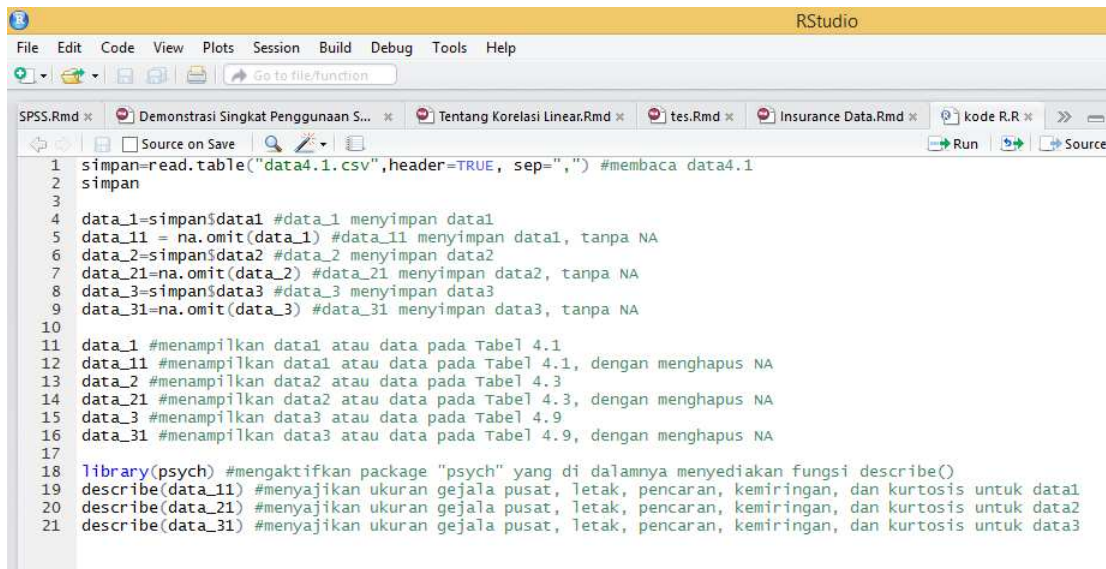
Data pada Tabel 4.1, Tabel 4.3, dan Tabel 4.9 disimpan terlebih dahulu dalam *Microsoft Excel*, dan disimpan dengan format **.CSV** (perhatikan Gambar 4.5).

	A	B	C	D
1	data1	data2	data3	
2	1	10	1	
3	2	10	2	
4	3	10	2	
5	4	20	3	
6	5	20	3	
7	6	30	3	
8	7	30	3	
9	8	30	3	
10	9	30	4	
11	10	30	4	
12	11	30	4	
13	12	40	5	
14	12	40	5	
15	13	40	6	
16	14	40	6	
17	15	50		
18	16	50		
19	17	50		
20	18	50		

Gambar 4.5

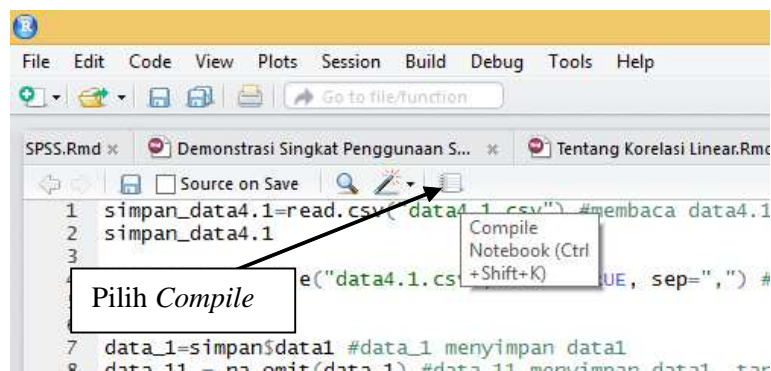
Berikut disajikan kode R (Gambar 4.6) untuk memperoleh hasil perhitungan ukuran gejala pusat, letak, pencaran, kemiringan, dan keruncingan, berdasarkan data pada Tabel 4.1 (data1, lihat Gambar 4.5), Tabel 4.3 (data2, lihat Gambar 4.5), dan Tabel 4.9 (data3, lihat Gambar 4.5).



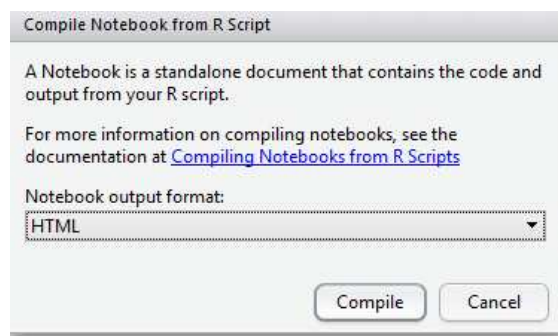


Gambar 4.6

Untuk mengeksekusi kode R pada Gambar 4.6, pilih *Compile* (Gambar 4.7). Pada Gambar 4.8, pilih format *output* HTML.



Gambar 4.7



Gambar 4.8

## Interpretasi Kode R

Berikut disajikan kode R, beserta *output* dari kode R tersebut.

```
[1] simpan=read.table("data4.1.csv",header=TRUE, sep=",") #membaca data4.1
    simpan
```

```
[2] data_1=simpan$data1 #data_1 menyimpan data1
data_11 = na.omit(data_1) #data_11 menyimpan data1, tanpa NA
data_2=simpan$data2 #data_2 menyimpan data2
data_21=na.omit(data_2) #data_21 menyimpan data2, tanpa NA
data_3=simpan$data3 #data_3 menyimpan data3
data_31=na.omit(data_3) #data_31 menyimpan data3, tanpa NA
data_1 #menampilkan data1 atau data pada Tabel 4.1
data_11 #menampilkan data1 atau data pada Tabel 4.1, dengan menghapus NA
data_2 #menampilkan data2 atau data pada Tabel 4.3
data_21 #menampilkan data2 atau data pada Tabel 4.3, dengan menghapus NA
data_3 #menampilkan data3 atau data pada Tabel 4.9
data_31 #menampilkan data3 atau data pada Tabel 4.9, dengan menghapus NA
```

```
[3] library(psych) #mengaktifkan package "psych" yang di dalamnya menyediakan
fungsi describe()
describe(data_11) #menyajikan ukuran gejala pusat, letak, pencaran, kemiringan, dan
kurtosis untuk data1
describe(data_21) #menyajikan ukuran gejala pusat, letak, pencaran, kemiringan, dan
kurtosis untuk data2
describe(data_31) #menyajikan ukuran gejala pusat, letak, pencaran, kemiringan, dan
kurtosis untuk data3
```

Gambar 4.9 merupakan hasil dari kode R [1]. Kode R [1] dapat diartikan variabel **simpan** ditugaskan untuk menyimpan data pada variabel **data1**, **data2**, dan **data3** dalam *file data4.1.csv*. Kemudian menampilkan data pada variabel **data1**, **data2**, dan **data3**. Perhatikan bahwa pada variabel **data1**, jumlah data sebanyak 19. Begitu juga pada variabel **data2**. Untuk variabel **data3**, jumlah data sebanyak 15, selebihnya adalah NA.

```
simpan=read.table("data4.1.csv",header=TRUE, sep=",") #membaca data4.1
simpan

##      data1 data2 data3
## 1         1    10     1
## 2         2    10     2
## 3         3    10     2
## 4         4    20     3
## 5         5    20     3
## 6         6    30     3
## 7         7    30     3
## 8         8    30     3
## 9         9    30     4
## 10        10   30     4
## 11        11   30     4
## 12        11   30     5
## 13        12   40     5
## 14        13   40     6
## 15        14   40     6
## 16        15   50    NA
## 17        16   50    NA
## 18        17   50    NA
## 19        18   50    NA
```

Gambar 4.9

Gambar 4.10 merupakan hasil dari kode R [2]. Sebagai contoh pada kode R **data\_1=simpan\$data1 #data\_1 menyimpan data1**, dapat diartikan variabel **data\_1** ditugaskan untuk menyimpan **data1** di dalam variabel **simpan**. Sehingga nilai dari **data\_1** adalah 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 11, 12, 13, 14, 15, 16, 17, 18. Pada kode R **data\_11 = na.omit(data\_1) #data\_11 menyimpan data1, tanpa NA** dapat diartikan variabel **data\_11** ditugaskan untuk menyimpan **data\_1**, dengan mengabaikan **NA**. Jadi nilai dari **data\_11** adalah 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 11, 12, 13, 14, 15, 16, 17, 18.

Kode R **data\_3=simpan\$data3 #data\_3 menyimpan data3**, dapat diartikan variabel **data\_3** ditugaskan untuk menyimpan **data3** di dalam variabel **simpan**. Sehingga nilai dari **data\_3** adalah 1, 2, 2, 3, 3, 3, 3, 3, 4, 4, 4, 5, 5, 6, 6, NA, NA, NA, NA. Pada kode R **data\_31 = na.omit(data\_3) #data\_31 menyimpan data3, tanpa NA** dapat diartikan variabel **data\_31** ditugaskan untuk menyimpan **data\_3**, dengan mengabaikan **NA**. Jadi nilai dari **data\_31** adalah 1, 2, 2, 3, 3, 3, 3, 3, 4, 4, 4, 5, 5, 6, 6. Kode R **data\_1** berarti menampilkan nilai pada variabel **data\_1**, kode R **data\_11** berarti menampilkan nilai pada variabel **data\_11**, dan seterusnya.

```
data_1 #menampilkan data1 atau data pada Tabel 4.1

## [1] 5 6 7 8 9 10 11 11 12 13 14 15 16 17 18 NA NA NA NA

data_11 #menampilkan data1 atau data pada Tabel 4.1, dengan menghapus NA

## [1] 5 6 7 8 9 10 11 11 12 13 14 15 16 17 18
## attr(,"na.action")
## [1] 16 17 18 19
## attr(,"class")
## [1] "omit"

data_2 #menampilkan data2 atau data pada Tabel 4.3

## [1] 10 10 10 20 20 30 30 30 30 30 30 30 40 40 40 50 50 50 50

data_21 #menampilkan data2 atau data pada Tabel 4.3, dengan menghapus NA

## [1] 10 10 10 20 20 30 30 30 30 30 30 30 40 40 40 50 50 50 50

data_3 #menampilkan data3 atau data pada Tabel 4.9

## [1] 1 2 2 3 3 3 3 3 4 4 4 5 5 6 6 NA NA NA NA

data_31 #menampilkan data3 atau data pada Tabel 4.9, dengan menghapus NA
```

**Gambar 4.10**

Gambar 4.11 merupakan hasil dari kode R [3]. Perhatikan kode R berikut.

```
library(psych)
describe(data_11)
```

Pada kode R tersebut, akan digunakan fungsi **describe**. Fungsi tersebut terdapat dalam *package* **psych**. Oleh karena itu, kode R **library(psych)** dapat diartikan mengaktifkan *package* **psych**. Setelah *package* **psych** diaktifkan, barulah fungsi **describe** dapat digunakan. Fungsi **describe** dalam hal ini digunakan untuk menentukan banyaknya data (**n**), rata-rata aritmatik (*mean*), standar deviasi (**sd**), median, minimum (**min**), maksimum (**max**), *range*, kemiringan (*skew*), dan kurtosis.

```

library(psych) #mengaktifkan package "psych" yang di dalamnya menyediakan fungsi describe()
describe(data_11) #menyajikan ukuran gejala pusat, letak, pencaran, kemiringan, dan kurtosis untuk d

## vars n mean sd median trimmed mad min max range skew kurtosis se
## 1 1 19 9.58 5.2 10 9.59 5.93 1 18 17 -0.04 -1.32 1.19

describe(data_21) #menyajikan ukuran gejala pusat, letak, pencaran, kemiringan, dan kurtosis untuk d

## vars n mean sd median trimmed mad min max range skew kurtosis
## 1 1 19 31.58 13.44 30 31.76 14.83 10 50 40 -0.14 -1.13
## se
## 1 3.08

describe(data_31) #menyajikan ukuran gejala pusat, letak, pencaran, kemiringan, dan kurtosis untuk d

## vars n mean sd median trimmed mad min max range skew kurtosis se
## 1 1 15 3.6 1.45 3 3.62 1.48 1 6 5 0.14 -1.01 0.38

```

**Gambar 4.11**

Pada hasil R Gambar 4.11, nilai kemiringan dihitung dengan rumus sebagai berikut.

$$Kemiringan = \frac{\sum(X - \bar{X})^3}{ns^3} - 3$$

Sementara, dalam *Microsoft excel*, nilai kemiringan dihitung dengan rumus sebagai berikut.

$$Kemiringan = \frac{n}{(n-1)(n-2)} \left( \frac{\sum(X - \bar{X})^3}{s^3} \right)$$

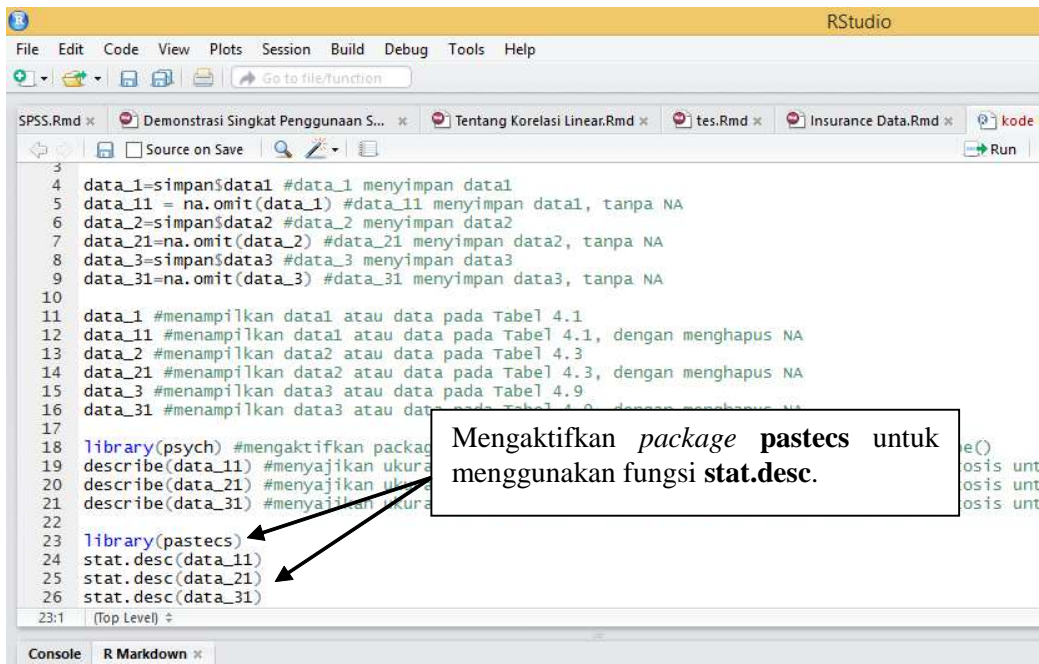
Pada hasil R Gambar 4.11, nilai kurtosis dihitung dengan rumus sebagai berikut.

$$Kurtosis = \frac{\sum(X - \bar{X})^4}{ns^4} - 3$$

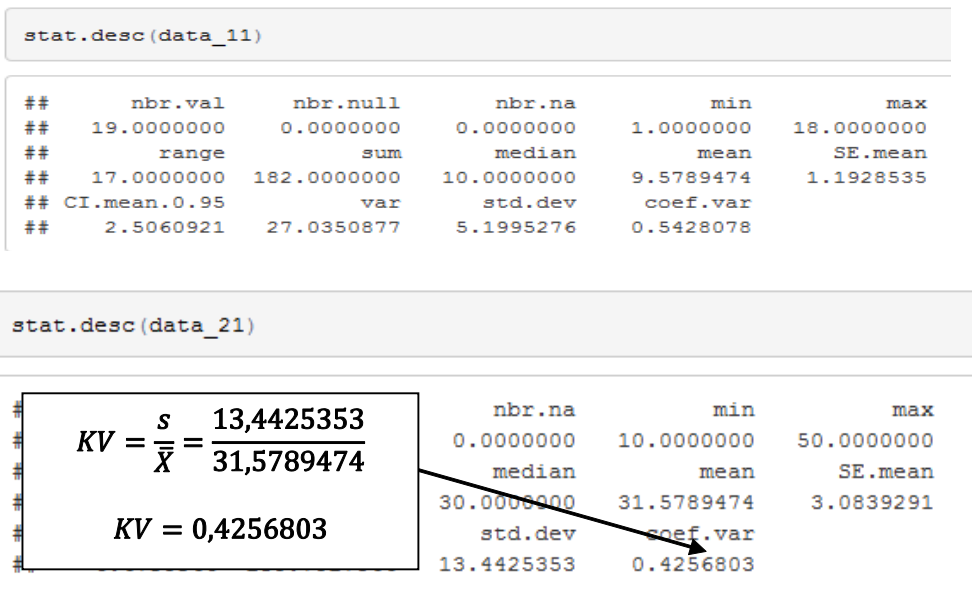
Sementara, dalam *Microsoft excel*, nilai kurtosis dihitung dengan rumus sebagai berikut.

$$Kurtosis = \left\{ \frac{(n)(n+1) \sum(X - \bar{X})^4}{(n-1)(n-2)(n-3)s^4} \right\} - \frac{3(n-1)^2}{(n-2)(n-3)}$$

Pada Gambar 4.12, mengaktifkan *package* **pastecs** untuk menggunakan fungsi **stat.desc**. Hasilnya diperlihatkan pada Gambar 4.13. Pada penggunaan fungsi **stat.desc**, menyajikan beberapa nilai statistik, seperti *variance*, jumlah keseluruhan (*sum*), *standard error mean* (*SE.mean*), dan koefisien variasi, yang **sebelumnya tidak tersaji** pada penggunaan fungsi **describe**.



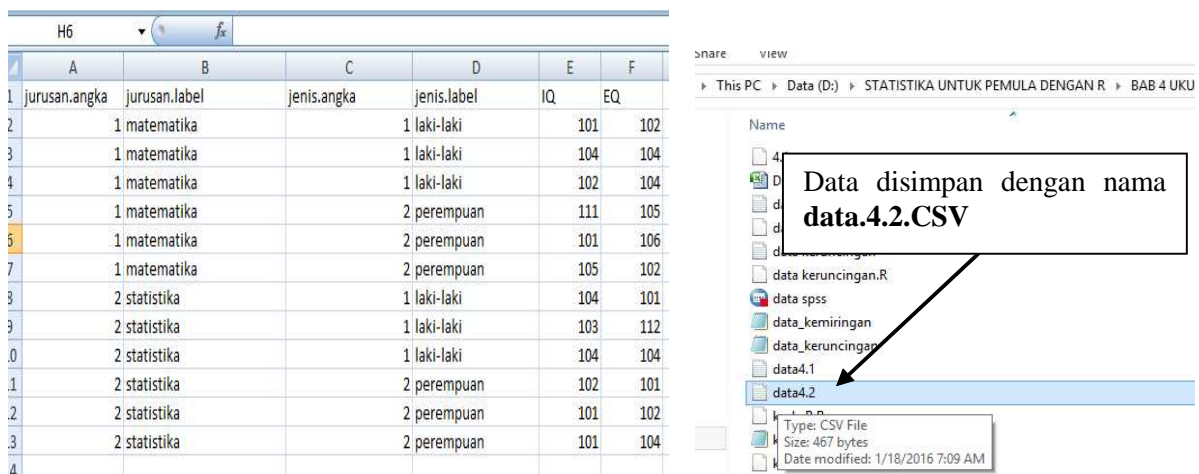
Gambar 4.12



Gambar 4.13

## Aplikasi dalam R (Data Berkelompok)

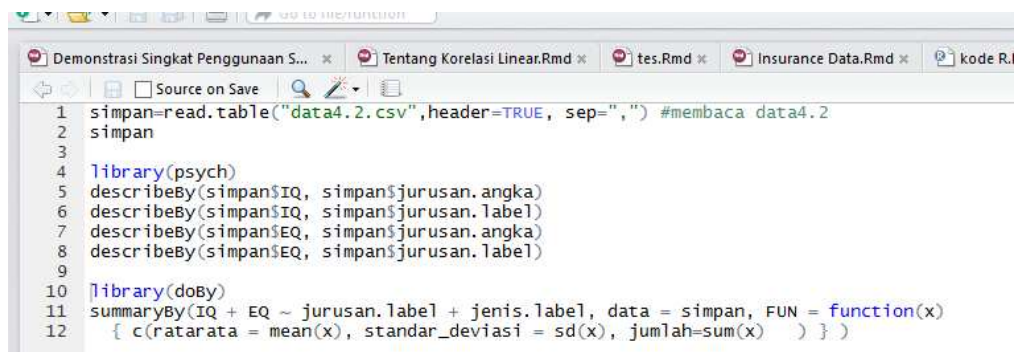
Andaikan diberikan data, seperti pada Gambar 4.14. Data tersebut disimpan dengan nama **data4.2.CSV** (perhatikan Gambar 4.14).



	A	B	C	D	E	F
1	jurusan.angka	jurusan.label	jenis.angka	jenis.label	IQ	EQ
2	1	matematika	1	laki-laki	101	102
3	1	matematika	1	laki-laki	104	104
4	1	matematika	1	laki-laki	102	104
5	1	matematika	2	perempuan	111	105
6	1	matematika	2	perempuan	101	106
7	1	matematika	2	perempuan	105	102
8	2	statistika	1	laki-laki	104	101
9	2	statistika	1	laki-laki	103	112
10	2	statistika	1	laki-laki	104	104
11	2	statistika	2	perempuan	102	101
12	2	statistika	2	perempuan	101	102
13	2	statistika	2	perempuan	101	104

Gambar 4.14

Berikut disajikan kode R (Gambar 4.15).



```
1 simpan=read.table("data4.2.csv",header=TRUE, sep=",") #membaca data4.2
2 simpan
3
4 library(psych)
5 describeBy(simpan$IQ, simpan$jurusan.angka)
6 describeBy(simpan$IQ, simpan$jurusan.label)
7 describeBy(simpan$EQ, simpan$jurusan.angka)
8 describeBy(simpan$EQ, simpan$jurusan.label)
9
10 library(doby)
11 summaryBy(IQ + EQ ~ jurusan.label + jenis.label, data = simpan, FUN = function(x)
12 { c(ratarata = mean(x), standar_deviasi = sd(x), jumlah=sum(x) ) } )
```

Gambar 4.15

Untuk mengeksekusi kode R pada Gambar 4.15, pilih *Compile* dan pilih format *output HTML*.

### Interpretasi Kode R

Berikut disajikan kode R, beserta *output* dari kode R tersebut.

```
[1] simpan=read.table("data4.2.csv",header=TRUE, sep=",") #membaca data4.2
```

**Simpan**

```
[2] library(psych)
describeBy(simpan$IQ, simpan$jurusan.angka)
describeBy(simpan$IQ, simpan$jurusan.label)
describeBy(simpan$EQ, simpan$jurusan.angka)
describeBy(simpan$EQ, simpan$jurusan.label)
```

[3] library(doBy)

```
summaryBy(IQ + EQ ~ jurusan.label + jenis.label, data = simpan, FUN = function(x)
{ c(ratarata = mean(x), standar_deviasi = sd(x), jumlah=sum(x) ) } )
```

Gambar 4.16 merupakan hasil dari kode R [1]. Kode R [1] dapat diartikan variabel **simpan** ditugaskan untuk menyimpan data pada variabel **jurusan.angka**, **jurusan.label**, **jenis.angka**, **jenis.label**, **IQ**, dan **EQ** dalam file **data4.2.csv**. Kemudian menampilkan data yang tersimpan pada variabel **simpan**.

```
simpan=read.table("data4.2.csv",header=TRUE, sep=",") #membaca data4.2
simpan

##   jurusan.angka jurusan.label jenis.angka jenis.label IQ EQ
## 1             1   matematika             1 laki-laki 101 102
## 2             1   matematika             1 laki-laki 104 104
## 3             1   matematika             1 laki-laki 102 104
## 4             1   matematika             2 perempuan 111 105
## 5             1   matematika             2 perempuan 101 106
## 6             1   matematika             2 perempuan 105 102
## 7             2   statistika             1 laki-laki 104 101
## 8             2   statistika             1 laki-laki 103 112
## 9             2   statistika             1 laki-laki 104 104
## 10            2   statistika             2 perempuan 102 101
## 11            2   statistika             2 perempuan 101 102
## 12            2   statistika             2 perempuan 101 104
```

Gambar 4.16

```
library(psych)
describeBy(simpan$IQ, simpan$jurusan.angka)

## group: 1
##   vars n mean  sd median trimmed mad min max range skew kurtosis se
## 1     1 6 104 3.79  103    104 2.97 101 111   10 0.86   -0.93 1.55
## -----
## group: 2
##   vars n mean  sd median trimmed mad min max range skew kurtosis se
## 1     1 6 102.5 1.38 102.5  102.5 2.22 101 104    3 0   -2.06 0.56

describeBy(simpan$IQ, simpan$jurusan.label)

## group: matematika
##   vars n mean  sd median trimmed mad min max range skew kurtosis se
## 1     1 6 104 3.79  103    104 2.97 101 111   10 0.86   -0.93 1.55
## -----
## group: statistika
##   vars n mean  sd median trimmed mad min max range skew kurtosis se
## 1     1 6 102.5 1.38 102.5  102.5 2.22 101 104    3 0   -2.06 0.56

describeBy(simpan$EQ, simpan$jurusan.angka)

## group: 1
##   vars n mean  sd median trimmed mad min max range skew kurtosis se
## 1     1 6 103.83 1.6   104  103.83 2.22 102 106    4 -0.02  -1.82 0.65
## -----
## group: 2
##   vars n mean  sd median trimmed mad min max range skew kurtosis se
## 1     1 6 104 4.15  103    104 2.22 101 112   11 1.05   -0.59 1.69
```

Gambar 4.17

Gambar 4.17 merupakan hasil dari kode R [2]. Kode R **library(psych)** dapat diartikan untuk mengaktifkan *package psych*. Pengaktifan *package psych* dimaksudkan untuk penggunaan

fungsi `describeBy`. Kode R `describeBy(simpan$IQ, simpan$jurusan.angka)` dapat diartikan akan disajikan nilai-nilai statistik, seperti rata-rata (*mean*), median, *range*, dan seterusnya, berdasarkan variabel **IQ** untuk kelompok pada variabel **jurusan.angka**. Kode R `describeBy(simpan$IQ, simpan$jurusan.label)` dapat diartikan akan disajikan nilai-nilai statistik, seperti rata-rata (*mean*), median, *range*, dan seterusnya, berdasarkan variabel **IQ** untuk kelompok pada variabel **jurusan.label**. Kode R `describeBy(simpan$EQ, simpan$jurusan.label)` dapat diartikan akan disajikan nilai-nilai statistik, seperti rata-rata (*mean*), median, *range*, dan seterusnya, berdasarkan variabel **EQ** untuk kelompok pada variabel **jurusan.label**.

Gambar 4.18 merupakan hasil dari kode R [3]. Kode R `library(doBy)` dapat diartikan untuk mengaktifkan *package doBy*. Pengaktifan *package doBy* dimaksudkan untuk penggunaan fungsi `summaryBy`. Kode R `summaryBy(IQ + EQ ~ jurusan.label + jenis.label, data = simpan, FUN = function(x) { c(ratarata = mean(x), standar_deviasi = sd(x), jumlah=sum(x) ) } )` dapat diartikan akan disajikan nilai-nilai statistik, seperti rata-rata (*mean*), median, *range*, dan seterusnya, berdasarkan variabel **IQ** dan **EQ**, untuk kombinasi kategori dari variabel **jurusan.label** dan **jenis.label**.

```
library(doBy)

## Loading required package: survival

summaryBy(IQ + EQ ~ jurusan.label + jenis.label, data = simpan, FUN = function(x)
  { c(ratarata = mean(x), standar_deviasi = sd(x), jumlah=sum(x) ) } )

## jurusan.label jenis.label IQ.ratarata IQ.standar_deviasi IQ.jumlah
## 1 matematika laki-laki 102.3333 1.5275252 307
## 2 matematika perempuan 105.6667 5.0332230 317
## 3 statistika laki-laki 103.6667 0.5773503 311
## 4 statistika perempuan 101.3333 0.5773503 304
## EQ.ratarata EQ.standar_deviasi EQ.jumlah
## 1 103.3333 1.154701 310
## 2 104.3333 2.081666 313
## 3 105.6667 5.686241 317
## 4 102.3333 1.527525 307
```

**Gambar 4.18**

## Referensi

1. Agresti, A. dan B. Finlay. 2009. *Statistical Methods for the Social Sciences, 4<sup>th</sup> Edition*. United States of America: Prentice Hall.
2. Field, A. 2009. *Discovering Statistics Using SPSS, 3<sup>rd</sup> Edition*. London: Sage.
3. Gio, P.U. dan E. Rosmaini, 2015. Belajar Olah Data dengan SPSS, Minitab, R, Microsoft Excel, EViews, LISREL, AMOS, dan SmartPLS. USUpress.
4. Johnson, R.A. dan G.K. Bhattacharyya. 2011. *Statistics, Principles and Methods, 6<sup>th</sup> Edition*. John Wiley and Sons, Inc.
5. Mann, P. S. dan C.J. Lacke. 2011. *Introductory Statistics, International Student Version, 7<sup>th</sup> Edition*. Asia: John Wiley & Sons, Inc.



6. Montgomery, D. C. dan G. C. Runger. 2011. *Applied Statistics and Probability for Engineers, 5<sup>th</sup> Edition*. United States of America: John Wiley & Sons, Inc.
7. Ott, R.L. dan M. Longnecker. 2001. *An Introduction to Statistical Methods and Data Analysis, 5<sup>th</sup> Edition*. United States of America: Duxbury.
8. Smidth, R. K. dan D. H. Sanders. 2000. *Statistics a First Course, 6<sup>th</sup> Edition*. United States of America: McGraw-Hill Companies.
9. Spiegel, M.R. dan L.J. Stephens. 2008. *Statistics, 4<sup>th</sup> Edition*. McGraw-Hill.
10. <https://cran.r-project.org/web/packages/psych/psych.pdf>
11. <https://cran.r-project.org/web/packages/pastecs/pastecs.pdf>
12. <https://cran.r-project.org/web/packages/doBy/doBy.pdf>