

1. INTRODUCTION

Statistics is the science of conducting studies to collect, organize, summarize, analyze, and draw conclusions from a group of numerical information.

A **variable** is a characteristic or attribute that can assume different values.

Data are the values (measurements or observations) that the variables can assume.

A **population** consists of all subjects (human or otherwise) that are being studied.

A **sample** is a group of subjects selected from a population.

Variables and Types of Data		
Qualitative characteristic or attribute	Quantitative Numerical data	
gender (male or female)	Discrete Specified isolated values	Continuous all possible values within a specific range
	number of students in each class	temperature height weight

2. FREQUENCY DISTRIBUTION

Raw data is the collection of all the observation.

Frequency is the number of occurrence of a certain observation.

Cumulative frequency is the sum of all the frequencies up to and including that class.

Math scores of 50 students.

42	72	42	100	48	66	54	78	48	54
60	78	90	60	48	60	24	48	36	48
42	48	66	78	48	54	100	96	30	36
30	54	99	48	30	30	66	78	24	98
42	42	99	100	80	70	99	100	95	76

A. Simple Frequency Distribution

Scores	Frequency (<i>f</i>)	Cumulative Frequency (<i>cf</i>)
24		
30		
36		
42		
48		
54		
60		
66		
70		
72		
76		
78		
80		
90		
95		
96		
98		
99		
100		
Total		

B. Grouped Frequency Distribution

Observations can be grouped into **classes**.

Scores	Frequency (<i>f</i>)	Cumulative Frequency (<i>cf</i>)
21-30		
31-40		
41-50		
51-60		
61-70		
71-80		
81-90		
91-100		
Total		

Class is each categories created as a result of grouping.

Class Limits are two ends class values.

(The smaller one is *lower class limit*, the greater is *upper class limit*). 51 → lower class limit 60 → upper class limit

Class Boundaries

51-60	$d = \frac{61-60}{2} = 0.5$	51-0.5 – 60+0.5	50.5 – 60.5
61-70		61-0.5 – 70+0.5	60.5 – 70.5

Mid-value is the middle value of class.

51-60	49.5 – 60.5	$\frac{51+60}{2} = \frac{50.5+60.5}{2} = 55.5$
61-70	60.5 – 70.5	$\frac{61+70}{2} = \frac{60.5+70.5}{2} = 65.5$

Width is the difference between upper and lower class boundaries.

51-60	49.5 – 60.5	60.5 – 50.5 = 10
-------	-------------	------------------

Fill the blanks in the following frequency distribution table.

Class (Scores)	<i>f</i>	Class Limits		Class Boundaries		Mid-Value	Class Width
		Lower	Upper	Lower	Upper		
21-30	2						
31-40	4						
41-50	6						
51-60	5						
61-70	1						
71-80	3						
81-90	8						
91-100	5						
Total							

Assignment:

1. Travel distance of 50 workers to their work.

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

Based on the data above, fill the following table.

Class (Scores)	f	Class Limits		Class Boundaries		Mid-Value	Class Width
		Lower	Upper	Lower	Upper		
1 – 3							
4 – 6							
7 – 9							
10 – 12							
13 – 15							
16 – 18							
Total							

2. Temperature values for 30 cities in Fahrenheit.

112	100	127	120	134	118	105	110	109	112
110	118	117	116	118	122	114	114	105	109
107	112	114	115	118	117	118	122	106	110

Based on the data above, fill the following table.

Class (Scores)	f	Class Limits		Class Boundaries		Mid-Value	Class Width
		Lower	Upper	Lower	Upper		
100–104							
105–109							
110–114							
115–119							
120–124							
125–129							
130–134							
Total							

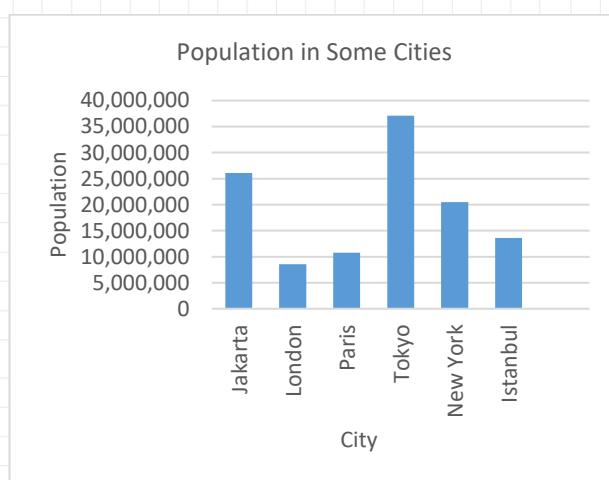
3. GRAPHICAL REPRESENTATION OF DATA

Data can be presented in graphic.

A. Bar Charts

The following table shows the population of some cities. Let's represent this data with bar charts.

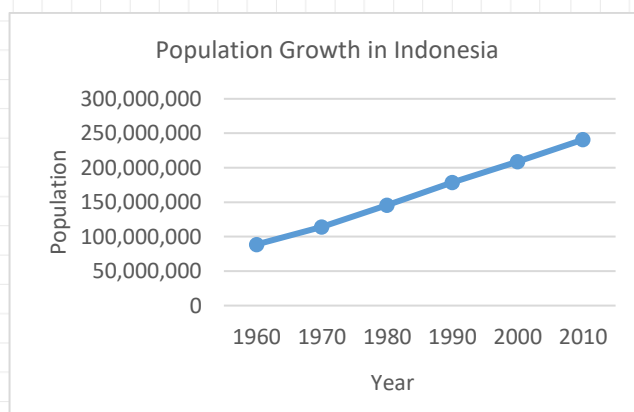
class	population
Jakarta	26,063,000
London	8,586,000
Paris	10,755,000
Tokyo	37,126,000
New York	20,464,000
Istanbul	13,576,000



B. Line Graphs

The following table is the population history of Indonesia.

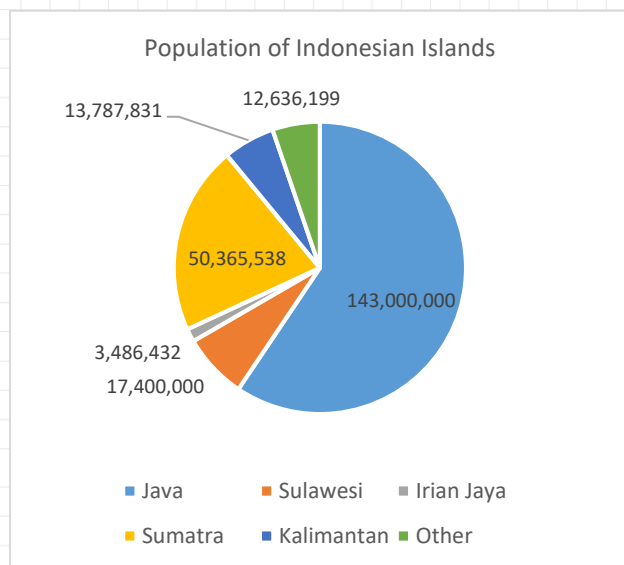
class	population
1960	88,693,000
1970	114,067,000
1980	145,494,000
1990	178,633,000
2000	208,939,000
2010	240,676,000



C. Pie Charts

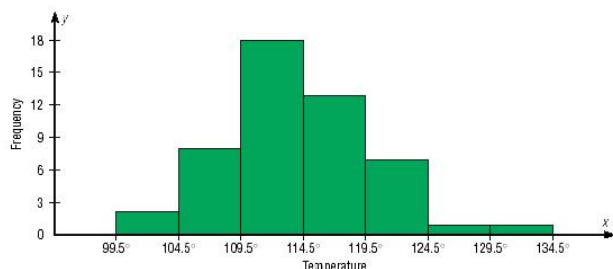
The following table is the population distribution of Indonesian islands.

class	population
Java	143,000,000
Sulawesi	17,400,000
Irian Jaya	3,486,432
Sumatra	50,365,538
Kalimantan	13,787,831
Other	12,636,199



D. Histogram

Class boundaries	frequency
99.5–104.5	2
104.5–109.5	8
109.5–114.5	18
114.5–119.5	13
119.5–124.5	7
124.5–129.5	1
129.5–134.5	1



4. MEASURE OF CENTRAL TENDENCY

A. Ungrouped Data

Mean (\bar{x}) is the sum of all values divided by the number of values.

Let x_1, x_2, \dots, x_n be n values, $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum x}{n}$

Example: Find the mean of 2, 14, 6, 10, 24, 13, 15.

Mean from frequency distribution table

Class	f
x_1	f_1
x_2	f_2
\vdots	\vdots
x_n	f_n
Total	

1st way: $\bar{x} = \frac{x_1 \cdot f_1 + x_2 \cdot f_2 + \dots + x_n \cdot f_n}{f_1 + f_2 + \dots + f_n} = \frac{\sum f \cdot x}{n}$

2nd way: $\bar{x} = \bar{x}_a + \frac{\sum f \cdot d}{n}$,

where $d = x - \bar{x}_a$ & \bar{x}_a are assumed mean chosen from the middle of frequency distribution.

Example: Find the mean age of the data given in the following frequency distribution table.

1st way:

x	f	$f \cdot x$
15	3	
17	2	
20	6	
27	1	
28	2	
29	2	
30	4	
Total	20	

2nd way:

x	f	$d = x - \bar{x}_a$	$f \cdot d$
15	3		
17	2		
20	6		
$\bar{x}_a = 27$	1		
28	2		
29	2		
30	4		
Total	20		

Let x_1, x_2, \dots, x_n be n ordered values.

Example: Find the median of the following data.

- 10, -4, 9, 4, 14, 11, 9, 18, -4, 5, 7
- 25, 32, 43, 54, 36, 18, 46, 24

Weight (kg)	65	67	68	70	72	74	75
No of person	7	4	3	4	1	3	2

x	f	cf
Total		

The third quartile (upper quartile) (Q_3) is the middle value between the median and the highest value of the data set.

$$Q_i = \left(\frac{i \cdot (n+1)}{4} \right)^{th} \text{ term}$$

10,11,12,	12	,15,17,18,	21	,21,21,22,	23	,23,25,32
$\left(\frac{1-(15+1)}{4} \right)^{10}$		$\left(\frac{2-(15+1)}{4} \right)^{21}$		$\left(\frac{3-(15+1)}{12} \right)^{23}$		

11 12 12 15 17 18 19 21 22 22 23 23 25 32

Q_1 median Q_3

$Q_3 =$

Weight (kg)	65	67	68	70	72	74	75
No of person	7	4	3	4	1	3	2

x	f	cf
Total		

$$D_i = \left(\frac{i \cdot (n+1)}{10} \right)^{th} \text{ term}$$
$$P_i = \left(\frac{i \cdot (n+1)}{100} \right)^{th} \text{ term}$$

Example: Find the 5th decile and 90th percentile in the following frequency distribution table.

x	f	cf
12	8	8
17	4	12
20	6	18
27	10	28
28	13	41
29	9	50
30	21	71
34	19	90
38	10	100
Total		

Mode is the value of data with maximum frequency.

x	f
12	2
17	6
20	3
27	5
28	1
29	4
30	2
Mode of the data above is 17.	

x	f
12	2
17	6
20	3
27	5
28	1
29	4
30	6
Mode of the data above is 17 and 30.	

x	f
12	2
17	2
20	2
27	2
28	2
29	2
30	2
Data above has no mode.	

Example: Find the mode of frequency distribution table given in previous example.

B. Grouped Data

Mean (\bar{x})

All values are assumed to be represented by mid-value of their class interval.

Work on the following data.

1st way:
$$\bar{x} = \frac{\sum f \cdot x}{n}$$

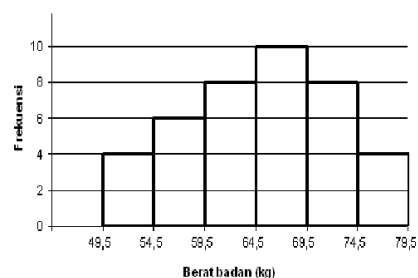
class	f	mid-value	$f \cdot x$
1-10	4		
11-20	3		
21-30	5		
31-40	3		
41-50	6		
Total	21		

2nd way:
$$\bar{x} = \bar{x}_a + \frac{\sum f \cdot d}{n}$$

class	f	mid-value	$d = x - \bar{x}_a$	$f \cdot d$
1-10	4			
11-20	3			
21-30	5			
31-40	3			
41-50	6			
Total	21			

Example: (UN 2007)

Perhatikan gambar berikut !



Berat badan siswa pada suatu kelas disajikan dengan histogram seperti pada gambar. Rataan berat badan tersebut adalah ... kg.

The data of students' weight in a class is shown by the histogram above. The average of weights is ...kg

Median

In the following data there are 21 values.

So, median is $\frac{21}{2} = 10,5^{th}$ value. It is in the class shown.

class boundaries	f	cf
0.5 – 10.5	4	4
10.5 – 20.5	3	7
20.5 – 30.5	5	12
30.5 – 40.5	3	15
40.5 – 50.5	6	21
Total	21	

$x_1, \dots, x_7, x_8, \dots, x_{10,5}, \dots, x_{12}, x_{13}, \dots, x_{21}$

$$x_{10,5} = 20,5 + \frac{10,5 - 7}{12 - 7} \cdot (30,5 - 20,5) = 27,5$$

median =

$$\left(\text{Lower boundary of median class} \right) + \frac{\frac{n}{2} - \left(\text{cum. frequency of previous class} \right)}{\left(\text{frequency of median class} \right)} \cdot \left(\text{width of median class} \right)$$

Example: (UN 2010 PAKET A)

Perhatikan tabel berikut!

Data	Frekuensi
10 – 19	2
20 – 29	8
30 – 39	12
40 – 49	7
50 – 59	3

Median dari data pada tabel adalah ...

Median of data above is ...

Quartiles

$Q_i =$

$$\left(\text{Lower boundary of quartile class} \right) + \frac{\frac{i \cdot n}{4} - \left(\text{cum. frequency of previous class} \right)}{\left(\text{frequency of quartile class} \right)} \cdot \left(\text{width of quartile class} \right)$$

class boundaries	f	cf
0.5 – 10.5	4	4
10.5 – 20.5	3	7
20.5 – 30.5	5	12
30.5 – 40.5	3	15
40.5 – 50.5	6	21
Total	21	

$$\frac{i \cdot n}{4} = \frac{1 \cdot 21}{4} = 5,25$$

$$Q_1 = 10,5 + \frac{\frac{1 \cdot 21}{4} - 4}{3} \cdot 10 = 14,67$$

$Q_2 =$

$Q_3 =$

Deciles

$$D_i = \left(\text{Lower boundary of decile class} \right) + \frac{\frac{i \cdot n}{10} - \left(\text{cum. frequency of previous class} \right)}{\left(\text{frequency of decile class} \right)} \cdot \left(\text{width of decile class} \right)$$

Percentiles

$P_i =$

$$\left(\text{Lower boundary of percentile class} \right) + \frac{\frac{i \cdot n}{100} - \left(\text{cum. frequency of previous class} \right)}{\left(\text{frequency of percentile class} \right)} \cdot \left(\text{width of percentile class} \right)$$

Example: Find the 4th decile and 70th percentile of the following frequency distribution.

class boundaries	f	cf
20.5 – 25.5	5	
25.5 – 30.5	8	
35.5 – 40.5	13	
40.5 – 45.5	9	
45.5 – 50.5	11	
Total	46	

Mode

For the grouped data with **equal class width** and only **one modal class**.

$$\text{Mode} = \left(\text{Lower boundary of modal class} \right) +$$

$$\frac{\left(\text{frequency of modal class} \right) - \left(\text{frequency of previous class} \right)}{2 \cdot \left(\text{frequency of modal class} \right) - \left(\text{frequency of previous class} \right) - \left(\text{frequency of following class} \right)} \cdot \left(\text{width of modal class} \right)$$

For the grouped data with **unequal class width** or with **more than one modal class**, use the following relation.

$$\text{Mean} - \text{Mode} = 3(\text{Mean} - \text{Median})$$

Example: Find the mode of the following data.

class	class boundaries	f
1 – 5		8
6 – 10		5
11 – 15		7
16 – 20		10
21 – 25		3
26 – 30		4
31 – 35	30,5 – 35,5	12
36 – 40		9
41 – 45		4
46 – 50		6

$$\text{mode} = 30,5 + \frac{12 - 4}{2 \cdot 12 - 4 - 9} \cdot 5 = 34,1$$

Example: (UN 2010 PAKET A)

Perhatikan tabel berikut!

Berat Badan (kg)	Frekuensi
40 – 45	5
46 – 51	7
52 – 57	9
58 – 63	12
64 – 69	7

Modus dari data pada tabel tersebut adalah ...

Mode of the data above is ...

Nilai	Frekuensi
50 – 54	2
55 – 59	4
60 – 64	8
65 – 69	16
70 – 74	10
75 – 79	2

Modus dari data pada tabel adalah ...

Mode of the data above is ...

Example: Find the mode of the following data.

class	class boundaries	f	cf	mid-value	f · x
1 – 10		8			
11 – 15		7			
16 – 22		10			
23 – 29		4			
30 – 42		12			
43 – 50		9			

Example: UN 2011 PAKET 46

Distribusi nilai ulangan matematika di kelas XIIA :

5. MEASURE OF DISPERSION

A. Range

Range = maximum value - minimum value

class	f
10	2
14	6
22	3
26	5
29	1
35	4
42	2
Range = 42 - 10 = 32	

<i>class</i>	<i>class boundaries</i>	<i>f</i>
120-129	119,5-129,5	3
130-139		1
140-149		2
150-159		5
160-169		4
170-179	169,5-179,5	7
Range = 179,5 - 119,5 = 60		

B. Average Deviation

Let x_1, x_2, \dots, x_n be n values, $Average\ Deviation = \frac{\sum |x_i - \bar{x}|}{n}$

Example: Find the average deviation of 5, 6, 5, 7, 8, 5, 11, 13.

C. Variance (σ^2) and Standard Deviation (σ)

Ungrouped Data

Let x_1, x_2, \dots, x_n be n values,

1st way:

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n} \quad \sigma^2 = \frac{\sum f_i \cdot (x_i - \bar{x})^2}{n} \text{ for table data}$$

2nd way:
$$\sigma^2 = \frac{\sum (x_i - \bar{x}_a)^2}{n} - \left(\frac{\sum (x_i - \bar{x}_a)}{n} \right)^2$$

\bar{x}_a : assumed mean

Example: Find the variance and standard deviation of 5, 6, 5, 7, 8, 5, 11, 13.

Example: Find the variance and standard deviation of the following data.

x	f				
5	3				
10	4				
15	6				
20	2				
25	5				
Total	20				

Grouped Data

Let x_1, x_2, \dots, x_n be mid-values of n classes,

1st way:
$$\sigma^2 = \frac{\sum f \cdot x_i^2}{\sum f} - \left(\frac{\sum f \cdot x_i}{\sum f} \right)^2$$

2nd way:
$$\sigma^2 = \frac{\sum f \cdot (x_i - \bar{x}_a)^2}{\sum f} - \left(\frac{\sum f \cdot (x_i - \bar{x}_a)}{\sum f} \right)^2$$

\bar{x}_a : assumed mean

Example: Find the variance and standard deviation of the following data.

class	f				
1 - 5	8				
6 - 10	5				
11 - 15	7				
16 - 20	10				
21 - 25	3				
26 - 30	4				
31 - 35	12				
36 - 40	9				

Review Test

1. Data: 5, 6, 9, 9, 7, 9, 5, 6, 9, 7, 7, 4.
What is the mean of the data?

A) 3 B) 4 C) 5 D) 6 E) 7

2. Data: 5, 6, 9, 9, 7, 9, 5, 6, 9, 7, 7, 4, 5, 8, 9.
What is the mean of the data?

A) 3 B) 5 C) 7 D) 9 E) 11

3. Data: 5, 6, 9, 9, 7, 9, 5, 6, 9, 7, 7, 4.
What is the first quartile (Q_1) of the data?

A) 5 B) 5,5 C) 7 D) 7,5 E) 8

4. Data: 5, 6, 9, 9, 7, 9, 5, 6, 9, 7, 7, 4.
What is the first quartile (Q_3) of the data?

A) 5 B) 5,5 C) 7 D) 7,5 E) 9

5. There are ten students in a group, mean of their marks is 7,5. If two students are out from the group, mean of eight students is 7. What is the mean of two students?

A) 9,5 B) 9 C) 8,5 D) 8 E) 7,5

6. Data: 5, 6, 9, 9, 7, 9, 5, 6, 9, 7, 7, 8.
What is the Semi Interquartile Range (SIR) of the data?

A) 3 B) 2 C) $\frac{3}{2}$ D) 1 E) $\frac{1}{2}$

7. What is the mean of the following data?

Height (cm)	f
130 - 134	5
135 - 139	2
140 - 144	3
145 - 149	3
150 - 154	7

- A) 143,25
B) 142,5
C) 140,5
D) 134,25
E) 133,25

8. According to data number in 7, what is the second quartile?
A) 142 B) 144,5 C) 145 D) 149 E) 149,5

9. (UN 2005) Berat badan dari 40 siswa dalam kg tercatat pada tabel di samping. Rataan berat badan tersebut adalah...
Average of the following data is ...

Berat (kg)	f
35 - 39	4
40 - 44	11
45 - 49	12
50 - 54	7
55 - 59	4
60 - 64	2

- A) 46,20 B) 47 C) 47,25 D) 47,50 E) 47,90

10. Pada ulangan matematika, diketahui nilai rata-rata kelas adalah 58. Jika rata-rata nilai matematika untuk siswa laki-

laki 64 dan rata-rata untuk siswa perempuan 56, maka perbandingan banyak siswa laki-laki dan perempuan adalah...

The average of Math scores is 58. If the average of the male students is 64 and the average of the female students is 56, then ratio of male students to female students is ...

- A) 1:6 B) 1:3 C) 2:3 D) 3:2 E) 3:4

11. (UN 2007 PAKET B) Perhatikan tabel berikut!

Nilai	Frekuensi
20 - 24	2
25 - 29	8
30 - 34	10
35 - 39	16
40 - 44	12
45 - 49	8
50 - 54	4

Median dari data yang disajikan berikut adalah...

Median of data above is ...

- A) 32
B) 37,625
C) 38,25
D) 43,25
E) 44,50

12. (UN 2012/A13) Data yang diberikan dalam tabel frekuensi sebagai berikut:

Nilai	Frekuensi
20 - 29	3
30 - 39	7
40 - 49	8
50 - 59	12
60 - 69	9
70 - 79	6
80 - 89	5

Nilai modus dari data pada tabel adalah...

Mode of the data above is ...

- A) $49,5 - \frac{40}{7}$
B) $49,5 - \frac{36}{7}$
C) $49,5 + \frac{36}{7}$
D) $49,5 + \frac{40}{7}$
E) $49,5 + \frac{48}{7}$

13. (UN 2011 PAKET 12) Modus dari data pada table berikut adalah...

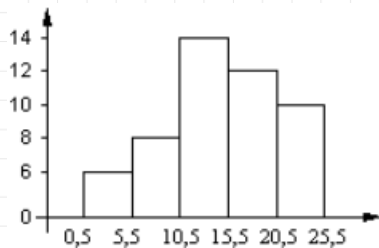
Mode of the following data is ...

Ukuran	Frekuensi
1 - 5	3
6 - 10	17
11 - 15	18
16 - 20	22
21 - 25	25
26 - 30	21
31 - 35	4

- A) $20,5 + \frac{3}{4} \cdot 5$
 B) $20,5 + \frac{3}{25} \cdot 5$
 C) $20,5 + \frac{3}{7} \cdot 5$
 D) $20,5 - \frac{3}{4} \cdot 5$
 E) $20,5 - \frac{3}{7} \cdot 5$

14. (UN 2004)

Modus dari data pada gambar adalah...
 Mode of the following data is ...



- A) 13,05 B) 13,50 C) 13,75 D) 14,05 E) 14,25

15. (UN 2009 PAKET A/B) Perhatikan table berikut!

Nilai	Frekuensi
40 - 49	7
50 - 59	6
60 - 99	10
70 - 79	8
80 - 89	9
Jumlah	40

Nilai kuartil atas (Q3) dari data yang disajikan adalah...
 Third quartile of data above is ...

- A) 54,50 B) 60,50 C) 78,25 D) 78,50 E) 78,75

16. (UN 2008 PAKET A/B) Perhatikan table berikut!

Nilai	Frekuensi
151 - 155	4
156 - 160	7
161 - 165	12
166 - 170	10
171 - 175	7

Nilai kuartil atas (Q3) dari data yang disajikan adalah...
 Third quartile of data above is ...

- A) 167 B) 167,5 C) 168 D) 168,5 E) 169

17. (UN 2010 PAKET B)

Data	Frekuensi
10 - 19	2
20 - 29	8
30 - 39	12
40 - 49	7
50 - 59	3

Median dari data pada tabel adalah...

Median of the data given in the table above is ...

- A) $34,5 + \frac{16-10}{12} \times 10$
 B) $34,5 + \frac{16-10}{12} \times 9$
 C) $29,5 + \frac{16-10}{12} \times 9$
 D) $29,5 + \frac{16-10}{12} \times 10$
 E) $38,5 + \frac{16-10}{12} \times 10$

18. (UN 2003) Perhatikan tabel berikut!

Nilai	Frekuensi
30 - 39	1
40 - 49	3
50 - 59	11
60 - 69	21
70 - 79	43
80 - 89	32
90 - 99	9

Kuartil bawah dari data yang tersaji pada tabel distribusi di atas adalah...

First quartile of data above is ...

- A) 66,9 B) 66,6 C) 66,2 D) 66,1 E) 66,0

19. The mean marks of 100 students is found to be 40.

Later on, it was discovered that a mark 53 was misread as 83. Now, the correct mean marks is...

- A) 38,7 B) 39 C) 39,4 D) 39,7 E) 40,7

20. The mean of the following distribution is...

Class	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70	70 - 80
Frequency	10	6	8	12	5	9

- A) 48,4 B) 48,6 C) 49,4 D) 49,6 E) 49,8