

# 6 - 1

## How to Analyze Data

### Average and Spread (1)

#### Instruction

Representative values.

The following table summarizes the results for the cultivation of two sets of tomatoes. Compare which record is heavier.

Set 1: Tomato Cultivation Record

Number	Weight	Number	Weight	Number	Weight	Number	Weight	Number	Weight
1	56	6	49	11	44	16	38	21	50
2	49	7	32	12	49	17	37	22	58
3	42	8	52	13	58	18	46	23	38
4	46	9	39	14	53	19	57	24	46
5	36	10	53	15	46	20	38	25	56

Set 2: Tomato Cultivation Record

Number	Weight	Number	Weight	Number	Weight	Number	Weight	Number	Weight
1	54	6	45	11	37	16	36	21	39
2	44	7	54	12	50	17	49	22	40
3	49	8	49	13	49	18	44	23	54
4	39	9	60	14	40	19	58	24	49
5	37	10	38	15	38	20	49		

- 1 Find the **mean** for the records of each set.  
The mean for Set 1. The mean for Set 2.

$$1168 \div 25 = 46.72$$

$$1101 \div 24 = 45.875$$

Average you learned is called mean.



The mean learned so far is called the mean value.

$$(\text{Mean value}) = (\text{Total of data values}) \div (\text{Number of data})$$

From point of mean values, Set 1 is heavier than Set 2.  
Is there any other way to compare?



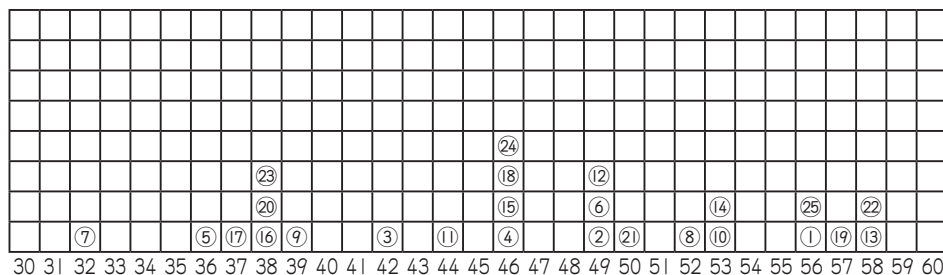
- 2 Which is the **maximum** score for each set? Also, find the **minimum** score.

	Maximum	Minimum
Set 1	58	32
Set 2	60	36

Set 1 has a larger mean value but has smaller maximum and minimum weight.



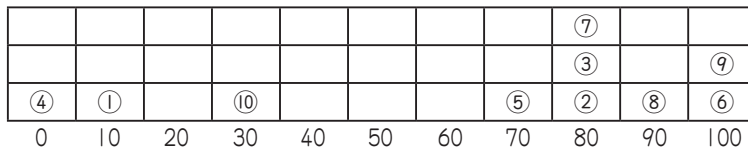
- 3 Think about the graph which shows the record of Set 1.



The graph shown on the previous page is called a **dot plot**. Vertically, you can see the amount of data. Horizontally, you can understand how the data is distributed. The number corresponds to the data numbers.

**Example**

The following plot shows the results of the language class exam. A girl got 70 scores and she compares the score with classmates. Answer the following questions



1 Find the mean

$$(0 + 10 + 30 + 70 + 3 \times 80 + 90 + 2 \times 100) \div 10 = 64$$

2 Which score is the most frequent value?

64

80

3 When the data is aligned in size order, which value is located in the middle?

$$(80 + 80) \div 2 = 80$$

80

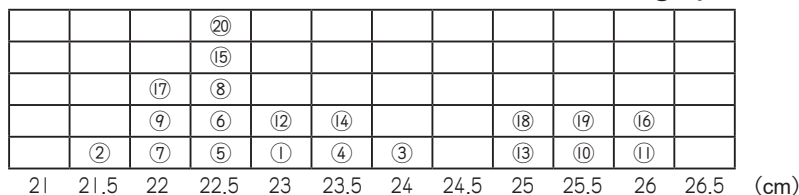
Find the middle value counting from top and bottom and calculate—the 5th value in this case.

Her score is higher than the mean value, but she is 7<sup>th</sup> from the top. The mean value is not appropriate when she compares the scores this time.



- Within the data, the most frequent value is called the **mode** value.
- When the data is aligned in size order, the value that is located in the middle is called the **median** value. When the number of data is odd, the value is in the exact middle. On the other hand, when the number of data is even, the mean value of the two values is in the middle.
- Values that represent data such as the mean, mode, median are called **representative values**. Depending on data use, it is chosen among the three values.

The following data plot shows the results of a survey on students' shoe size in a class. A shoemaker wants to know which size of shoes he should make to sell more shoes. Answer the following questions.



1 Find the mean, mode and median value

Mean  Mode  Median

2 Which size of shoes he should make to sell more shoes? Choose an appropriate representative value for the situation.

Not many students have the same shoe size as the mean.

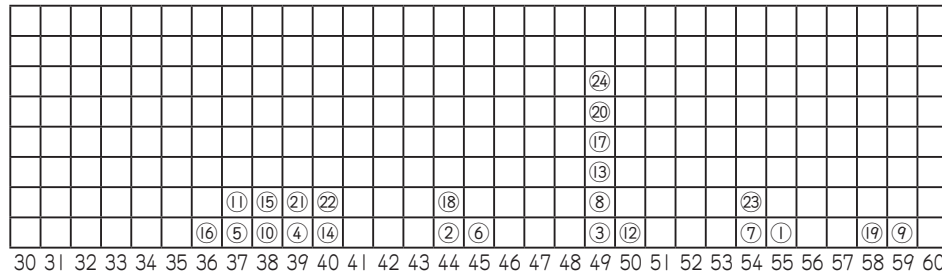


# 6 - 2

## How to Analyze Data

### Average and Spread (2)

**Example** The records of Set 2 on page 66 are summarized in the dot plot below. Answer the following questions.



- Complete the table.
- From what number to what number represents the largest number of tomatoes?

More than or equal to  and less than   
 or equal to  than

- How many tomatoes are more than or equal to 35 g and less than 40 g?

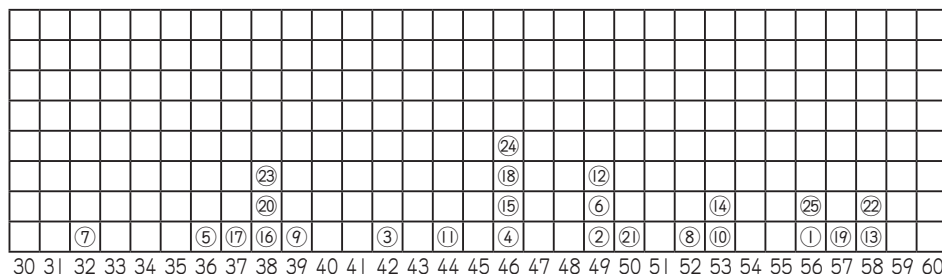
- Which class has 4 tomatoes?

More than or equal to  and less than

Weight (g)		Number of tomatoes
more than or equal to	30 ~ less than 35	0
	35 ~ 40	7
	40 ~ 45	4
	45 ~ 50	7
	50 ~ 55	3
	55 ~ 60	3
Total		24

- A section (delimiter) such as “more than or equal to 30 g and less than 35 g” is called a **class**, and the size of the section is called a **class interval**.
- The number of data counted for each class is called **frequency**. A table that shows the distribution by class or frequency is called a **frequency distribution table**.
- You can decide the class interval and use a frequency distribution table.

- The records of Set 1 on page 66 are summarized in the dot plot below. Answer the following questions.



1 Complete the table where there are blank spaces.

2 From what number to what number represents the largest number of tomatoes?

More than  or equal to  and less than

3 How many tomatoes are more than or equal to 35 g and less than 40 g?

4 Which class has 4 tomatoes?

More than or equal to  and less than

Weight (g)				Number of tomatoes	
more than or equal to	30	~	less than	35	
	35	~		40	
	40	~		45	
	45	~		50	
	50	~		55	
	55	~		60	
Total					

2 The following table summarizes the results of mathematics class exam for boys and girls. Compare which record is better.

Score of Mathematics Class Exam in Boy

Number	Score	Number	Score	Number	Score	Number	Score	Number	Score
1	92	3	88	5	78	7	70	9	88
2	84	4	100	6	94	8	86	10	98

Score of Mathematics Class Exam in Girl

Number	Score	Number	Score	Number	Score	Number	Score	Number	Score
1	84	3	72	5	94	7	84	9	90
2	100	4	100	6	92	8	78		

Score				Number of students	
more than or equal to	70	~	less than	75	
	75	~		80	
	80	~		85	
	85	~		90	
	90	~		95	
	95	~			
Total					

Score				Number of students	
more than or equal to	70	~	less than	75	
	75	~		80	
	80	~		85	
	85	~		90	
	90	~		95	
	95	~			
Total					

# 6 - 3

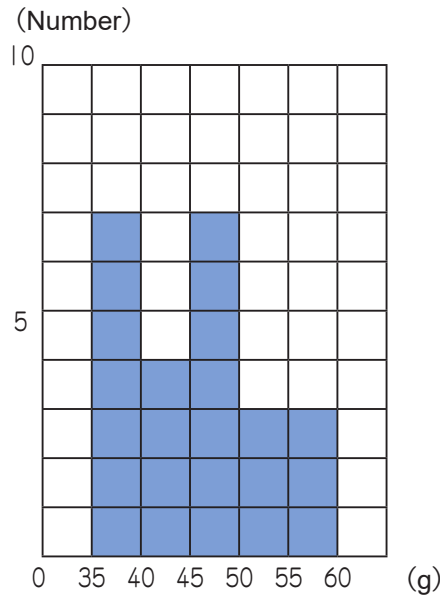
## How to Analyze Data

### Histograms (1)

#### Example

The graph below was drawn based on the distribution of records of Set 2 on page 66 and its corresponding frequency distribution table. Answer the following questions.

Weight (g)	Number of tomatoes
more than or equal to 0 ~ less than 35	0
35 ~ 40	7
40 ~ 45	4
45 ~ 50	7
50 ~ 55	3
55 ~ 60	3
<b>Total</b>	<b>24</b>



1 How many tomatoes are heavier than or equal to 50 g and less than 55 g?

3

2 Which class has 4 tomatoes?

More than or equal to

40 g

and less than

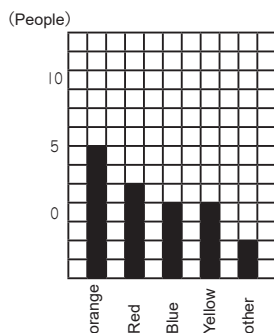
45 g

- A graph like the one above is called a **histogram**. If you look at the histogram it is easy to understand the distribution.
- In a histogram, the horizontal axis represents the class interval and the vertical axis represents how many tomatoes were produced by that class.

#### Instruction

Difference with a bar graph.

Bar graph and histogram use bars. What differences are there?



There is no gap between bars on histogram.



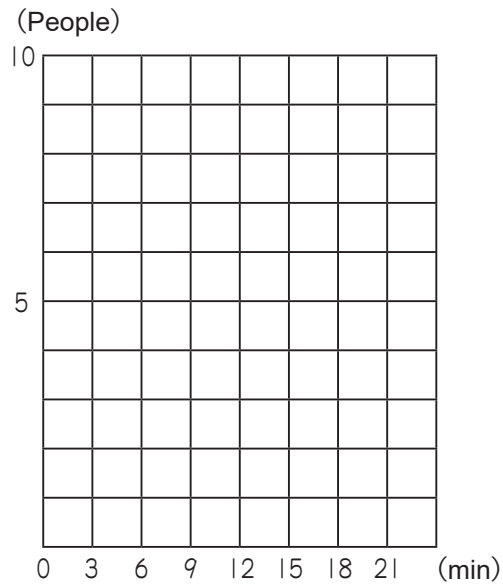
This is because it shows how many data there are within a certain interval.



The horizontal axis of a histogram is always a numerical value.

**1** The table below shows commuting time of classmates. Answer the following questions.

Time (min)		Number of student
more than or equal to	0 ~ less than 3	2
	3 ~ 6	6
	6 ~ 9	9
	9 ~ 12	7
	12 ~ 15	3
	15 ~ 18	1
<b>Total</b>		<b>28</b>



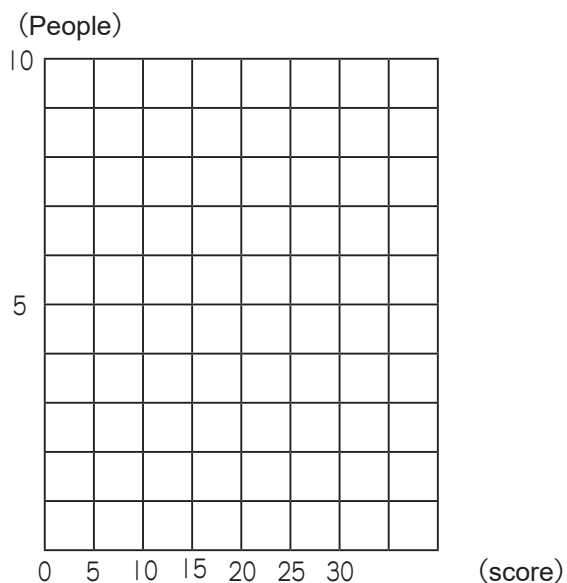
- 1 Complete histogram.
- 2 Which time frame represents the most students?

More than or equal to  and less than

- 3 How many students take more than or equal to 12 minutes?

**2** The following table summarizes the results of a language class test for girls. Complete the histogram.

Time (score)		Number of student
more than or equal to	0 ~ less than 5	1
	5 ~ 10	4
	10 ~ 15	2
	15 ~ 20	1
	20 ~ 25	0
	25 ~ 30	1
<b>Total</b>		<b>10</b>



# 6 - 4

## How to Analyze Data

### Histograms (2)

**Example** The following frequency distribution tables below were from the same records at throwing a softball of a class.

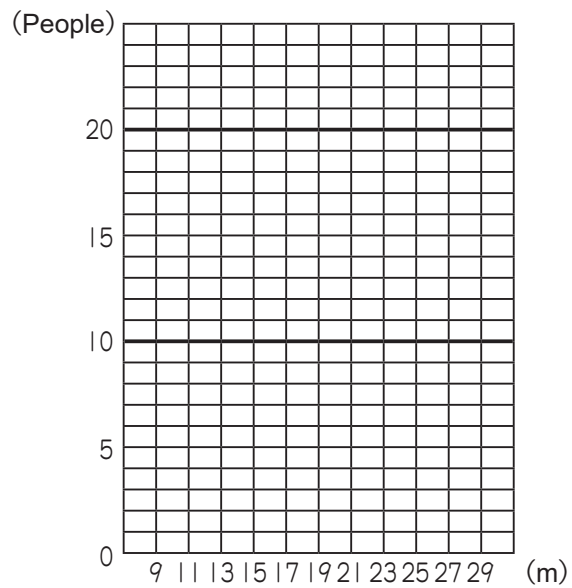
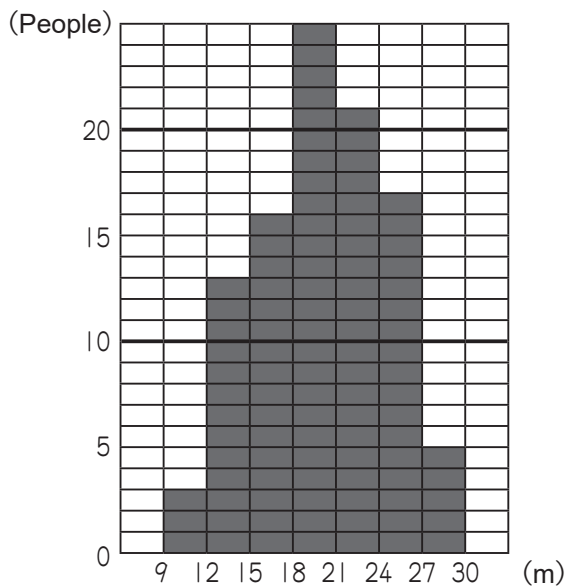
interval is 3 m

Distance (m)		Number of students
more than or equal to	9 ~ less than 12	3
	12 ~ 15	13
	15 ~ 18	16
	18 ~ 21	25
	21 ~ 24	21
	24 ~ 27	17
	27 ~ 30	5
Total		100

interval is 2 m

Distance (m)		Number of students
more than or equal to	9 ~ less than 11	1
	11 ~ 13	5
	13 ~ 15	10
	15 ~ 17	16
	17 ~ 19	19
	19 ~ 21	6
	21 ~ 23	11
	23 ~ 25	17
	25 ~ 27	10
	27 ~ 29	5
Total		100

Complete the histogram on the left.



Answer the questions on the data above.

- 1 Complete the histogram for 2 m interval on the right.
- 2 How many mountain shapes are there?

3 m interval

2 m interval

Even using the same data, set of a class makes a shape of graph different. For example, while you can observe the most people throw a ball from 19 m to 20 m on the left graph, we can not say so on the right graph.



# 6 - 5

## How to Analyze Data

### Various Graphs

#### Example

The graph on the right shows the relative populations of men and women in a country by age group in 2005

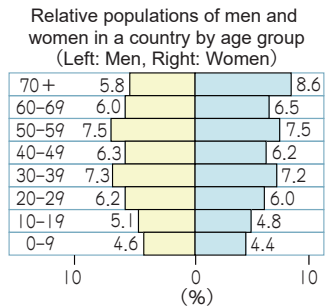
1 Which group has the highest number of people? Answer the question for both men and women.

Men

50-59

Women

70+



2 What % of the total population is younger than age of 20?

Math sentence

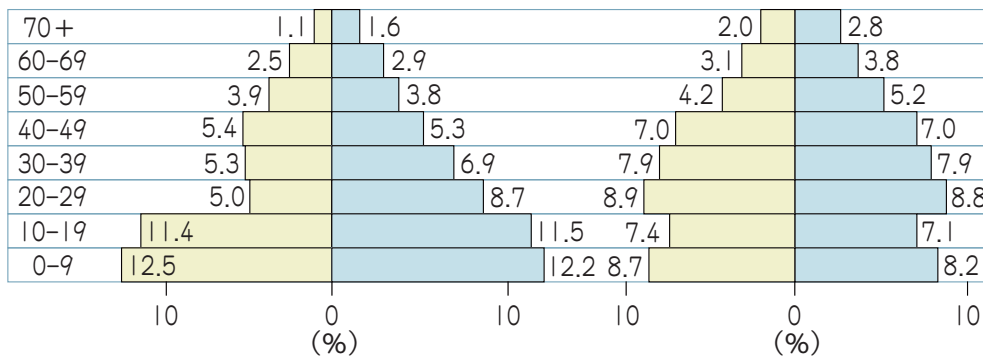
$$5.1 + 4.6 + 4.8 + 4.4 = 18.9$$

Answer

18.9 %

1 The graph below shows the relative population of men and women in a country by age group for 1945 and 1975.

Relative populations of men and women in a country by age group for 1945 (Left) and 1975 (Right)  
 Total population: 71990000 (Left: Men, Right: Women) Total population: 111940000



1 For each year, which group has the highest number of people? Answer the question for both men and women

	1945	1975
Men		
Women		

2 Choose appropriate observation.

A The proportion of the population younger than age of 20 doesn't change from 1945 to 1975.

B Comparing graphs in 1945 and 1975, the proportion of the population in the age of 0-9 and 10-19 turns small.





# 6 - 6

## How to Analyze Data

### Review

- 1 The following table summarizes the records of the number of books students in two classes borrowed. Compare which record is better.

Record of Class A Student's Borrowed Books

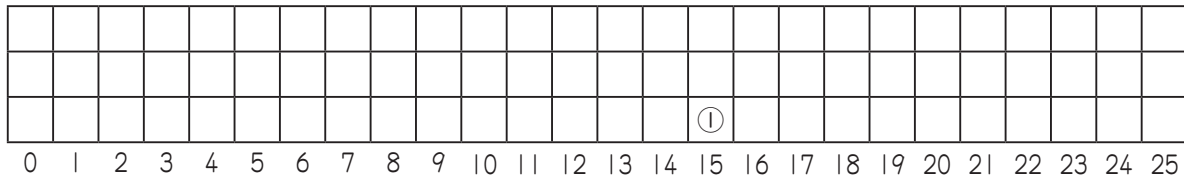
Number	Books	Number	Books	Number	Books	Number	Books
1	15	6	14	11	11	16	12
2	7	7	18	12	18	17	10
3	4	8	8	13	16	18	13
4	24	9	2	14	20	19	15
5	18	10	14	15	13	20	3

Record of Class B Student's Borrowed Books

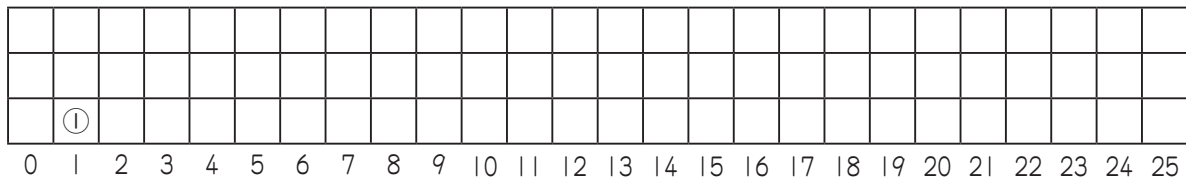
Number	Books	Number	Books	Number	Books	Number	Books
1	1	6	12	11	15	16	15
2	22	7	15	12	8	17	10
3	18	8	14	13	17	18	9
4	14	9	18	14	16	19	13
5	23	10	6	15	20		

- 1 Represent the records using data plot.

Class A



Class B



- 2 Complete the table below to find the mean, mode and median values of Class A and Class B.

	Mean	Mode	Median
Class A			
Class B			

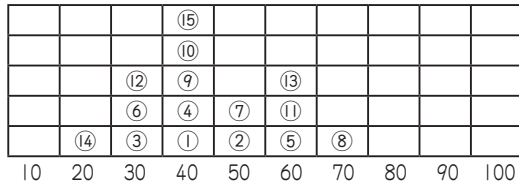
Finding mean value for Class A and Class B

Math  
sentence

Math  
sentence

**2** The table below shows the result of a survey on learning time for G5 and G6 students at home.

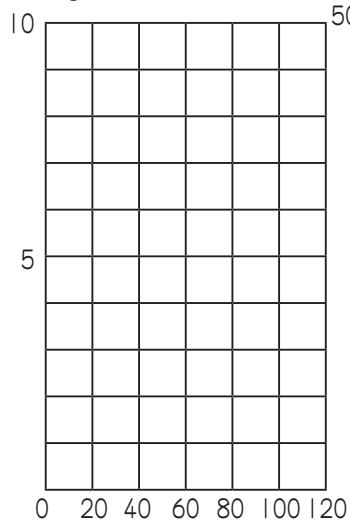
Grade 5



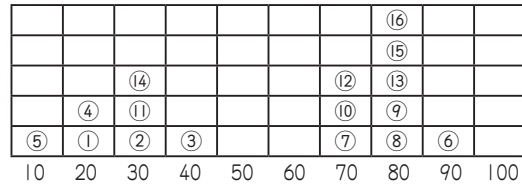
Grade 5 Students

Learning time (minutes)	Number of students
more than or equal to 0 ~ less than 20	0
20 ~ 40	4
40 ~ 60	7
60 ~ 80	4
80 ~ 100	0
<b>Total</b>	<b>15</b>

Learning time for Grade 5 Students



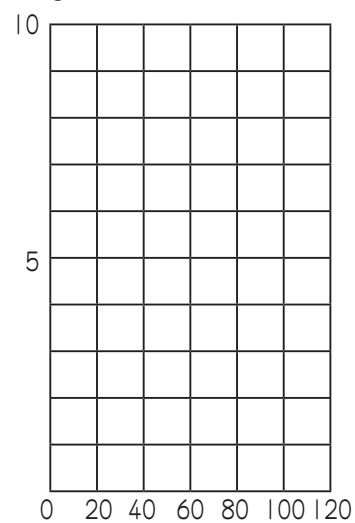
Grade 6



Grade 6 Students

Learning time (minutes)	Number of students
more than or equal to 0 ~ less than 20	1
20 ~ 40	5
40 ~ 60	1
60 ~ 80	3
80 ~ 100	6
<b>Total</b>	<b>16</b>

Learning time for Grade 6 Students



**1** Complete histogram above.

**2** Complete the table below to find the mean, mode and median values of Grade 5 and Grade 6.

	Mean	Mode	Median
Grade 5			
Grade 6			

What can you observe from the table on the left and histogram?



**3** Choose appropriate observation.

**A** Since mean, mode, and median values of Grade 5 students is lower than that of Grade 6 students, all the Grade 5 students have shorter learning time than that of Grade 6 students.

**B** While mean, mode, and median values of Grade 6 students is higher than that of Grade 5 students, about half of Grade 6 students have less learning time than the mean.