12 Statistics

CASIO
Essential Materials

Introduction

These teaching materials were created with the hope of conveying to many teachers and students the appeal of scientific calculators.

(1) Change awareness (emphasizing the thinking process) and boost efficiency in learning mathematics

- By reducing the time spent on manual calculations, we can have learning with a focus on the thinking process that is more efficient.
- This reduces the aversion to mathematics caused by complicated calculations, and allows students to experience the joy of thinking, which is the essence of mathematics.

(2) Diversification of learning materials and problem-solving methods

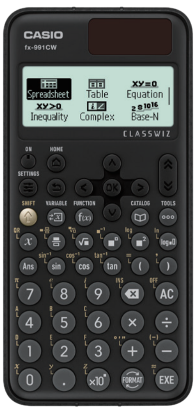
- Making it possible to do difficult calculations manually allows for diversity in learning materials and problem-solving methods.

(3) Promoting understanding of mathematical concepts

- By using the various functions of the scientific calculator in creative ways, students are able to deepen their understanding of mathematical concepts through calculations and discussions from different perspectives than before.
- This allows for exploratory learning through easy trial and error of questions.
- Listing and graphing of numerical values by means of tables allows students to discover laws and to understand visually.

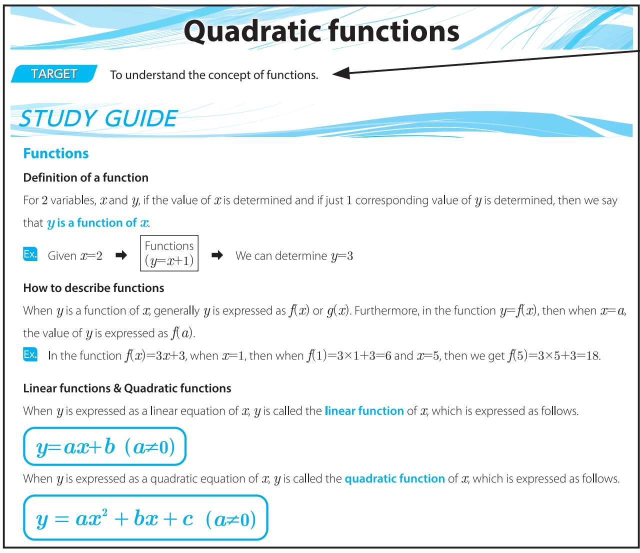
Features of this book

- As well as providing first-time scientific calculator users with opportunities to learn basic scientific calculator functions from the ground up, the book also has material to show people who already use scientific calculators the appeal of scientific calculators described above.
- You can also learn about functions and techniques that are not available on conventional Casio models or other brands of scientific calculators.
- This book covers many units of high school mathematics, allowing students to learn how to use the scientific calculator as they study each topic.
- This book can be used in a variety of situations, from classroom activities to independent study and homework by students.



Better Mathematics Learning with Scientific Calculator

Structure



Quadratic functions

TARGET To understand the concept of functions.

STUDY GUIDE

Functions

Definition of a function

For 2 variables, x and y, if the value of x is determined and if just 1 corresponding value of y is determined, then we say that **y is a function of x.**

Ex. Given $x=2$ ➡ Functions ($y=x+1$) ➡ We can determine $y=3$

How to describe functions

When y is a function of x, generally y is expressed as $f(x)$ or $g(x)$. Furthermore, in the function $y=f(x)$, then when $x=a$, the value of y is expressed as $f(a)$.

Ex. In the function $f(x)=3x+3$, when $x=1$, then when $f(1)=3\times1+3=6$ and $x=5$, then we get $f(5)=3\times5+3=18$.

Linear functions & Quadratic functions

When y is expressed as a linear equation of x, y is called the **linear function** of x, which is expressed as follows.

$$y=ax+b\ (a\neq0)$$

When y is expressed as a quadratic equation of x, y is called the **quadratic function** of x, which is expressed as follows.

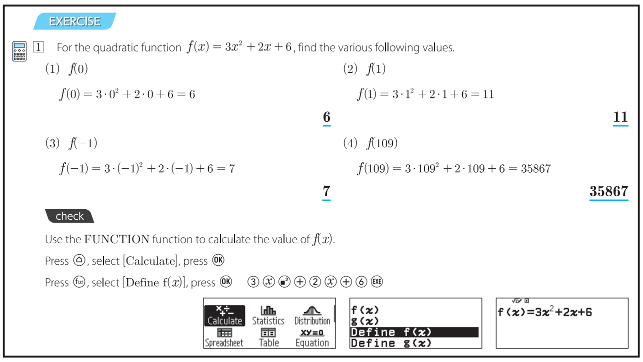
$$y=ax^2+bx+c\ (a\neq0)$$

TARGET

Students can identify the objective to learn in each section.

STUDY GUIDE

Mathematical theorems and concepts are explained in detail. A scientific calculator is used to check and derive formulas according to the topic.



EXERCISE

1 For the quadratic function $f(x)=3x^2+2x+6$, find the various following values.

(1) $f(0)$

$f(0)=3\cdot0^2+2\cdot0+6=6$

6

(2) $f(1)$

$f(1)=3\cdot1^2+2\cdot1+6=11$

11

(3) $f(-1)$

$f(-1)=3\cdot(-1)^2+2\cdot(-1)+6=7$

7

(4) $f(109)$

$f(109)=3\cdot109^2+2\cdot109+6=35867$

35867

check

Use the FUNCTION function to calculate the value of $f(x)$.

Press ⓐ, select [Calculate], press OK

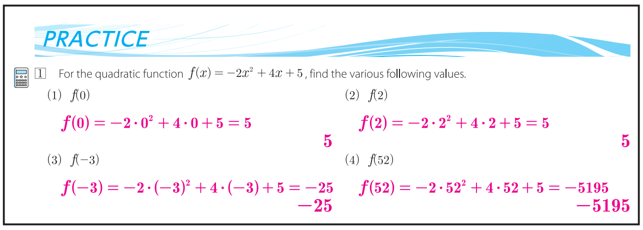
Press f(x), select [Define f(x)], press OK ③ ⓧ ⓧ² ⊕ ② ⓧ ⊕ ⑥ EXE

EXERCISE

Students learn basic examples based on the explanation in Study Guide.

check

Explains how to use the scientific calculator to solve problems and check answers.



PRACTICE

1 For the quadratic function $f(x)=-2x^2+4x+5$, find the various following values.

(1) $f(0)$

$f(0)=-2\cdot0^2+4\cdot0+5=5$

5

(2) $f(2)$

$f(2)=-2\cdot2^2+4\cdot2+5=5$

5

(3) $f(-3)$

$f(-3)=-2\cdot(-3)^2+4\cdot(-3)+5=-25$

-25

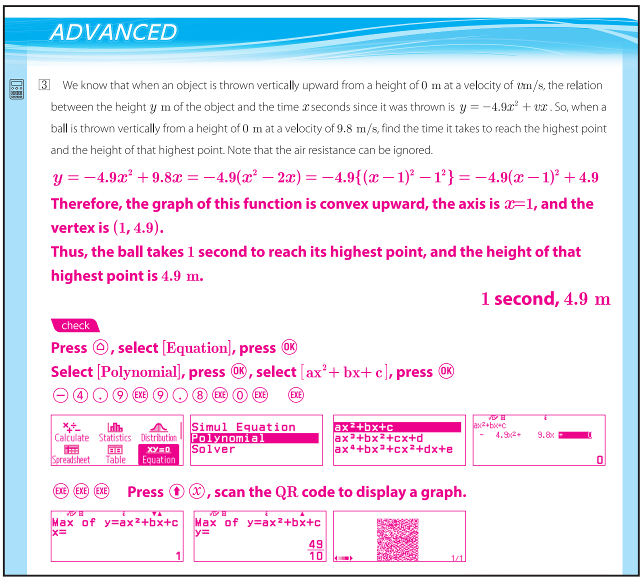
(4) $f(52)$

$f(52)=-2\cdot52^2+4\cdot52+5=-5195$

-5195

PRACTICE

Students can do practice problems similar to those in EXERCISE. They can also practice using the scientific calculator as they learned to in Check.



ADVANCED

3 We know that when an object is thrown vertically upward from a height of 0 m at a velocity of v m/s, the relation between the height y m of the object and the time x seconds since it was thrown is $y=-4.9x^2+vx$. So, when a ball is thrown vertically from a height of 0 m at a velocity of 9.8 m/s, find the time it takes to reach the highest point and the height of that highest point. Note that the air resistance can be ignored.

$y=-4.9x^2+9.8x=-4.9(x^2-2x)=-4.9\{(x-1)^2-1^2\}=-4.9(x-1)^2+4.9$

Therefore, the graph of this function is convex upward, the axis is $x=1$, and the vertex is (1, 4.9).

Thus, the ball takes 1 second to reach its highest point, and the height of that highest point is 4.9 m.

1 second, 4.9 m

check

Press ⓐ, select [Equation], press OK

Select [Polynomial], press OK, select [ax²+bx+c], press OK

⊖ ④ ⊙ ⑨ EXE ⑨ ⊙ ⑧ EXE ⓪ EXE EXE

EXE EXE EXE **Press ⇧ ⓧ, scan the QR code to display a graph.**

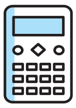
ADVANCED

Practical problems have been included in several topics. Solutions using a scientific calculator are also presented as necessary.

Other marks

Ex.	Simple examples on how to apply equations and theorems
explanation	Formulas and their supplementary explanations
proof	Proofs and checks of mathematical formulas
EXTRA Info.	Knowledge and information on formulas and other supplementary information in other units
OTHER METHODS	Alternative solutions and different verification methods for previously presented problems

Calculator mark



Where to use the scientific calculator

Colors of fonts in the teaching materials

- In STUDY GUIDE, important mathematical terms and formulas are printed in blue.
- In PRACTICE and ADVANCED the answers are printed in red.
  (Separate data is also available without the red parts, so it can be used for exercises.)

Applicable models

The applicable model is fx-991CW.
(Instructions on how to do input are for the fx-991CW, but in many cases similar calculations can be done on other models.)

Related Links

- Information and educational materials relevant to scientific calculators can be viewed on the following site.
  https://edu.casio.com
- The following video can be viewed to learn about the multiple functions of scientific calculators.
  https://www.youtube.com/playlist?list=PLRgxo9AwbIZLurUCZnrbr4cLfZdqY6aZA

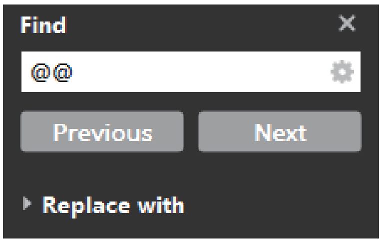
How to use PDF data

About types of data

- Data for all unit editions and data for each unit are available.
- For the above data, the PRACTICE and ADVANCED data without the answers in red is also available.

How to find where the scientific calculator is used

(1) Open a search window in the PDF Viewer.
(2) Type in "@@" as a search term.
(3) You can sequentially check where the calculator marks appear in the data.



How to search for a unit and section

(1) Search for units of data in all unit editions

- The data in all unit editions has a unit table of contents.
- Selecting a unit in the table of contents lets you jump to the first page of that unit.
- There is a bookmark on the first page of each unit, so you can jump from there also.

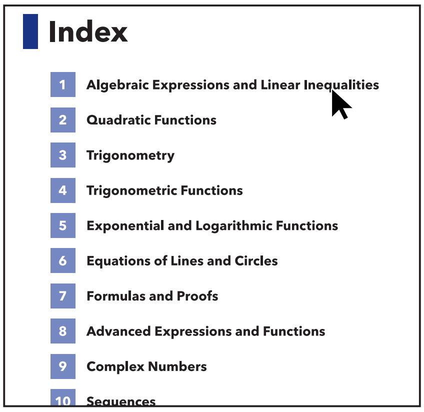
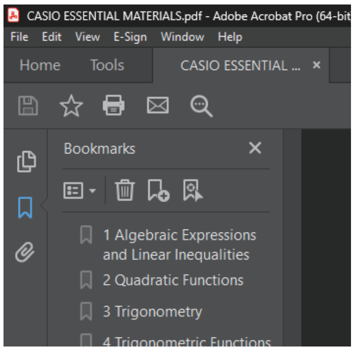


Table of contents of unit



Bookmark of unit

(2) Search for sections

- There are tables of contents for sections on the first page of units.
- Selecting a section in the table of contents takes you to the first page of that section.

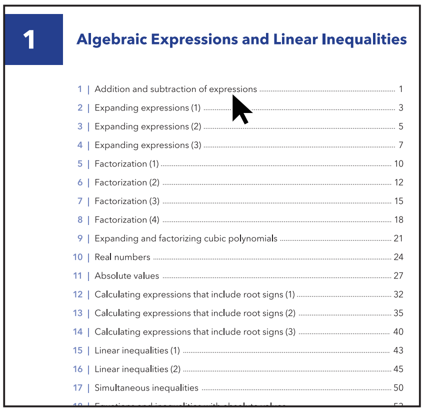


Table of contents of section

Class and frequency

TARGET To learn how to organize collected data into formats suitable for analysis.

STUDY GUIDE

Class, class width, frequency, frequency table, and cumulative frequency

Data is organized by separating it into a number of intervals.

Each interval is called a **class**, and the width of each interval is called **class width**.

The number of occurrences of the data in each class is called the **frequency**.

When data is arranged by class width and frequency in a chart, we call that chart a **frequency table**.

The sum of the frequencies from the first class to a certain class is called the **cumulative frequency**.

EXERCISE

◆ Table 1 below shows the records of a 50 meter race by a group of students. Solve the following problems for this data.

Table 1 Records of 50-m race

Number	Record (sec.)	Number	Record (sec.)	Number	Record (sec.)	Number	Record (sec.)
1	7.2	6	8.1	11	8.2	16	7.4
2	8.6	7	9.3	12	7.5	17	6.9
3	7.8	8	6.7	13	8.7	18	7.1
4	8.4	9	8.5	14	8.8	19	7.9
5	7.7	10	7.0	15	8.4	20	8.2

(1) Organize the records in table 1 into a frequency table as in table 2.

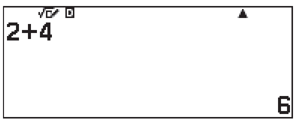
Find the cumulative frequencies.

Find the cumulative frequency of times under 7.5 seconds (add the frequencies of times greater than or equal to 6.0 seconds and under 6.5 seconds, those greater than or equal to 6.5 seconds and under 7.0 seconds, and those greater than or equal to 7.0 seconds and under 7.5).

Table 2 Records of 50-m race

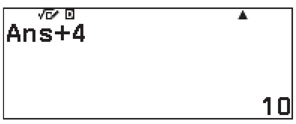
Class (sec.)			Frequency (people)	Cumulative frequency (people)
Greater than or equal to		Under		
6.0	-	6.5	**0**	**0**
6.5	-	7.0	**2**	**2**
7.0	-	7.5	**4**	**6**
7.5	-	8.0	**4**	**10**
8.0	-	8.5	**5**	**15**
8.5	-	9.0	**4**	**19**
9.0	-	9.5	**1**	**20**
9.5	-	10.0	**0**	**20**
Total			**20**	

② ⊕ ④ EXE



Add the frequency of times greater than or equal to 7.5 seconds and under 8.0 seconds to the above total.

⊕ ④ EXE



Add the following frequencies, in order, the same way.

⊕ ⑤ EXE ⊕ ④ EXE ⊕ ① EXE

(2) Find which class has the highest frequency.

Interpret the answer from the frequency table. The most frequent is 5 people who ran greater than or equal to 8.0 seconds and under 8.5 seconds.

Class that is greater than or equal to 8.0 seconds and under 8.5 seconds

(3) Find how many students have records under 8.0 seconds.

Interpret the cumulative frequency of those who ran greater than or equal to 7.5 seconds and under 8.0 seconds.

10 people

PRACTICE

1 Solve the following problems.

(1) When some data are organized and separated into several classes, what do you call the number of occurrences of the data in each of those classes?

Frequency

(2) What do you call a table in which the distribution of frequencies is organized for easy understanding?

Frequency table

(3) What do you call the sum of the frequencies from the first class to a certain class in the table in item (2)?

Cumulative frequency

2 Table 1 below shows the records of throwing a handball by a group of students. Solve the following problems for this data.

Table 1 Records of throwing a handball

Number	Record (m)	Number	Record (m)	Number	Record (m)	Number	Record (m)
1	25	6	17	11	19	16	14
2	12	7	22	12	23	17	29
3	21	8	9	13	32	18	16
4	15	9	26	14	18	19	34
5	24	10	24	15	39	20	21

(1) Organize the records in table 1 into a frequency table as in table 2.

Table 2 Records of throwing a handball

Record (m) Greater than or equal to		Record (m) Under	Frequency (people)	Cumulative frequency (people)
5	-	10	**1**	**1**
10	-	15	**2**	**3**
15	-	20	**5**	**8**
20	-	25	**6**	**14**
25	-	30	**3**	**17**
30	-	35	**2**	**19**
35	-	40	**1**	**20**
Total			20	

Find the cumulative frequencies.
Find the cumulative frequency of distances under 15 m (add the frequencies of distances greater than or equal to 5 m and under 10 m, and those greater than or equal to 10 m and under 15 m).

① ⊕ ② (EXE)

```
1+2
                3
```

Add the frequency of distances greater than or equal to 15 m and under 20 m to the above total.

⊕ ⑤ (EXE)

```
Ans+5
                8
```

Add the following frequencies, in order, the same way.

⊕ ⑥ (EXE) ⊕ ③ (EXE)

```
Ans+3
               17
```

⊕ ② (EXE) ⊕ ① (EXE)

```
Ans+1
               20
```

(2) Find which class has the highest frequency.

Interpret the answer from the frequency table.

Class that is greater than or equal to $20\ \text{meters}$ and under $25\ \text{meters}$

(3) Find how many students have records under $30\ \text{meters}$.

Interpret the cumulative frequency of the class greater than or equal to $25\ \text{meters}$ and under $30\ \text{meters}$.

17 people

Histograms

TARGET To learn how to show frequency tables as histograms and frequency line graphs.

STUDY GUIDE

Histograms and frequency line graphs

Histograms are **graphs** based on data from frequency tables **that show a consecutive distribution of frequencies** using rectangles whose height is the frequency and base is the class width.

Frequency line graphs are **graphs that connect the center of the top of each rectangle** in a histogram.

EXERCISE

◆ The table on the right is a frequency table for a 50 meter race by a group of students. Solve the following problems for this data.

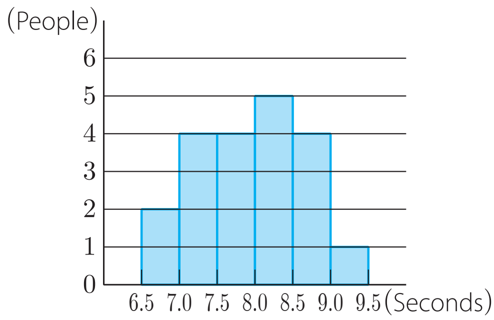
Records of 50-meter race

Class (sec.) Greater than or equal to		Under	Frequency (people)
6.0	-	6.5	0
6.5	-	7.0	2
7.0	-	7.5	4
7.5	-	8.0	4
8.0	-	8.5	5
8.5	-	9.0	4
9.0	-	9.5	1
9.5	-	10.0	0
Total			20

(1) Show this frequency table in a histogram.

<How to create a histogram from a frequency table>

Create a histogram by drawing consecutive rectangles whose height is the frequency and base is the class width.



check

Press ⌂, select [Statistics], press OK, select [1-Variable], press OK

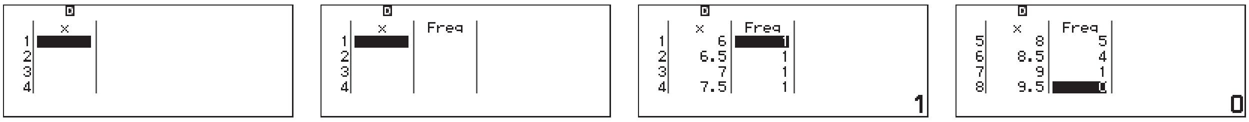
Press ⋯, select [Frequency], press OK, select [On], press OK, press AC

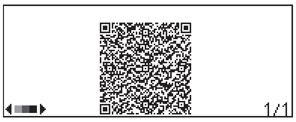
After inputting the minimum values of each class into x and the frequencies into Freq, scan the QR code.

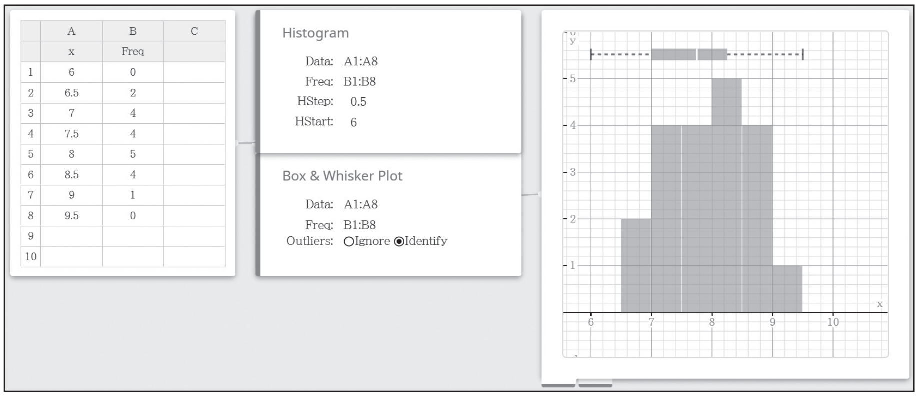
6 EXE 6 . 5 EXE 7 EXE 7 . 5 EXE 8 EXE 8 . 5 EXE 9 EXE 9 . 5 EXE

∨ > 0 EXE 2 EXE 4 EXE 4 EXE 5 EXE 4 EXE 1 EXE 0 EXE ⇧ x

Set the histogram interval to 0.5.



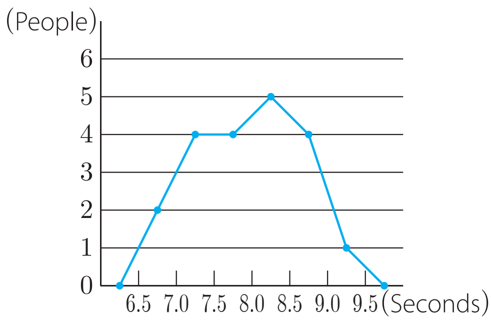




(2) Draw a frequency line graph from the histogram.

<How to create a frequency line graph from a histogram>

Use line segments to connect the midpoint of the tops of each rectangle in the histogram to create a line graph. Connect **the midpoint of the base of a class if its frequency is 0**. Draw the graph so the frequency of the rightmost class is 0 and the frequency of the leftmost class is 0.



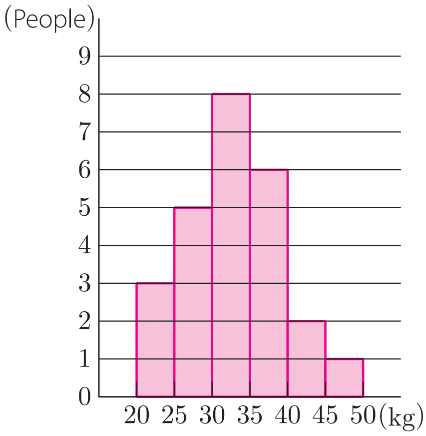
PRACTICE

◆ The table on the right is a frequency table for the grip strengths of a group of students. Solve the following problems for this data.

Records of grip strength

Class (kg)			Frequency (people)
Greater than or equal to		Under	
20	-	25	3
25	-	30	5
30	-	35	8
35	-	40	6
40	-	45	2
45	-	50	1
Total			25

(1) Show this frequency table in a histogram.



check

Press [HOME], select [Statistics], press [OK], select [1-Variable], press [OK]

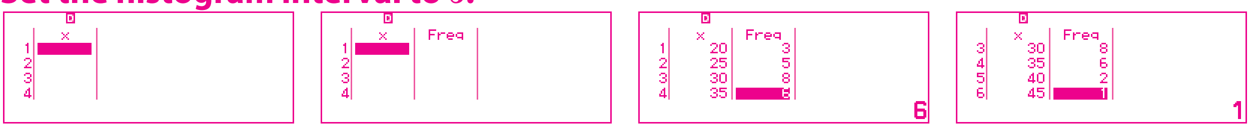
Press [○○○], select [Frequency], press [OK], select [On], press [OK], press [AC]

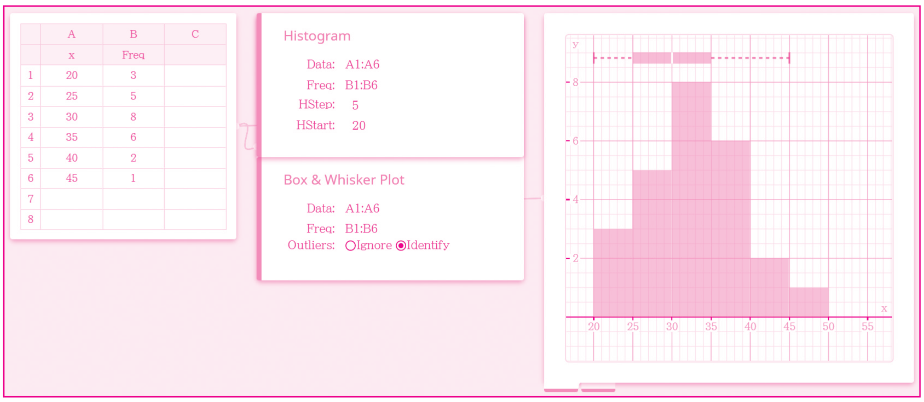
After inputting the minimum values of each class into x and the frequencies into Freq, scan the QR code.

[2] [0] [EXE] [2] [5] [EXE] [3] [0] [EXE] [3] [5] [EXE] [4] [0] [EXE] [4] [5] [EXE]

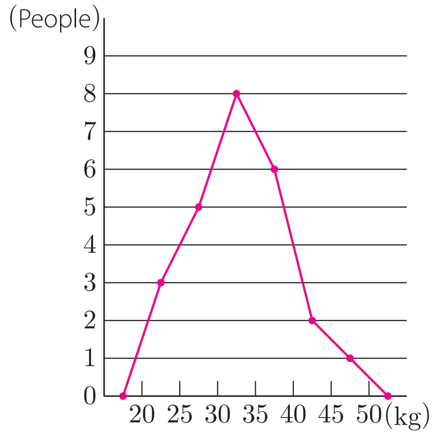
[⋁] [>] [3] [EXE] [5] [EXE] [8] [EXE] [6] [EXE] [2] [EXE] [1] [EXE] [SHIFT] [𝑥]

Set the histogram interval to 5.





(2) Draw a frequency line graph from the histogram.



Relative frequency

TARGET To learn about relative frequency and cumulative relative frequency.

STUDY GUIDE

Relative frequency and cumulative relative frequency

The proportion of the frequency of a class to that of everything else is called **relative frequency**. For each class, the sum of the relative frequencies from the first class to a certain class is called the **cumulative relative frequency**.

$$(\text{Relative frequency of a class}) = \frac{(\text{Frequency of that class})}{(\text{Total of frequencies})}$$

EXERCISE

◆ On the right, table 1 is a frequency table of records of side-step tests by homeroom 1 in a certain grade. Table 2 is the frequency table of records of side-step tests for all the students in that grade. Solve the following problems for this data.

Table 1 Record of side-step test for homeroom 1

Class (times) Greater than or equal to		Under	Frequency (people)
30	-	35	2
35	-	40	3
40	-	45	5
45	-	50	7
50	-	55	2
55	-	60	1
Total			20

Table 2 Records of side-step tests for all students in that grade

Class (times) Greater than or equal to		Under	Frequency (people)
30	-	35	5
35	-	40	13
40	-	45	27
45	-	50	24
50	-	55	8
55	-	60	3
Total			80

(1) Find the relative frequencies (to 2 decimal places) for table 1 and table 2 each, and then write them in tables 3 and 4.

Find

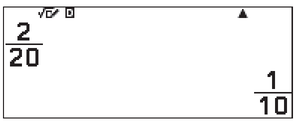
$$(\text{Relative frequency of a class}) = \frac{(\text{Frequency of that class})}{(\text{Total of frequencies})}.$$

Ex.

For the class of greater than or equal to 30 and under 35, the frequency is 2, so enter 20 for the total frequency of all classes.

Press ⌂, select [Calculate], press OK

② ⊟ ② ⓪ EXE



Press FORMAT, select [Decimal], press OK, show as decimal.

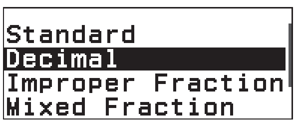


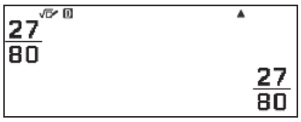
Table 3 Record of side-step test for homeroom 1

Class (times) Greater than or equal to		Under	Frequency (people)	Relative frequency formula	Relative frequency
30	-	35	2	$\frac{2}{20}$	**0.10**
35	-	40	3	$\frac{3}{20}$	**0.15**
40	-	45	5	$\frac{5}{20}$	**0.25**
45	-	50	7	$\frac{7}{20}$	**0.35**
50	-	55	2	$\frac{2}{20}$	**0.10**
55	-	60	1	$\frac{1}{20}$	**0.05**
Total			20	$\frac{20}{20}$	**1.00**

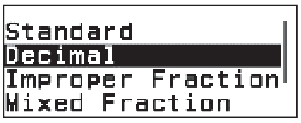
Ex.

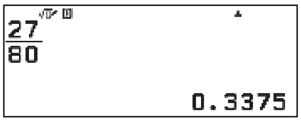
For the class of greater than or equal to 40 and under 45, the frequency is 27, so enter 80 for the total frequency of all classes.

② ⑦ ⊟ ⑧ ⓪ EXE



Press FORMAT, select [Decimal], press OK, show as decimal.





Round off the output value 0.3375 to 2 decimal places.

Table 4 Records of side-step tests for all students in that grade

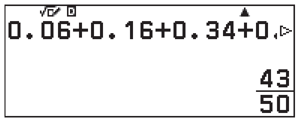
Class (times)			Frequency (people)	Relative frequency formula	Relative frequency
Greater than or equal to		Under			
30	-	35	5	$\frac{5}{80}$	**0.06**
35	-	40	13	$\frac{13}{80}$	**0.16**
40	-	45	27	$\frac{27}{80}$	**0.34**
45	-	50	24	$\frac{24}{80}$	**0.30**
50	-	55	8	$\frac{8}{80}$	**0.10**
55	-	60	3	$\frac{3}{80}$	**0.04**
Total			80	$\frac{80}{80}$	**1.00**

(2) Find cumulative relative frequency of the class of greater than or equal to 45 and under 50 of all students in that grade.

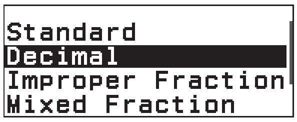
Add all the subsequent relative frequencies from the relative frequency of the class of greater than or equal to 30 and under 35 up to that of the class of greater than or equal to 45 and under 50.

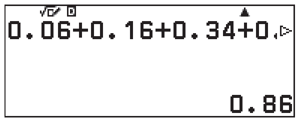
$0.06+0.16+0.34+0.30=0.86$

⓪ ⊙ ⓪ ⑥ ⊕ ⓪ ⊙ ① ⑥ ⊕ ⓪ ⊙ ③ ④ ⊕ ⓪ ⊙ ③ ⓪ EXE



Press FORMAT, select [Decimal], press OK, show as decimal.





0.86

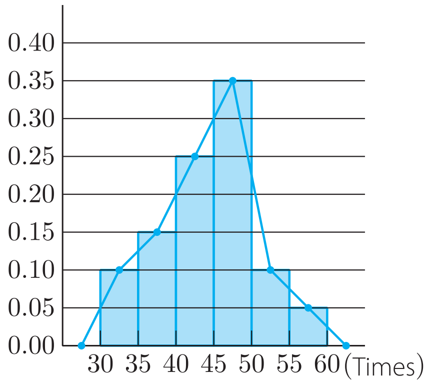
(3) Draw a histogram and a frequency line graph using the relative frequency in table 3 as the vertical axis.

<How to create a histogram from a relative frequency table>

Create a histogram by drawing consecutive rectangles whose height is the relative frequency and base is the class width.

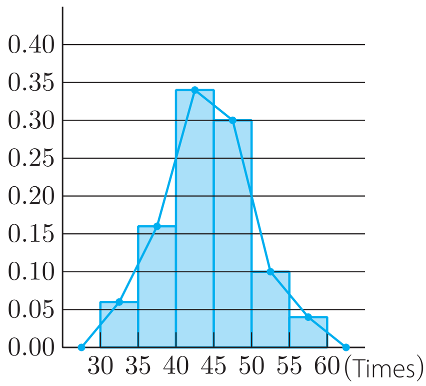
<How to create a frequency line graph from a histogram>

Use line segments to connect the midpoint of the tops of each rectangle in the histogram to create a line graph. Connect the midpoint of the base of a class if its frequency is 0. Draw the graph so the frequency of the rightmost class is 0 and the frequency of the leftmost class is 0.



(4) Draw a histogram and a frequency line graph using the relative frequency in table 4 as the vertical axis.

Draw it the same as for (3).



(5) Based on (3) and (4), select from the following the correct characteristic for the record of homeroom 1 compared with all students in that grade.

(a) All students in that grade had a larger proportion of greater than or equal to 40 and under 45, but homeroom 1 had a larger proportion of greater than or equal to 45 and under 50.

(b) All students in that grade had a larger proportion of greater than or equal to 45 and under 50, but homeroom 1 had a larger proportion of greater than or equal to 40 and under 45.

(c) All students in that grade and homeroom 1 had the same proportion of greater than or equal to 35 and under 40.

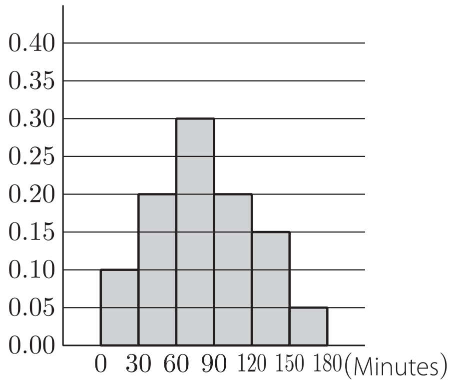
Interpret the graphs.

(a)

PRACTICE

◆ The records for time spent reading on Sundays by homeroom 1 and all students in that grade are organized in the following frequency table. In addition, there is a histogram of homeroom 1, created by interpreting the frequency table below. Solve the following problems for this data.

Class (minutes)			Frequency (people)		Relative frequency	
			Homeroom 1	Entire grade	Homeroom 1	Entire grade
Greater than or equal to		Under				
0	-	30	2	9	0.10	**0.15**
30	-	60	4	24	0.20	**0.40**
60	-	90	6	12	0.30	**0.20**
90	-	120	4	6	0.20	**0.10**
120	-	150	3	6	0.15	**0.10**
150	-	180	1	3	0.05	**0.05**
Total			20	60	1.00	1.00



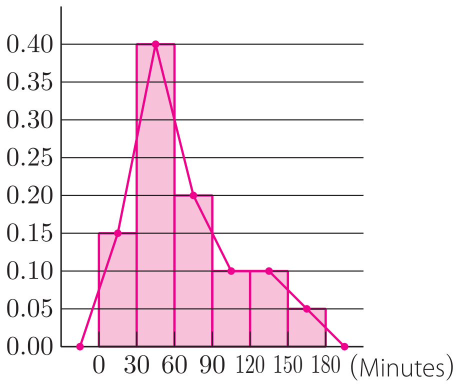
(1) Find the relative frequencies of each class for all students in that grade and then write them in the table.

For the class of greater than or equal to 0 and under 30, divide the frequency 9 by the total frequency 60.

⑨ ÷ ⑥ ⓪ EXE

Press FORMAT, select [Decimal], press OK, show as decimal.

(2) Draw a histogram and a frequency line graph using the relative frequency of all students in that grade.



(3) Select from the following the correct characteristic for the record of homeroom 1 compared with all students in that grade.

(a) All students in that grade had a larger proportion of greater than or equal to 60 and under 90, but homeroom 1 had a larger proportion of greater than or equal to 90 and under 120.

(b) All students in that grade had a larger proportion of greater than or equal to 30 and under 60, but homeroom 1 had a larger proportion of greater than or equal to 60 and under 90.

(c) All students in that grade and homeroom 1 had the same proportion of greater than or equal to 0 and under 30.

(b)

How to describe the characteristics of data distribution

TARGET To learn how to compare numerical values for the characteristics of data distribution.

STUDY GUIDE

Maximum/minimum values, ranges, and representative values in data distribution

The difference between the **maximum value** and **minimum value** of the data is called the **range** of the distribution. The distribution of data can be shown by its range.

When a feature of an entire set of data is represented by 1 value, that value is called the **representative value**, some of which are shown here.

(a) **Average** The sum of the data values divided by the total number of data values. This value indicates an average.

(b) **Median** The middle value of a set of data values arranged in order of magnitude.

(c) **Mode** The value that appears most often in a set of data values. **The mode is the median value of the class with the highest frequency in a frequency table.**

EXERCISE

◆ The data below show the vertical jump records (unit: cm), of 20 students, arranged in ascending order. Now, solve the following problems.

19, 20, 22, 23, 26, 28, 30, 31, 31, 35, 36, 36, 36, 39, 39, 42, 43, 43, 44, 45

(1) Find the range of distribution.

(Range of distribution)=(Maximum value)−(Minimum value)$=45-19=26$ (cm)

26 cm

(2) Find the average.

The sum of the data values divided by the total number of data values is the average.

$$\frac{19+20+22+23+26+28+30+31+31+35+36+36+36+39+39+42+43+43+44+45}{20}$$

$=33.4$ (cm)

33.4 cm

(3) Find the median.

For an odd number of data, the (median)=(middle value), and for an even number of data, the (median)=(the average of the 2 middle values).

$$\frac{35+36}{2}=35.5(\text{cm})$$

35.5 cm

(4) Find the mode.

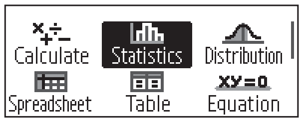
The value that appears most often in this set of data values is 36.

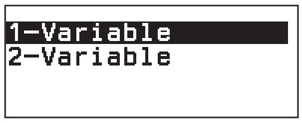
36 cm

check

Press (HOME), select [Statistics], press (OK), select [1-Variable], press (OK)

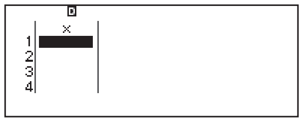
Press (TOOLS), select [Frequency], press (OK), select [Off], press (OK), press (AC)





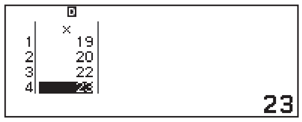
Input data in the x column from line 1 in order downward.

19, 20, 22, 23, 26, 28, 30, 31, 31, 35, 36, 36, 36, 39, 39, 42, 43, 43, 44, 45



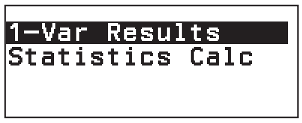
① ⑨ (EXE) ② ⓪ (EXE) ② ② (EXE) ② ③ (EXE)

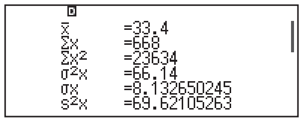
② ⑥ (EXE) ② ⑧ (EXE)



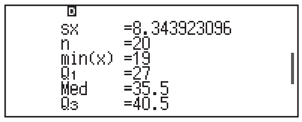
Do the same input below, and after inputting the final data, press (EXE),

select [1-Var Results], press (OK)



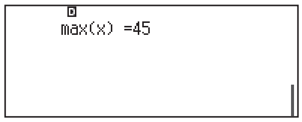


Average : $\bar{x} = 33.4$



Minimum : min(x)=19

Median : Med=35.5



Maximum : max(x)=45

PRACTICE

1 Solve the following problems.

(1) What is the difference between the maximum value and minimum value of the data called?

Range

(2) What is the middle value of a set of data values arranged in order of magnitude called?

Median

(3) What is the value that appears most often in a set of data values called?

Mode

2 The data below show the vertical jump records (unit: cm), of 20 students, arranged in ascending order. Now, solve the following problems.

21, 23, 25, 27, 28, 30, 31, 33, 33, 35, 36, 36, 36, 38, 38, 40, 41, 41, 43, 45

(1) Find the range of distribution.

(Range of distribution)=(Maximum value)−(Minimum value)=45−21=24 (cm)

24 cm

(2) Find the average.

$$\frac{680}{20}=34\ (\text{cm})$$

34 cm

(3) Find the median.

There are 20 values in the data, which is even, so find the average of the 10th and 11th values.

35.5 cm

(4) Find the mode.

The value that appears most often in this set of data values is 36.

36 cm

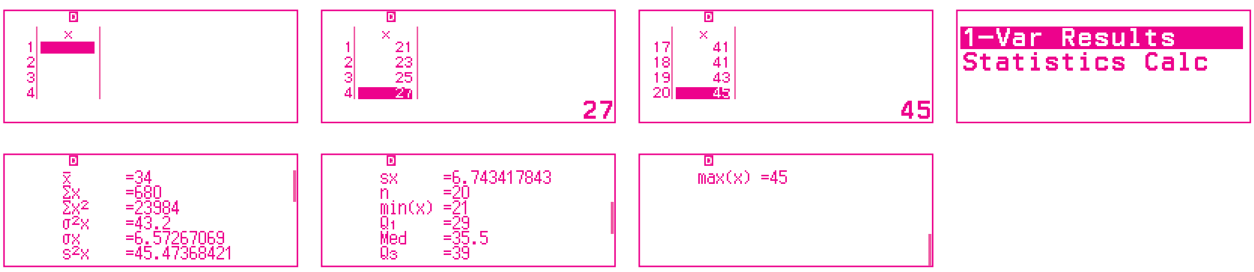
check

Press ⌂, select [Statistics], press OK, select [1-Variable], press OK

Input data in the x column.

② ① EXE ② ③ EXE ② ⑤ EXE ② ⑦ EXE

Do the same input below, and after inputting the final data, press EXE, select [1-Var Results], press OK



How to use data

TARGET To learn how to solve problems by analyzing data using representative values, etc.

EXERCISE

◆ The table below shows the vertical jump records (unit: occurrences), of 2 groups of students, arranged in order of occurrences. Now, solve the following problems.

	Data 1	Data 2	Data 3	Data 4	Data 5	Data 6	Data 7	Data 8	Data 9	Data 10
Group A	23	29	29	30	30	31	31	31	39	43
Group B	24	28	29	30	31	32	33	33	36	40

(1) Find the range of distribution of each of the groups' records.

Use (Range of distribution)=(Maximum value)−(Minimum value) to find the range of distribution.

Range of distribution of group A $43-23=20$ (occurrence)

Range of distribution of group B $40-24=16$ (occurrence)

Group A ... 20 occurrences, group B ... 16 occurrences

(2) Find the median of each of the groups' records.

Since the number of data is even, find the average of the 2 numbers in the middle.

Median of group A $\frac{30+31}{2}=30.5$ (occurrence)

Median of group B $\frac{31+32}{2}=31.5$ (occurrence)

Group A ... 30.5 occurrences, group B ... 31.5 occurrences

(3) Find the mode of each of the groups' records.

Select the value that appears most often in the table.

Group A ... 31 occurrences, group B ... 33 occurrences

(4) Find the total and average of each of the groups' records.

Find the sum (= total) of each of the groups' data values, and then divide that by the total number of data values to find the average.

Total of group A $23+29+29+30+30+31+31+31+39+43=316$ (occurrence)

Average of group A $\frac{316}{10}=31.6$ (occurrence)

Total of group B $24+28+29+30+31+32+33+33+36+40=316$ (occurrence)

Average of group B $\frac{316}{10}=31.6$ (occurrence)

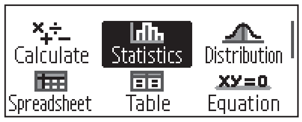
Group A … total 316 occurrences, average 31.6 occurrences,
group B … total 316 occurrences, average 31.6 occurrences

(5) Describe the characteristics of the 2 groups from the numbers in (1) to (4). Furthermore, describe the overall comparison of the 2 groups.

(Example) From (1), group A has records distributed over a wider range than group B.
Furthermore, group A has a smaller minimum value and a larger maximum value than group B.
From (2) and (3), group A has a smaller median and mode than group B.
From (4), the total and average are the same.
From the above, we can see overall that group B has better records than group A.

check

Press (HOME), select [Statistics], press (OK), select [2-Variable], press (OK)



Press (⋯), select [Frequency], press (OK), select [Off], press (OK), press (AC)

Input group A data in the x column and group B in the y column.

Input data in the x column from line 1 in order downward.

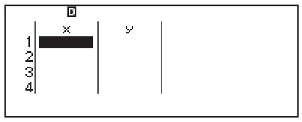
	Data 1	Data 2	Data 3	Data 4	Data 5	Data 6	Data 7	Data 8	Data 9	Data 10
Group A	23	29	29	30	30	31	31	31	39	43
Group B	24	28	29	30	31	32	33	33	36	40

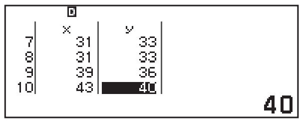
Input values for group A. (Input values for group B in the same way.)

② ③ (EXE) ② ⑨ (EXE) ② ⑨ (EXE) ③ ⓪ (EXE)

③ ⓪ (EXE) ③ ① (EXE) ③ ① (EXE) ③ ① (EXE)

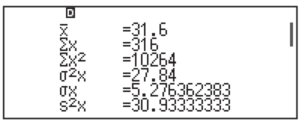
③ ⑨ (EXE) ④ ③ (EXE)

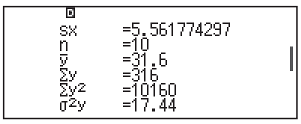


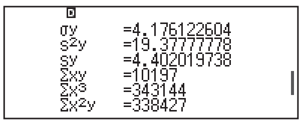


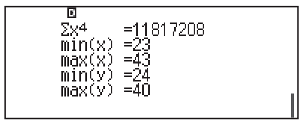
Input the y column in the same way, and after inputting the final data, press (EXE),

select [2-Var Results], press (OK)









PRACTICE

◆ The table below shows the weight (unit: g) of eggs in 2 containers A and B, arranged in order of lightness. Now, solve the following problems.

	Data 1	Data 2	Data 3	Data 4	Data 5	Data 6	Data 7	Data 8	Data 9	Data 10
Container A	48.8	49.4	49.7	51.1	51.5	51.7	52.2	53.1	53.2	54.3
Container B	44.2	45.9	46.8	51.2	51.3	51.9	54.6	54.7	56.3	58.1

(1) Find the median of each container.

Find the average of the 5^{th} and 6^{th} values.

Container A…51.6 g, container B…51.6 g

(2) Find the average of each container.

The average value of container A is $\frac{515}{10}=51.5$ (g), and the average value of container B is $\frac{515}{10}=51.5$ (g)

Container A…51.5 g, container B…51.5 g

(3) Find the range of distribution of each container.

Range of distribution of container A is $54.3-48.8=5.5$ (g)

Range of distribution of container B is $58.1-44.2=13.9$ (g)

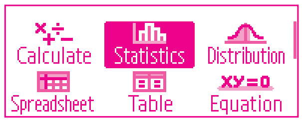
Container A…5.5 g, container B…13.9 g

(4) What can you say about the differences in the weights of the eggs in container A and B?

(Example) Compared with container A, the weights of eggs in container B are more dispersed.

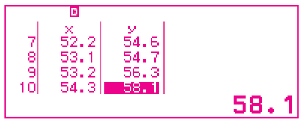
check

Press ⌂, select [Statistics], press OK, select [2-Variable], press OK
Press ◦◦◦, select [Frequency], press OK, select [Off], press OK, press AC

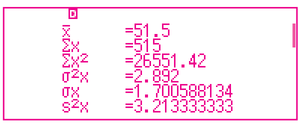


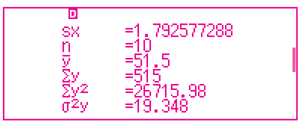
Input group A data in the x column and group B in the y column.

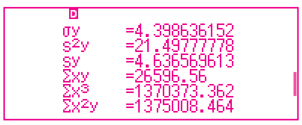
4 8 . 8 EXE 4 9 . 4 EXE

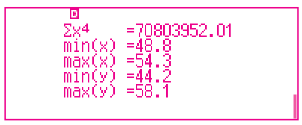


Do the same input below, and after inputting the final data, press EXE, select [2-Var Results], press OK









Relative frequency and probability

TARGET To learn to show numerically the likelihood of something happening.

STUDY GUIDE

Defining probability

Take a coin toss as an example. Before you throw the coin, no one can say absolutely whether it will be heads or tails. However, if we toss the coin many times repeatedly, we can predict the proportion of how many times heads comes up. We call the number that indicates the likelihood of this happening the **probability** of the coin coming up heads.

Probability: **A number indicating the degree to which something is likely to happen**.

Increasing the number of times we do an experiment brings the relative frequency of something happening closer to the value of the probability.

EXERCISE

◆ The following table shows the data collected from the number of times a coin was tossed. Now, solve the following problems.

Number of tosses (occurrences)	10	50	100	200	400	600	800	1000
Number of heads (occurrences)	4	18	41	73	153	229	305	382
Relative frequency formula	$\frac{4}{10}$	$\frac{18}{50}$	$\frac{41}{100}$	$\frac{73}{200}$	$\frac{153}{400}$	$\frac{229}{600}$	$\frac{305}{800}$	$\frac{382}{1000}$
Relative frequency	**0.40**	**0.36**	**0.41**	**0.37**	**0.38**	**0.38**	**0.38**	**0.38**

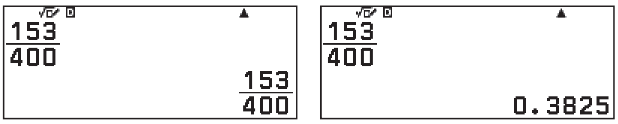
(1) Find the relative frequencies (to 2 decimal places) for heads coming up for each number of tosses, and write them in the table.

Use $(\text{Relative frequency}) = \frac{(\text{Number of heads})}{(\text{Number of tosses})}$ to find the answer.

Ex. The coin came up heads 153 times out of 400 tosses (rounded off at 3rd decimal place).

Press ⌂, select [Calculate], press OK, after inputting the values, select [Decimal] in FORMAT, press OK, show as decimal.

⌂ OK   1 5 3 ▭ 4 0 0 EXE   FORMAT ∨ OK



(2) What do you think is the approximate probability of heads coming up when a coin is tossed? Solve to 2 decimal places.

As the number of tosses increases, the relative frequency approaches 0.38.

0.38

PRACTICE

◆ The following table shows the data collected from the number of times a thumbtack was tossed and landed facing up. Now, solve the following problems.

Number of tosses (occurrences)	700	800	900	1000
Number facing up (occurrences)	390	441	510	573
Relative frequency formula	**$\frac{390}{700}$**	**$\frac{441}{800}$**	**$\frac{510}{900}$**	**$\frac{573}{1000}$**
Relative frequency	**0.56**	**0.55**	**0.57**	**0.57**

(1) Find the relative frequencies (to 2 decimal places) for facing up for each number of tosses, and write them in the table.

The thumbtack faced up 390 times out of 700 tosses (rounded off at 3rd decimal place).

Press ⌂, select [Calculate], press OK, after inputting the values, select [Decimal] in FORMAT, press OK, show as decimal.

⌂ OK　3 9 0 ⊟ 7 0 0 EXE　FORMAT ∨ OK

(2) What do you think is the approximate probability of the thumbtack facing up when it is tossed? Solve to 2 decimal places.

As the number of tosses increases, the relative frequency approaches 0.57.

0.57

Quartiles and box-and-whisker plots

TARGET To learn quartiles and box-and-whisker plots as procedures to compare data.

STUDY GUIDE

Quartiles, 5-number summaries, and box-and-whisker plots

Quartiles

Quartiles are the 3 values that divide the data into 4 equal parts by arranging the data in order of size. They are called the **1st quartile**, **2nd quartile**, and **3rd quartile**, in order from the smaller value.

Ex. 2 , 5 , 9 , 11 , 16 , 22 , 31

5 ↑ 1st quartile; 11 ↑ 2nd quartile; 22 ↑ 3rd quartile

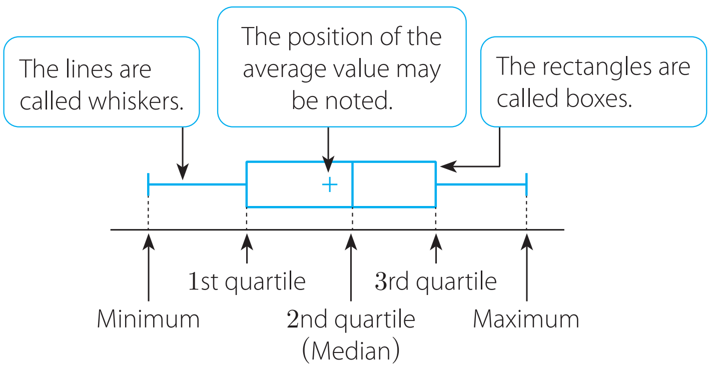
5-number summaries and box-and-whisker plots

The **5-number summary** is a method to show the dispersion of data by using 5 values: the minimum value, 1st quartile, 2nd quartile, 3rd quartile, and maximum value. Showing a 5-number summary graphically is called a **box-and-whisker plot**.

The difference between the 3rd quartile and the 1st quartile is called the **interquartile range**, half of the interquartile range is called the **quartile deviation**.

(Interquartile range)=(3rd quartile)−(1st quartile)

(Quartile deviation)=$\dfrac{\text{(Interquartile range)}}{2}$



How to find quartiles

Quartiles are found as follows.

(Median of the group)=(Quartile)

(2nd quartile)=(Median of entire group)

(1st quartile)=(Median of first half of group)

(3rd quartile)=(Median of last half of group)

EXTRA Info.

How to find the median

For an odd number of data

→The center value is the median

For an even number of data

→The average of the 2 numbers in the center is the median

EXERCISE

1 Find the minimum, 1st quartile, 2nd quartile, 3rd quartile, maximum, and interquartile range of the data on the right. Also, draw a box-and-whisker plot.

9, 13, 7, 2, 19, 17, 6, 1, 10, 4, 18

First, arrange the data values in order of magnitude, from the smallest.

1 , 2 , 4 , 6 , 7 , 9 , 10 , 13 , 17 , 18 , 19

↑ Minimum (1) ↑ 1st quartile (4) ↑ 2nd quartile (9) ↑ 3rd quartile (17) ↑ Maximum (19)

Since the total number of data is 11, the 2nd quartile (median) is $11 \div 2 = 5$ with a remainder of 1, which is the 6th smallest.

The 2nd quartile is a boundary between the lowest half of the group and the highest half of the group of data.

The lowest half of the group has 5 data, so the 1st quartile is $5 \div 2 = 2$ with a remainder of 1, which is the 3rd smallest.

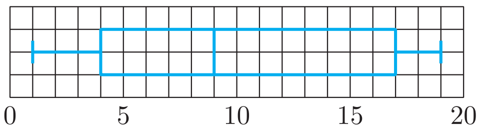
The highest half of the group has 5 data, so the 3rd quartile is the 9th smallest.

From (interquartile range)=(3rd quartile)−(1st quartile), we get $17 - 4 = 13$

Minimum …1, 1st quartile…4, 2nd quartile …9, 3rd quartile…17, maximum …19,

interquartile range…13

The box-and-whisker plot is shown on the right.



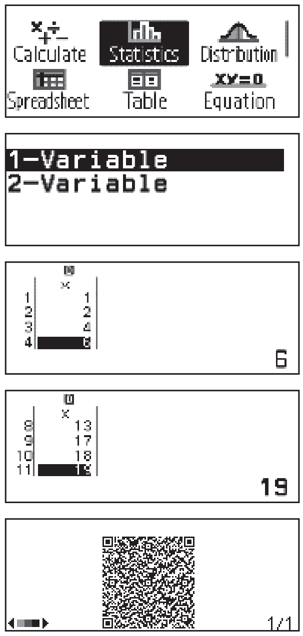
check

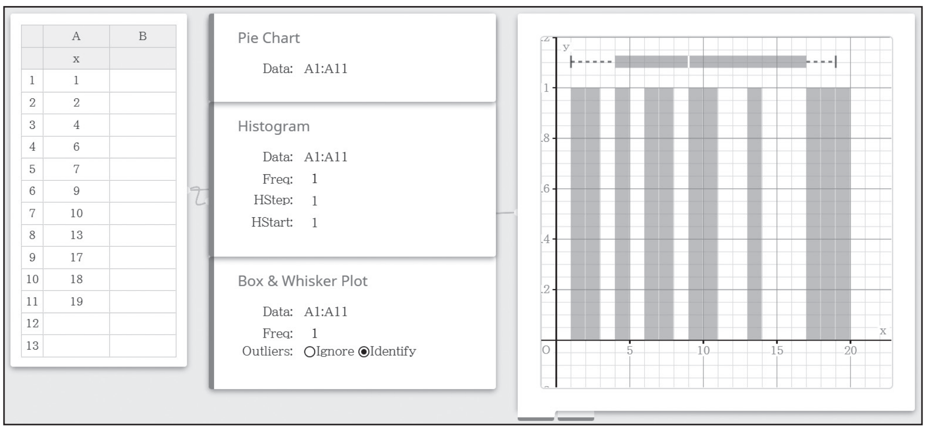
Press ⌂, select [Statistics], press OK, select [1-Variable], press OK, input data in the x column.

1, 2, 4, 6, 7, 9, 10, 13, 17, 18, 19

① EXE ② EXE ④ EXE ⑥ EXE

Do the same input below, and after inputting the final data, press ⇧, press ⓧ, use the QR code to confirm the box-and-whisker plot.

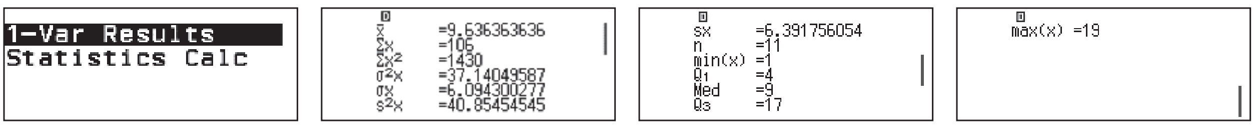




Press ↩, press EXE, select [1-Var Results], press OK, confirm the values.

Minimum: Min (x)=1　　1st quartile: $Q_1=4$　　2nd quartile (Q_2)=Median: Med=9

3rd quartile: $Q_3=17$　　Maximum: Max (x)=19



2 Find the minimum, 1st quartile, 2nd quartile, 3rd quartile, maximum, and interquartile range of the data on the right. Also, draw a box-and-whisker plot.

15, 7, 12, 9, 4, 18, 5, 11, 2, 10, 3, 16

First, arrange the data values in order of magnitude, from the smallest.

2 , 3 , 4 , 5 , 7 , 9 , 10 , 11 , 12 , 15 , 16 , 18

Minimum (2) — 1st quartile (between 4 and 5) — 2nd quartile (between 9 and 10) — 3rd quartile (between 12 and 15) — Maximum (18)

Since the total number of data is 12, the 2nd quartile (median) is $12 \div 2 = 6$, so it is the average of the 6th and 7th smallest.

$$(\text{2nd quartile}) = \frac{9+10}{2} = 9.5$$

The 2nd quartile is a boundary between the lowest half of the group and the highest half of the group of data.

The lowest half of the group has 6 data, so the 1st quartile is $6 \div 2 = 3$, making it the average of the 3rd and 4th smallest.

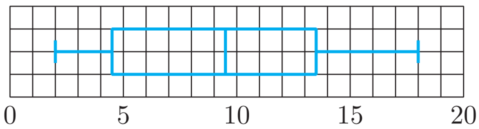
The highest half of the group has 6 data, so the 3rd quartile is the average of the 9th and 10th smallest.

$$(\text{1st quartile}) = \frac{4+5}{2} = 4.5,\ (\text{3rd quartile}) = \frac{12+15}{2} = 13.5$$

From (interquartile range)=(3rd quartile)−(1st quartile), we get $13.5 - 4.5 = 9$

Minimum …2, 1st quartile…4.5, 2nd quartile …9.5, 3rd quartile…13.5, maximum …18, interquartile range…9

The box-and-whisker plot is shown on the right.



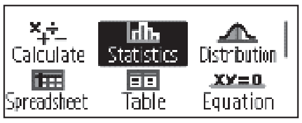
check

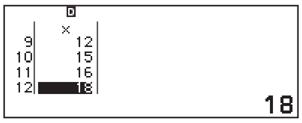
Press ⌂, select [Statistics], press OK, select [1-Variable], press OK, input data in the x column.

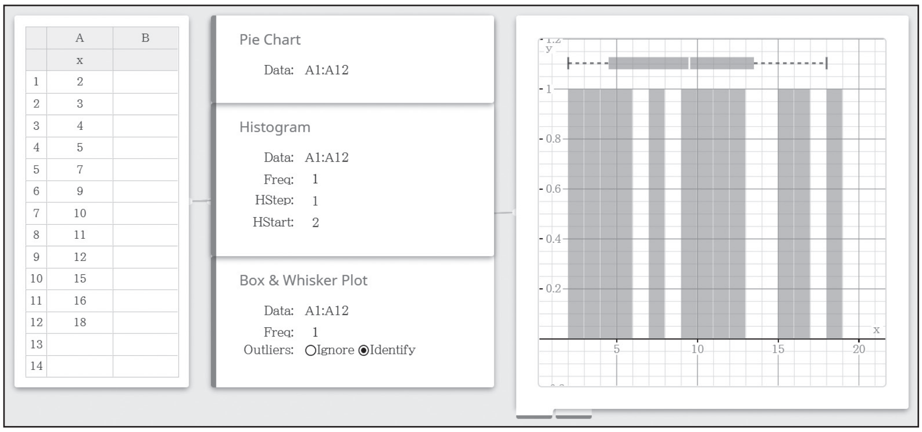
2, 3, 4, 5, 7, 9, 10, 11, 12, 15, 16, 18

② EXE ③ EXE ④ EXE ⑤ EXE

Do the same input below, and after inputting the final data, press ⇧, press ⓧ, use the QR code to confirm the box-and-whisker plot.



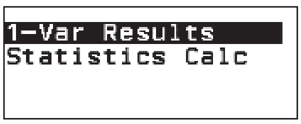


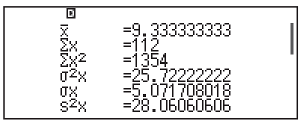


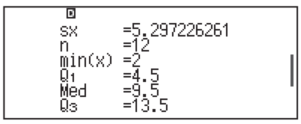
Press ↺, press EXE, select [1-Var Results], press OK, confirm the values.

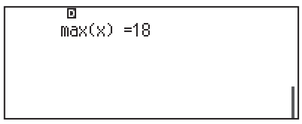
Minimum: Min (x)=2 1st quartile: Q_1=4.5 2nd quartile (Q_2)=Median: Med=9.5

3rd quartile: Q_3=13.5 Maximum: Max (x)=18









PRACTICE

1 Solve the following problems.

(1) What do you call the values in positions that divide a group of data values, which have been arranged in order of magnitude, into 4 equal parts?

Quartiles

(2) What do you call the 3 values from (1), from the smallest to the largest?

1st quartile, 2nd quartile, and 3rd quartile

(3) What do you call figures using boxes and lines to show the minimum and maximum values, and the values from (1)?

Box-and-whisker plots

2 The data below show the handball throwing records of 11 students, arranged in ascending order. Find the quartiles of this data.

14, 16, 17, 19, 19, 20, 21, 22, 23, 28, 33 (units: m)

There are 11 values in the data, so the median is the 6th value, 20 m. The 1st quartile is the 3rd value, 17 m. The 3rd quartile is the 9th value, 23 m.

1st quartile…17 m, 2nd quartile…20 m, 3rd quartile…23 m

3 The data below show the English test results of 10 students, arranged in ascending order. Find the quartiles of this data.

36, 41, 49, 54, 59, 61, 72, 77, 85, 91 (units: points)

There are 10 values in the data, so the median is the average of the 5th value, 59 points, and 6th value, 61 points,

$$\frac{59+61}{2}=60\text{ (points)}$$

The 1st quartile is the 3rd value, 49 points. The 3rd quartile is the 8th value, 77 points.

1st quartile…49 points, 2nd quartile…60 points, 3rd quartile…77 points

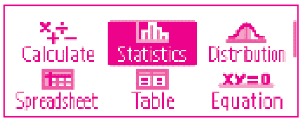
check

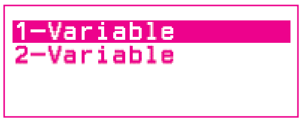
Press ⌂, select [Statistics], press OK, select [1-Variable], press OK

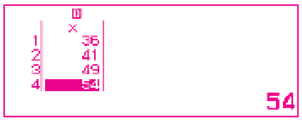
Input data in the x column.

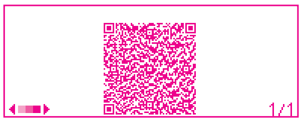
③ ⑥ EXE ④ ① EXE ④ ⑨ EXE

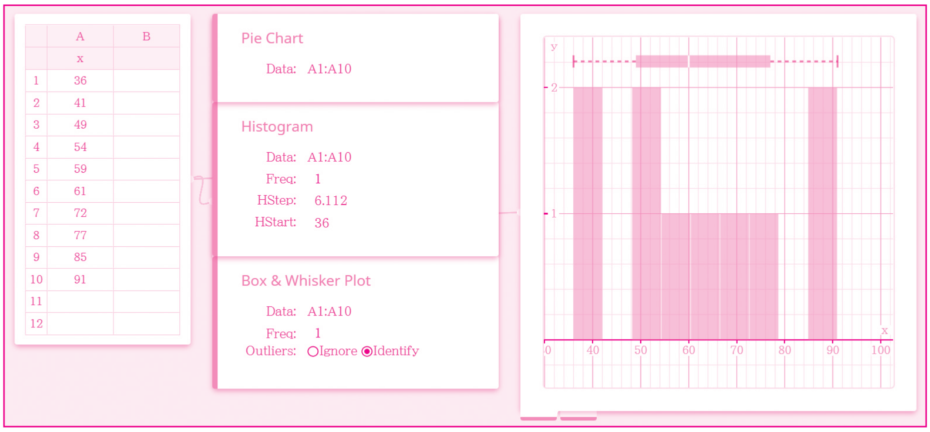
Do the same input below, and after inputting the final data, press ⬆, press x, use the QR code to confirm the box-and-whisker plot.



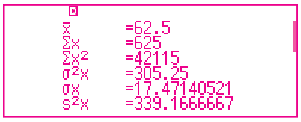


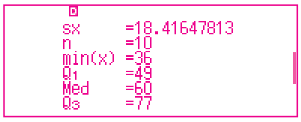


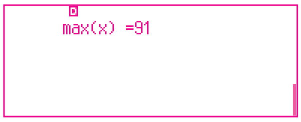




Press ↺, press EXE, select [1-Var Results], press OK, confirm the values.

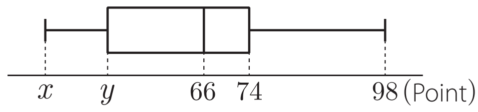






PRACTICE

[4] The figure on the right is a box-and-whisker plot showing the dispersion of the results of a high-school math test. Find the values of x and y in the figure, for a data range of 60 points and a quartile deviation of 12.5 points.



A range of $(98-x)$ points is shown, so from $98-x=60$, we get $x=38$(points)

The interquartile range is $12.5\times2=25$ (points), so from $74-y=25$, we get $y=49$ (points)

$x=38$, $y=49$

[5] Figure 1 and figure 2 on the right are box-and-whisker plots showing the distribution of average temperatures for 1 month at location A and location B. Now, solve the following problems.

Figure 1

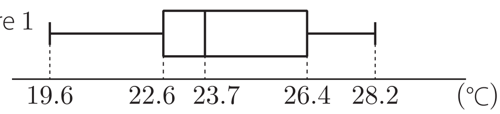
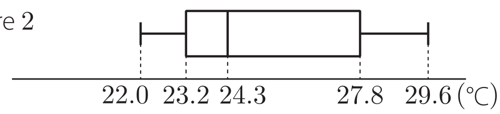


Figure 2



(1) Find the range of the data for location A.

From figure 1, we get $28.2-19.6=8.6$ (°C)

8.6°C

(2) The following summarizes what we can understand by comparing figure 1 and figure 2. Fill in the blanks [a] to [c] with "Location A" or "Location B" to complete the following text.

The median of [a] is larger, and the quartile deviation of [b] is larger. Furthermore, [c] has the larger temperature difference.

From figure 1 and figure 2, we can see the median at location A is 23.7°C and at location B it is 24.3°C, so location B is larger.

And, since the quartile deviation is half of the interquartile range, the inequality relation can be obtained by considering the quartile range, so location A is $26.4-22.6=3.8$(°C) and location B is $27.8-23.2=4.6$(°C), making location B larger.

The temperature difference is larger where the range is larger. Location A is $28.2-19.6=8.6$ (°C) and location B is $29.6-22.0=7.6$ (°C). So, the temperature difference is larger at location A.

[a]…Location B, [b]…Location B, [c]…Location A

Using histograms and box-and-whisker plots

TARGET To read data distribution from histograms and box-and-whisker plots.

STUDY GUIDE

How to read the characteristics and distribution of histograms and box-and-whisker plots

Characteristics of histograms

· Mode is easy to understand.

Characteristics of box-and-whisker plots

· Median is easy to understand.

How to read box-and-whisker plots

· Long whiskers. →Data distribution is large.

· Short whiskers. →Data distribution is small.

· Box positioned to the left. →Peak of histogram is to the left.

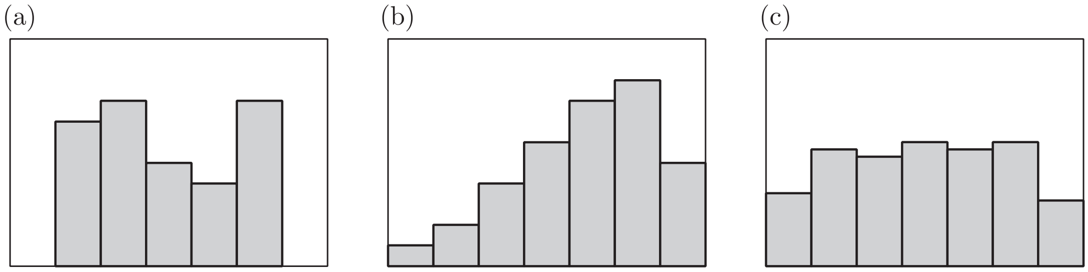
· Box positioned to the right. →Peak of histogram is to the right.

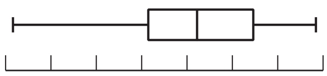
· Box positioned in the center. →Peak of histogram is in the center.

· Interquartile range is small. →Data is centralized.

EXERCISE

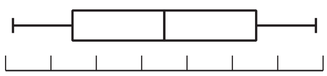
1 Below are the histograms (a) to (c). Solve by selecting which box-and-whisker plot corresponds to each histogram.



(1) 

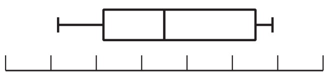
The median is to the right, and the range is large, so the corresponding histogram is (b).

(b)

(2) 

It is symmetrical, so the corresponding histogram is (c).

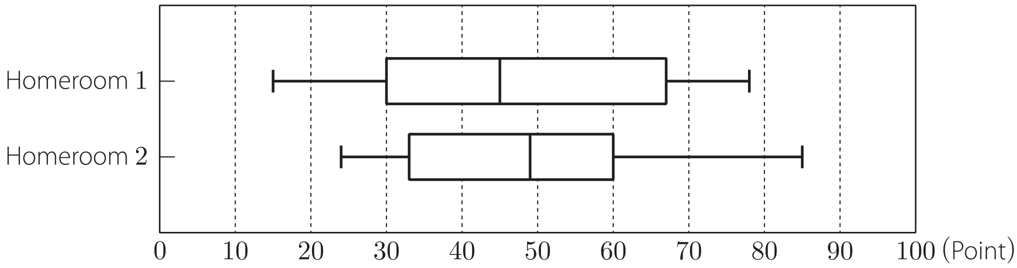
(c)

(3) 

The range is small, so the corresponding histogram is (a).

(a)

2 The figure below shows box-and-whisker plots showing the dispersion of the scores of a high-school math test taken by homeroom 1 and homeroom 2, which have 40 people each. Select all the answers from (a) to (d) that correctly describe these box-and-whisker plots.



(a) In both homerooms, greater than or equal to 20 students scored greater than or equal to 30 and less than 70 points.

(b) In both homerooms, no students scored less than or equal to 20 points.

(c) In both homerooms, the average score was greater than or equal to 40 points.

(d) Homeroom 2 has more centralized data.

In both homerooms, the 1st quartile scored greater than or equal to 30 points and the 3rd quartile scored less than 70 points, so greater than or equal to 20 students scored greater than or equal to 30 and less than 70 points. Therefore, (a) is correct.

Since the lowest score in homeroom 1 is less than or equal to 20 points, a student in homeroom 1 scored less than or equal to 20 points. Therefore, (b) is not correct.

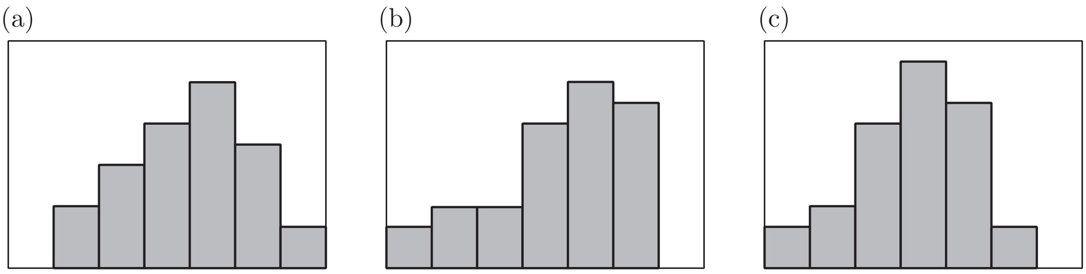
These box-and-whisker plots do not show the position of the average, so the average scores cannot be compared. Therefore, (c) is not correct.

By comparing the interquartile range (horizontal length of the boxes), we can see that of homeroom 2 is smaller, so the data in homeroom 2 is more centralized. Therefore, (d) is correct.

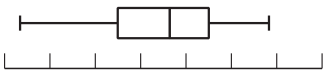
(a), (d)

PRACTICE

1 Below are the histograms (a) to (c). Solve by selecting which box-and-whisker plot corresponds to each histogram.



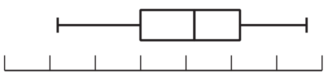
(1)



Since the minimum value is in the 1st column and the maximum value is in the 6th column, the peak of the data is centralized, so the answer is (c).

(c)

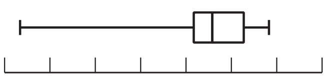
(2)



The minimum value is in the 2nd column, so the answer is (a).

(a)

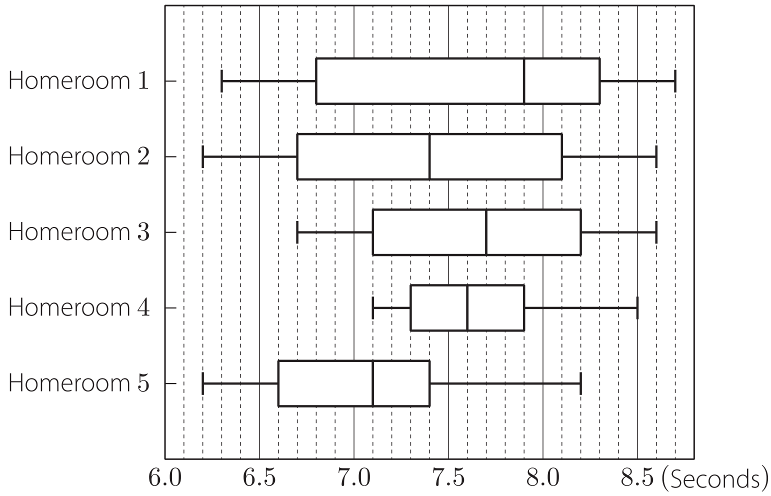
(3)



The minimum value is in the 1st column and the maximum value is in the 6th column, and the data is to the right, so the answer is (b).

(b)

2 The figure below shows box-and-whisker plots summarizing the records of 50-meter races by students in homeroom 1 to homeroom 5, each having 40 students. Select all the homerooms that correspond to the conditions (1) to (4) for these box-and-whisker plots.



(1) No students ran in less than 7 seconds.

Answer with the homerooms whose minimum value is greater than or equal to 7 seconds.

Homeroom 4

(2) Greater than or equal to 30 students ran in less than 7.5 seconds.

Answer with the homerooms whose 3rd quartile is less than 7.5 seconds.

Homeroom 5

(3) There are definitely greater than or equal to 10 students who ran in greater than or equal to 8 seconds.

Answer with the homerooms whose 3rd quartile is greater than or equal to 8 seconds.

Homerooms 1, 2, and 3

(4) Definitely greater than or equal to half of the students ran in greater than or equal to 7.5 seconds.

Answer with the homerooms whose 2nd quartile is greater than or equal to 7.5 seconds.

Homerooms 1, 3, and 4

Distribution of data

TARGET To find the variance s^2 and standard deviation s from the values of data.

STUDY GUIDE

Variance and standard deviation

Consider a number n of data $x_1, x_2, x_3, \cdots, x_n$ for variate x.

The average value is denoted as $\overline{x}$, such that $\overline{x} = \dfrac{x_1 + x_2 + x_3 + \cdots + x_n}{n}$.

The difference between the value of the data and the average value of the data ($x_1 - \overline{x}, x_2 - \overline{x}, x_3 - \overline{x}, \cdots, x_n - \overline{x}$) is called the **deviation**.

If we calculate the average of the deviation here, it is always 0, so it is not suitable for representing the state of distribution.

Then, when we consider the average of the square of the deviation $\left(\dfrac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + (x_3 - \overline{x})^2 + \cdots + (x_n - \overline{x})^2}{n} \right)$,

those values are all positive values, so the larger those values are the greater the distribution of the data.

This value is called **variance**, and the **positive square root of variance** is called **standard deviation**.

Variance $$s^2 = \frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + (x_3 - \overline{x})^2 + \cdots + (x_n - \overline{x})^2}{n}$$

Standard deviation $$s = \sqrt{\frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + (x_3 - \overline{x})^2 + \cdots + (x_n - \overline{x})^2}{n}}$$

The variance of data is the value obtained by squaring the original data; for example, when the original data expresses lengths in cm, the variance becomes the square of cm and the units do not agree. Then, we find the square root so the units agree, and we use the standard deviation to discover the distribution state. We use the **variance and the standard deviation** as **statistics to show the state of distribution of data**.

EXERCISE

◆ The data below shows the math test scores of 10 students. Find the average, variance, and standard deviation of this data.

67, 73, 54, 81, 47, 59, 62, 86, 49, 71 (units: points)

											Total
Score (points)	67	73	54	81	47	59	62	86	49	71	649
Deviation (points)	2.1	8.1	−10.9	16.1	−17.9	−5.9	−2.9	21.1	−15.9	6.1	0
Deviation squared	4.41	65.61	118.81	259.21	320.41	34.81	8.41	445.21	252.81	37.21	1546.9

Average $\bar{x} = \frac{649}{10} = 64.9$ (points)

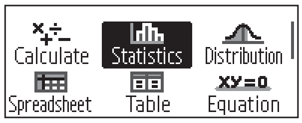
Variance $s^2 = \frac{1546.9}{10} = 154.69$

Standard deviation $s = \sqrt{154.69} \simeq 12.44$ (points)

Average…64.9 points, variance…154.69, standard deviation…12.44 points

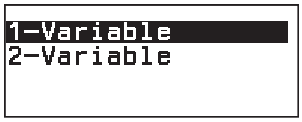
check

Press ⌂, select [Statistics], press (OK), select [1-Variable], press (OK), input data in the x column.

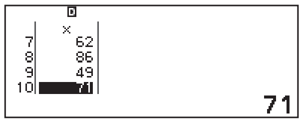


67, 73, 54, 81, 47, 59, 62, 86, 49, 71

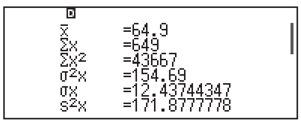
⑥ ⑦ (EXE) ⑦ ③ (EXE) ⑤ ④ (EXE) ⑧ ① (EXE)

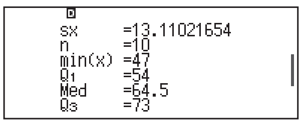


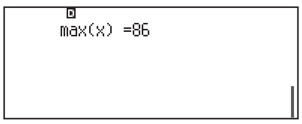
Do the same input below, and after inputting the final data, press (EXE), select [1-Var Results], press (OK)



Average (points): $\bar{x} = 64.9$ Variance: $\sigma^2 x = 154.69$ Standard deviation (points): $\sigma x = 12.44$

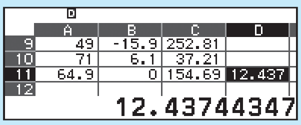






EXTRA Info.

There is also a function for doing the above calculation in a single process. (results only)



PRACTICE

1 The data below shows the science test scores of 10 students. Find the average, variance, and standard deviation of this data.

66, 76, 57, 89, 84, 63, 62, 82, 49, 92 (units: points)

											Total
Score (points)	66	76	57	89	84	63	62	82	49	92	720
Deviation (points)	−6	4	−15	17	12	−9	−10	10	−23	20	0
Deviation squared	36	16	225	289	144	81	100	100	529	400	1920

Average $\bar{x} = \frac{720}{10} = 72$ (points)

Variance $s^2 = \frac{1920}{10} = 192$

Standard deviation $s = \sqrt{192} \simeq 13.9$ (points)

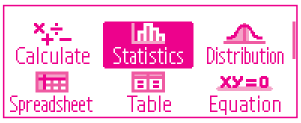
Average…72 points, variance…192, standard deviation…13.9 points

check

Press ⌂, select [Statistics], press OK, select [1-Variable], press OK, input data in the x column.

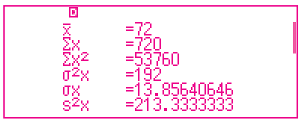
⑥ ⑥ EXE ⑦ ⑥ EXE ⑤ ⑦ EXE ⑧ ⑨ EXE

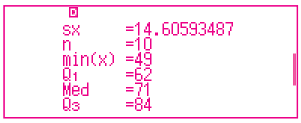
Do the same input below, and after inputting the final data, press EXE, select [1-Var Results], press OK

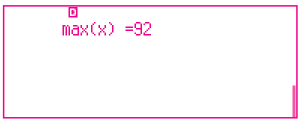


Average (points): $\bar{x} = 72$ **Variance:** $\sigma^2_x = 192$

Standard deviation (points): $\sigma_x = 13.9$







ADVANCED

PRACTICE

[2] The data below shows the records of throwing a ball for group A and group B. Solve by finding the average, variance, and standard deviation of the groups, as well as which group has the largest distribution.

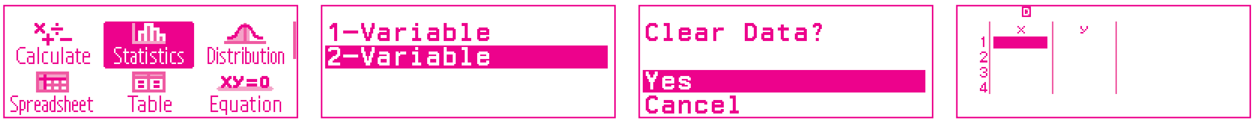
Group A	26, 18, 28, 31, 23	(units m)
Group B	20, 32, 23, 28, 14	(units m)

Group A　Average…25.2 (m), variance…19.76, standard deviation…4.45 (m)
Group B　Average…23.4 (m), variance…39.04, standard deviation…6.25 (m)
From the standard deviation, A<B, so group B has the larger distribution.

Group B

check

Press ⌂, select [Statistics], press OK, select [2-Variable], press OK, select [Yes], press OK
Press ○○○, select [Frequency], press OK, select [Off], press OK, press AC
Input group A data in the x column and group B in the y column.

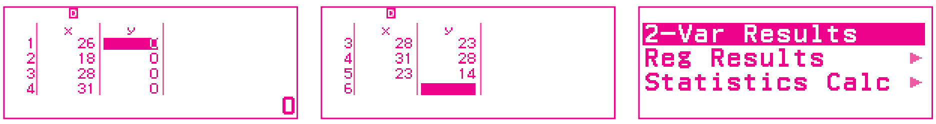


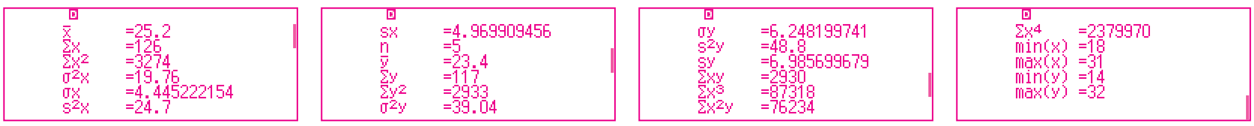
2 6 EXE 1 8 EXE 2 8 EXE 3 1 EXE 2 3 EXE

Use the cursor to move to row 1 in column y.

2 0 EXE 3 2 EXE 2 3 EXE 2 8 EXE 1 4 EXE

After inputting the final data, press EXE, select [2-Var Results], press OK





Group A　Average (m): $\bar{x}=25.2$, variance: $\sigma^2{}_x=19.76$, standard deviation (m): $\sigma_x=4.45$

Group B　Average (m): $\bar{y}=23.4$, variance: $\sigma^2{}_y=39.04$, standard deviation (m): $\sigma_y=6.25$

Data correlation

TARGET To learn about the correlation between 2 variates.

STUDY GUIDE

Correlation

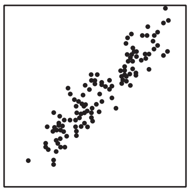
We say that 2 variates **have a correlation** when as one increases the other tends to increase or decrease linearly.

A graph that shows 2 variates as points on a plane is called a **scatter plot**, which makes the relation between 2 variates easy to grasp visually.

The following relations between 2 variates can be read from a scatter plot.

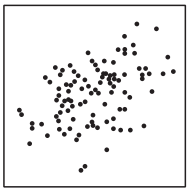
(a)	**Have a positive correlation**	As one increases, the other also increases linearly. …Up to the right
(b)	**Have a negative correlation**	As one increases, the other decreases linearly. …Down to the right
(c)	**Have no correlation**	When there is no tendency as in (a) or (b).
(d)	**Strength of correlation**	There is a correlation, which is strong when a linear slope clearly appears. If not, then it is weak.

(a) Positive correlation



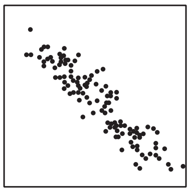
(d) Strong

(a) Positive correlation



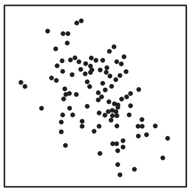
(d) Weak

(b) Negative correlation



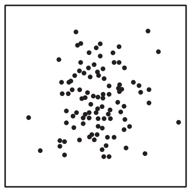
(d) Strong

(b) Negative correlation



(d) Weak

(c) No correlation



Correlation coefficient

For a number n of data for 2 variates x and y, we get $(x_1, y_1), (x_2, y_2), \cdots\cdots, (x_n, y_n)$.

Given the average of $x_1, x_2, \cdots\cdots, x_n$ is $\overline{x}$, with a standard deviation of s_x, and the average of $y_1, y_2, \cdots\cdots, y_n$ is $\overline{y}$, with a standard deviation of s_y, then we say that

$$s_{xy} = \frac{(x_1-\overline{x})(y_1-\overline{y})+(x_2-\overline{x})(y_2-\overline{y})+\cdots\cdots+(x_n-\overline{x})(y_n-\overline{y})}{n}$$

is the **covariance** of variate x and variate y.

When the covariance is positive, **there is a positive correlation** between variate x and variate y.

When the covariance is negative, **there is a negative correlation** between variate x and variate y.

The following correlation coefficient is considered a guide indicating direction and strength of correlations.

Correlation coefficient $r=\dfrac{s_{xy}}{s_x s_y}$

$$s_x=\sqrt{\frac{(x_1-\overline{x})^2+(x_2-\overline{x})^2+\cdots\cdots+(x_n-\overline{x})^2}{n}}$$

$$s_y=\sqrt{\frac{(y_1-\overline{y})^2+(y_2-\overline{y})^2+\cdots\cdots+(y_n-\overline{y})^2}{n}}$$

$$s_{xy}=\frac{(x_1-\overline{x})(y_1-\overline{y})+(x_2-\overline{x})(y_2-\overline{y})+\cdots\cdots+(x_n-\overline{x})(y_n-\overline{y})}{n}$$

<Properties of correlation coefficients>

(1) $-1\leq r\leq 1$

(2) The closer the value of r is to 1, the stronger the positive correlation.

The closer the value of r is to -1, the stronger the negative correlation.

The closer the value of r is to 0, the weaker the correlation.

EXERCISE

◆ The figure on the right shows the scores of the science tests (x points) and math tests (y points) of 5 students. Find the correlation coefficient r for x and y rounded down to the 2nd decimal place. Furthermore, determine if x and y have a correlation and whether it is positive or negative and strong or weak.

Student	x	y
A	6	4
B	3	3
C	5	4
D	9	8
E	7	6

The average values of x and y are $\overline{x}$ and $\overline{y}$, the deviations are $x-\overline{x}$ and $y-\overline{y}$, the product of the deviations is $(x-\overline{x})(y-\overline{y})$, and the squares of the deviations are $(x-\overline{x})^2$ and $(y-\overline{y})^2$, which are calculated in order and summarized in the following table.

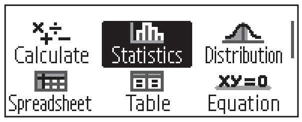
Student	x	y	$x-\overline{x}$	$y-\overline{y}$	$(x-\overline{x})(y-\overline{y})$	$(x-\overline{x})^2$	$(y-\overline{y})^2$
A	6	4	0	-1	0	0	1
B	3	3	-3	-2	6	9	4
C	5	4	-1	-1	1	1	1
D	9	8	3	3	9	9	9
E	7	6	1	1	1	1	1
Total	30	25	0	0	17	20	16
Average	6	5	0	0	3.4	4	3.2

From $s_{xy}=3.4$, $s_x=\sqrt{4}=2$, $s_y=\sqrt{3.2}$, we get $r=\dfrac{s_{xy}}{s_x s_y}=\dfrac{3.4}{2\times\sqrt{3.2}}=\dfrac{17\sqrt{5}}{40}\simeq 0.95$

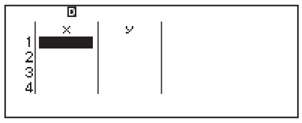
The correlation coefficient is $r=0.95$, and x and y have a strong positive correlation.

check

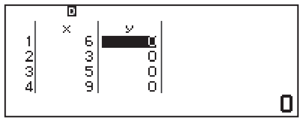
Press (HOME), select [Statistics], press (OK), select [2-Variable], press (OK)



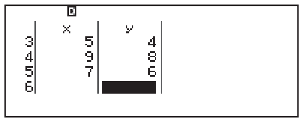
Input in the x column. (6) (EXE) (3) (EXE) (5) (EXE) (9) (EXE) (7) (EXE)



Return to row 1 in the y column. (>) (∨)

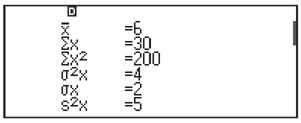


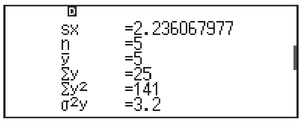
Input in the y column. (4) (EXE) (3) (EXE) (4) (EXE) (8) (EXE) (6) (EXE)

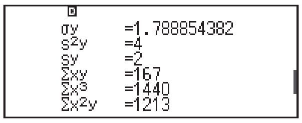


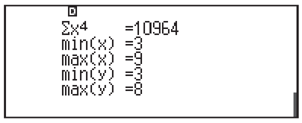
After inputting the final data, press (EXE), select [2-Var Results], press (OK)

Standard deviation of x: $\sigma_x = 2$ Standard deviation of y: $\sigma_y = 1.8$









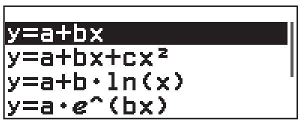
Since covariance is not shown in the above table, it is found by (covariance) = (average of product) − (product of averages).

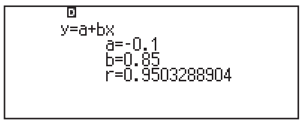
$$s_{xy} = \frac{\Sigma xy}{n} - \overline{x}\overline{y} = \frac{167}{5} - 6 \cdot 5 = 3.4$$

[2-Var Results] calculates the various values in the above tables.

[Reg Results] calculates a regression curve when there is a correlation in the value of the correlation coefficient and the scatter plot of the data.

Press (BACK), press (EXE), select [Reg Results], press (OK), select [y=a+bx], press (OK)





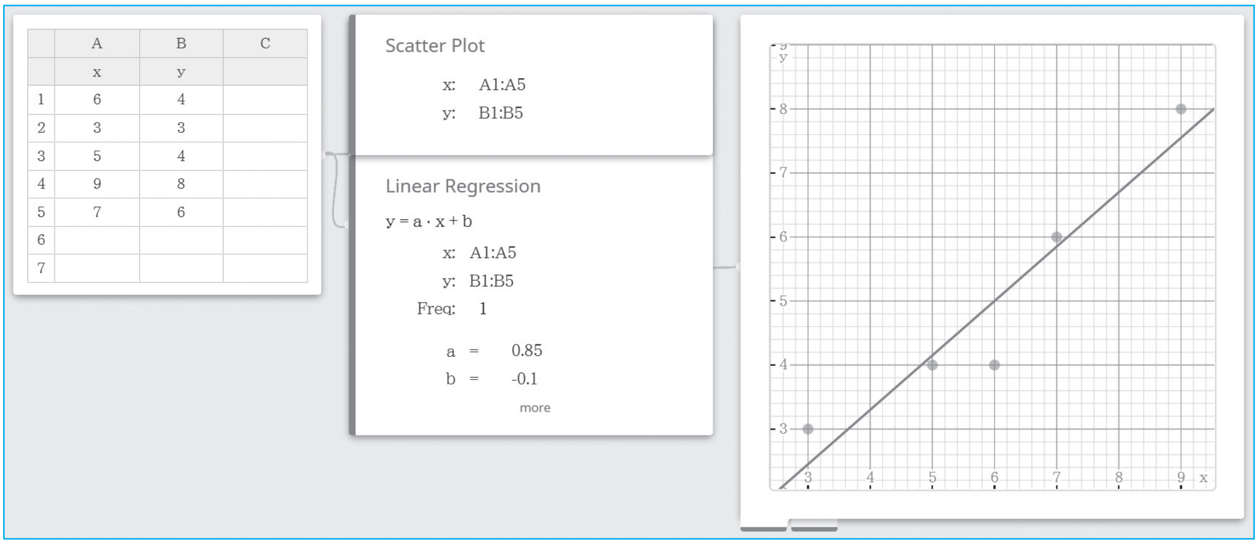
EXTRA Info.

Scatter plots and correlation

Select the QR code function.

Read the QR code with your smartphone.

A line appears showing the correlation calculated from the XY scatter plot, data table, and regression calculation.



PRACTICE

◆ The table on the right is the data for 2 variates x and y. Draw a scatter plot and find the correlation coefficient. Furthermore, determine if x and y have a correlation and whether it is positive or negative and strong or weak.

	x	y
A	3	4
B	8	7
C	12	11
D	21	14

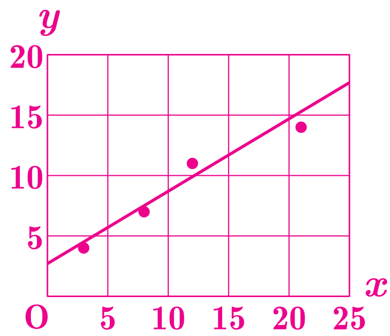
The average values of x and y are $\overline{x}$ and $\overline{y}$, the deviations are $x-\overline{x}$ and $y-\overline{y}$, the product of the deviations is $(x-\overline{x})(y-\overline{y})$, and the squares of the deviations are $(x-\overline{x})^2$ and $(y-\overline{y})^2$, which are calculated in order and summarized in the following table.

	x	y	$x-\overline{x}$	$y-\overline{y}$	$(x-\overline{x})(y-\overline{y})$	$(x-\overline{x})^2$	$(y-\overline{y})^2$
A	3	4	−8	−5	40	64	25
B	8	7	−3	−2	6	9	4
C	12	11	1	2	2	1	4
D	21	14	10	5	50	100	25
Total	44	36	0	0	98	174	58
Average	11	9	0	0	24.5	43.5	14.5

From $s_{xy}=24.5$, $s_x=\sqrt{43.5}\simeq 6.6$, $s_y=\sqrt{14.5}\simeq 3.8$, **we get**

$$r=\frac{s_{xy}}{s_x s_y}=\frac{24.5}{6.6\times 3.8}\simeq 0.98$$

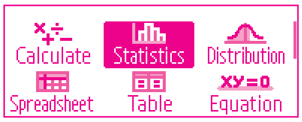
The scatter plot is shown on the right.



The correlation coefficient is $r=0.98$, and x and y have a strong positive correlation.

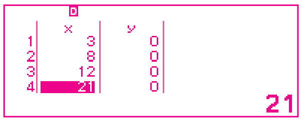
check

Press ⌂, select [Statistics], press OK, select [2-Variable], press OK



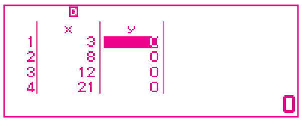
Input in the x column.

③ EXE ⑧ EXE ① ② EXE ② ① EXE



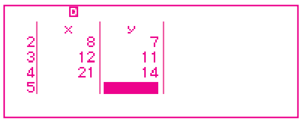
Return to row 1 in the y column.

⌄ ›



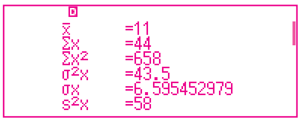
Input in the y column.

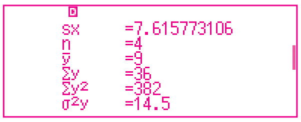
④ EXE ⑦ EXE ① ① EXE ① ④ EXE

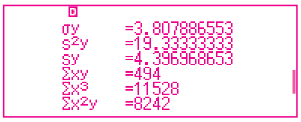


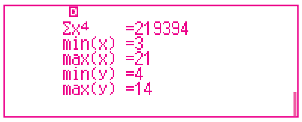
After inputting the final data, press EXE, select [2-Var Results], press OK

Standard deviation of x: $\sigma_x = 6.6$ **Standard deviation of y: $\sigma_y = 3.8$**







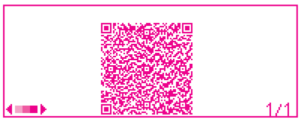


Since covariance is not shown in the above table, it is found by (covariance) = (average of product) − (product of averages).

$$s_{xy} = \frac{\Sigma xy}{n} - \bar{x}\bar{y} = \frac{494}{4} - 11 \cdot 9 = 24.5$$

EXTRA Info.

Select the QR code function.



Read the QR code with your smartphone.

	A	B	C
	x	y	
1	3	4	
2	8	7	
3	12	11	
4	21	14	
5			
6			

Scatter Plot

x: A1:A4
y: B1:B4

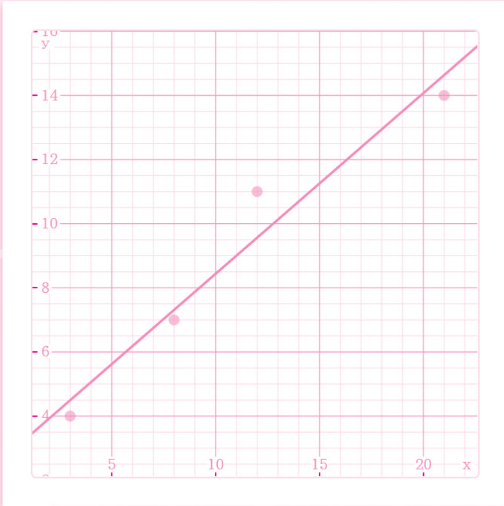
Linear Regression

y = a · x + b

x: A1:A4
y: B1:B4
Freq: 1

a = 0.5632183908
b = 2.804597701

more



CASIO Essential Materials

Publisher: CASIO Institute for Educational Development

Date of Publication: 2023/12/22 (1st edition)

https://edu.casio.com