

# Recognising the COVID-19 Pandemic from data

---

Senior Secondary Mathematics (Compulsory Part):  
Uses of Statistics

Curriculum Development Institute, Education Bureau

# Recognising the local epidemic



If we wish to have a deeper understanding of the local epidemic of COVID-19 and explore the infection of virus by age groups, how should we proceed?

Before finding relevant data for analysis, students are reminded that

- Data should be sought from **reliable sources** (e.g. websites of relevant Government Departments)
- Analyse data with **a careful and rigorous attitude** to avoid subjective and false conclusions

# Recognising the local epidemic



If we wish to have a deeper understanding of the local epidemic of COVID-19 and explore the infection of virus by age groups, how should we proceed?

## Reliable data source:

- For example, the data of government departments. The government's "DATA.GOV.HK" webpage provides data on COVID-19 from the Department of Health (in addition to accurate and reliable data, the webpage provides CSV files, which can be operated with spreadsheet software)

<https://data.gov.hk/tc-data/dataset/hk-dh-chpsebceddr-novel-infectious-agent>

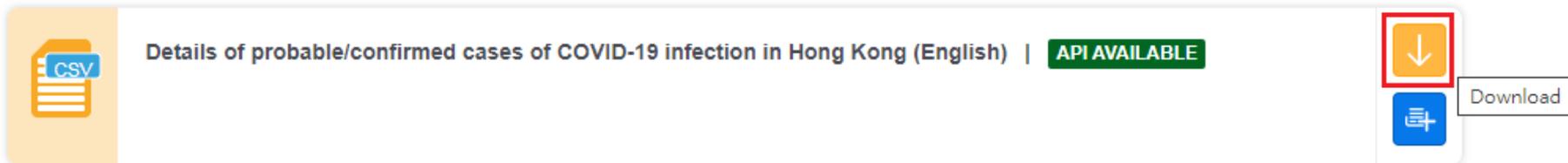
## Activity 1: Analysing the ages of confirmed patients using IT

Firstly download the data "Details of suspected/confirmed 2019 coronavirus cases in Hong Kong" from the "Data in Coronavirus Disease (COVID-19)" of the website "DATA.GOV.HK". In this activity, students will use a spreadsheet to analyse the data and complete the following table.

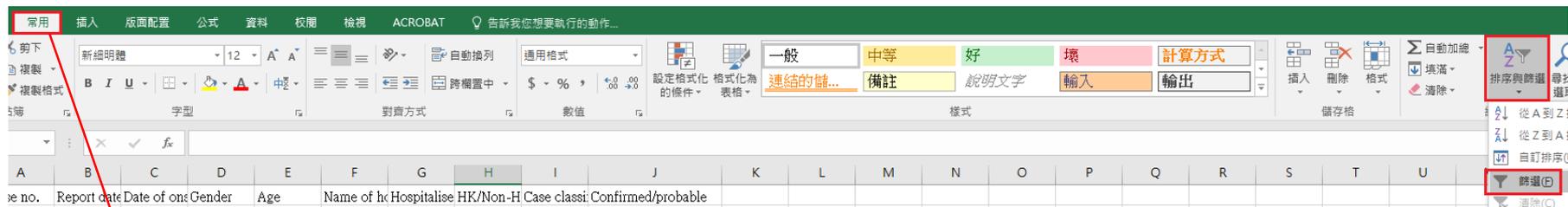
Confirmed cases		Male	Female	All
Number				
Age	Range			
	Mean			
	Mode			
	Median			
	Upper quartile			
	Lower quartile			

The following steps demonstrate how to analyse the data of Male's confirmed cases.

Step 1: Download the latest data from the "Data in Coronavirus Disease (COVID-19)" of the website "DATA.GOV.HK".



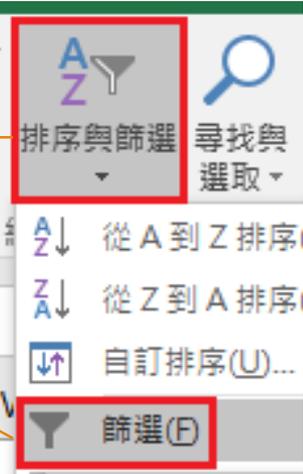
Step 2: Select "HOME" → "Sort & Filter" → "Filter"



HOME

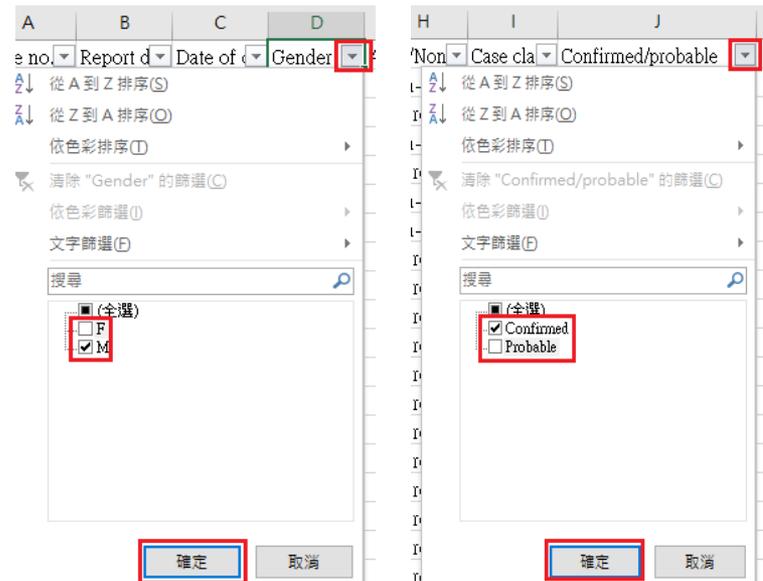
Sort & Filter

Filter



Step 3: Click the arrow button next to “Gender”, choose “M” only and click “confirm”. Then, click the arrow button next to “Confirmed/probable” and choose “Confirmed” and click “confirm”. The results are the confirmed cases for male only. It shows at the bottom of the table that there are 599 records, i.e. the number of the confirmed cases for male is 599 (use data as at 16 June 2020).

	A	B	C	D	E
1	Case no.	Report d	Date of c	Gender	Age
2	1	23/01/2020	21/01/2020	M	39
3	2	23/01/2020	18/01/2020	M	56
6	5	24/01/2020	23/01/2020	M	63
7	6	26/01/2020	21/01/2020	M	47
9	8	26/01/2020	25/01/2020	M	64
11	10	29/01/2020	25/01/2020	M	72
13	12	30/01/2020	22/01/2020	M	75
14	13	31/01/2020	29/01/2020	M	39
15	14	01/02/2020	23/01/2020	M	80
18	17	04/02/2020	22/01/2020	M	60



confirm

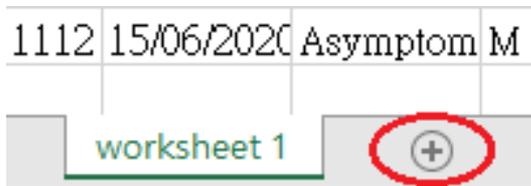
worksheet 1  
從 1113 中找出 599 筆記錄

599 records

Step 4: Use the mouse to click the first record of “Age” (i.e. cell E2), then press the three buttons “Shift”, “Ctrl” and down arrow together to select all the records of age for all male confirmed cases. Then, press “Ctrl” and “C” to copy these records.

	A	B	C	D	E
1	Case no. ▾	Report d ▾	Date of c ▾	Gender ▾	Age ▾
2		1 23/01/2020	21/01/2020	M	39
3		2 23/01/2020	18/01/2020	M	56

Step 5: Click the “plus” sign next to “worksheet 1” to create a new worksheet.



Step 6: Click cell A1 of “worksheet 2” and press “Ctrl” and “V” to paste all records of male confirmed cases under Column A of “worksheet 2” (as shown on the right).

Then, we can use different functions to find the values on the table below.

Confirmed cases		Male
Number		599
Age	Range	
	Average	
	Mode	
	Median	
	Upper Quartile	
	Lower Quartile	

	A	B	C	D
1	39			
2	56			
3	63			
4	47			
5	64			
6	72			
7	75			
8	39			
9	80			
10	60			
11	25			
12	56			
13	58			
14	42			
15	24			
16	70			
17	68			
18	22			
19	23			
20	69			
21	52			
22	59			
23	71			
24	75			
25	66			
26	37			
27	51			
28	43			
29	37			
30	70			
31	54			
32	69			
33	45			
34	58			
35	68			

Step 7a: Input related information under Column C and enter different formulae in the relevant cells under Column D.

✓ fx =max(A1:A599)	
C	D
Maximum	=max(A1:A599)
Minimum	
Range	
Average	
Mode	
Median	
Upper Quartile	
Lower Quartile	

✓ fx =min(A1:A599)	
C	D
Maximum	93
Minimum	=min(A1:A599)
Range	
Average	
Mode	
Median	
Upper Quartile	
Lower Quartile	

✓ fx =D1-D2	
C	D
Maximum	93
Minimum	0
Range	=D1-D2
Average	
Mode	
Median	
Upper Quartile	
Lower Quartile	

As this activity involves a large number of data, it is suggested to use information technology to calculate the statistics. Students can then focus on analysing the results.

Step 7b: Input related information under Column C and enter different formulae in the relevant cells under Column D.

✓	<i>f<sub>x</sub></i>	=average(A1:A599)
C		D
Maximum		93
Minimum		0
Range		93
Average		=average(A1:A599)
Mode		
Median		
Upper Quartile		
Lower Quartile		

✓	<i>f<sub>x</sub></i>	=mode(A1:A599)
C		D
Maximum		93
Minimum		0
Range		93
Average		37.13522538
Mode		=mode(A1:A599)
Median		
Upper Quartile		
Lower Quartile		

✓	<i>f<sub>x</sub></i>	=median(A1:A599)
C		D
Maximum		93
Minimum		0
Range		93
Average		37.13522538
Mode		20
Median		=median(A1:A599)
Upper Quartile		
Lower Quartile		

Step 7c: Input related information under Column C and enter different formulae in the relevant cells under Column D.

✓ fx		=quartile.inc(A1:A599,3)
C	D	
Maximum	93	
Minimum	0	
Range	93	
Average	37.13522538	
Mode	20	
Median	35	
Upper Quartile	=quartile.inc(A1:A599,3)	
Lower Quartile		

✓ fx		=quartile(A1:A599,1)
C	D	
Maximum	93	
Minimum	0	
Range	93	
Average	37.13522538	
Mode	20	
Median	35	
Upper Quartile	51	
Lower Quartile	=quartile(A1:A599,1)	

C	D
Maximum	93
Minimum	0
Range	93
Average	37.13522538
Mode	20
Median	35
Upper Quartile	51
Lower Quartile	22

## Activity 1: Analysing the ages of confirmed patients using IT

- a) Refer to the above steps, try to get the relevant values for female and all cases and complete the following table.

Confirmed cases		Male	Female	All
Number		599		
Age	Range	93		
	Mean	37.1		
	Mode	20		
	Median	35		
	Upper quartile	51		
	Lower quartile	22		

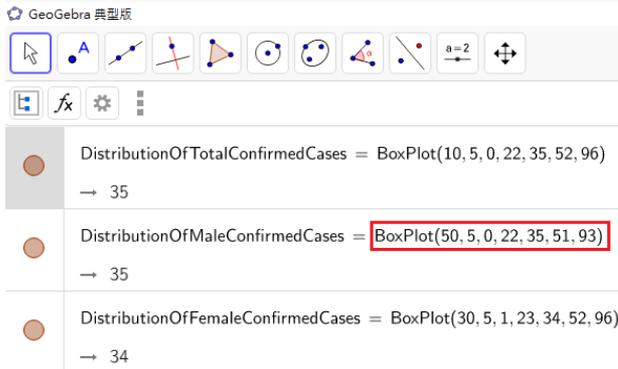
- b) Represent the ages of male, female, and all patients by boxplots and compare their distributions.

Suggested answers as follows (using data as at 16 June 2020)

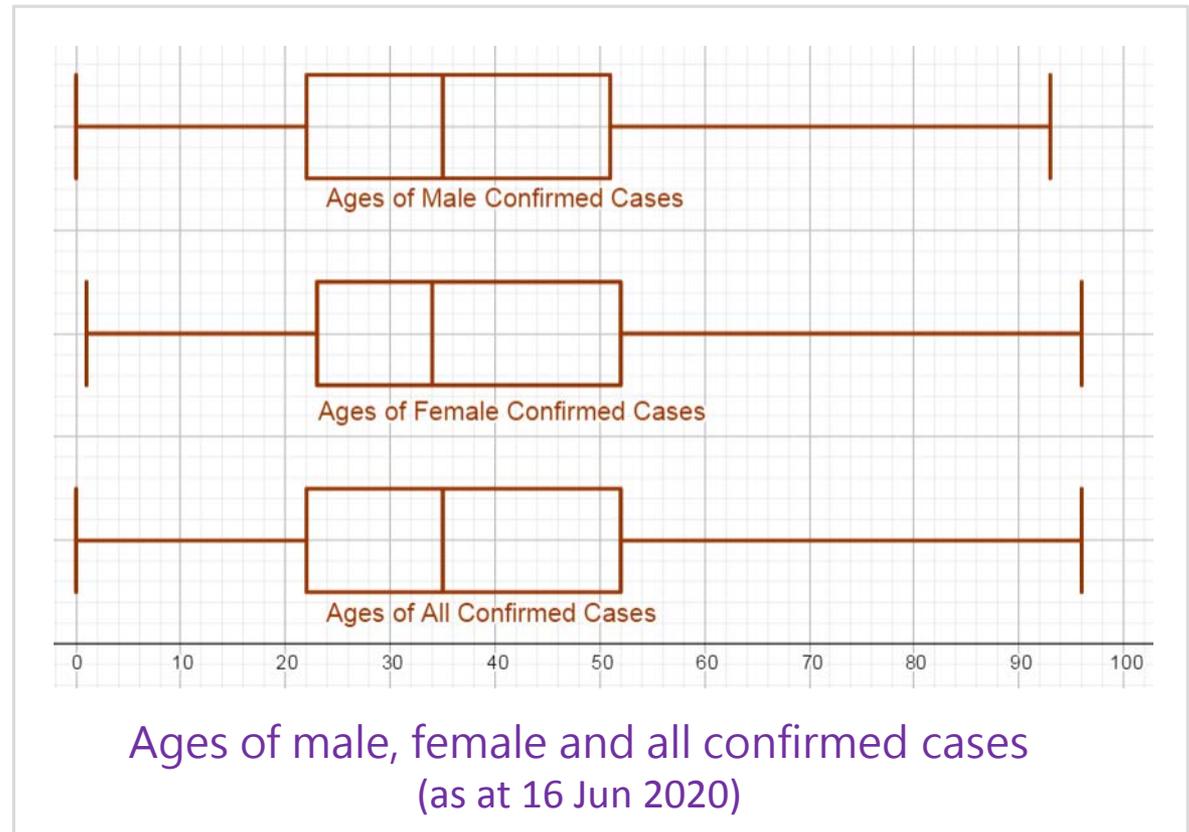
Confirmed cases		Male	Female	All
Number		599	513	1112
Age	Range	93	95	96
	Mean*	37.1	38.2	37.6
	Mode	20	20	20
	Median	35	34	35
	Upper quartile	51	52	52
	Lower quartile	22	23	22

\*correct to three significant figures

We can use the software “GeoGebra” and input the following command to draw the required boxplots.



The command of BoxPlot is “BoxPlot(50, 5, 0, 22, 35, 51, 93)”, the meanings, in sequence, of these seven numbers are: the position of the boxplot, the width of the boxplot, minimum values, lower quartile, median, upper quartile and maximum value.



By observing the boxplots, we find that the distributions of ages of male, female and all cases are very similar.

## Activity 2: Comparing the cases in different age groups

### Observe and guess

By observing the above table and boxplots, some students find that no matter whether it is male, female or all cases, 25% of the data are concentrated between the twenties and the thirties and the modal age of three groups are all 20. Do these results indicate that the age group of near 20 years old is more commonly infected with COVID-19 than other age groups?

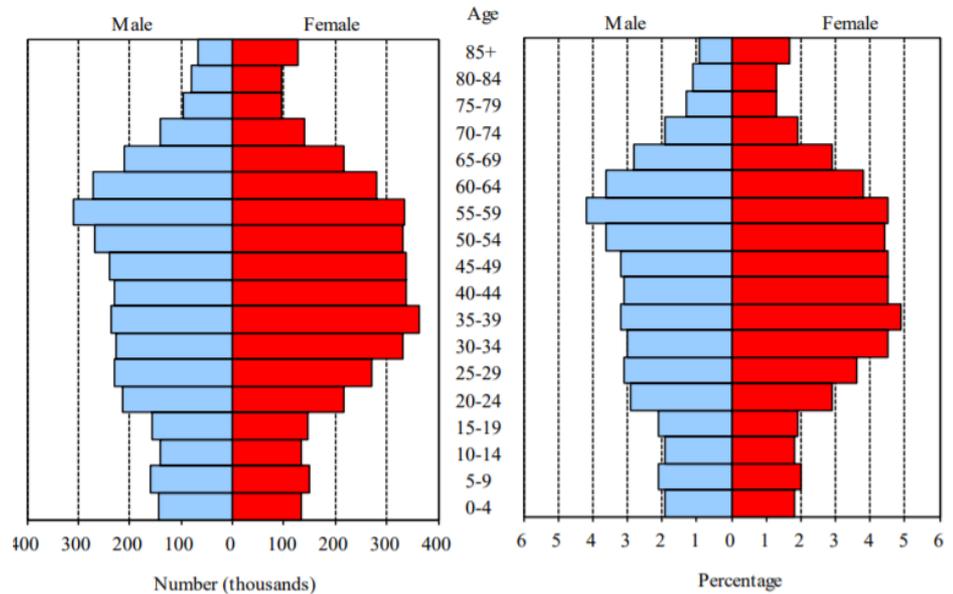


## Activity 2: Comparing the cases in different age groups

### Analysis

To know whether the guess is correct, we may first notice that numbers of people in different age groups are different. So, we need that data in order to further analyse which age group of people is more commonly infected with COVID-19.

Population of Hong Kong by age and sex, mid-2018



Sources: “Census and Statistics Department: Living with Statistics”  
<https://www.censtatd.gov.hk/hkstat/sub/sc460.jsp?productCode=B8XX0025>

## Activity 2: Comparing the cases in different age groups

- a) Complete the following table by the data “Details of suspected/confirmed 2019 coronavirus cases in Hong Kong“ used in Activity 1.

Age group	0-14	15-24	25-34	35-44	45-54	55-64	65+	Total number
Male								
Female								
All								

## Activity 2: Comparing the cases in different age groups

Age distribution of confirmed cases being infected with COVID-19  
in Hong Kong (use data as at 16 June 2020)

Age group	0-14	15-24	25-34	35-44	45-54	55-64	65+	Total number
Male	24	165	105	111	63	88	43	599
Female	17	128	114	76	56	76	46	513
All	41	293	219	187	119	164	89	1112

Do you know how to get the above figures using a spreadsheet?

## Activity 2: Comparing the cases in different age groups

- b) Try to get the "population by age group" data from the Census and Statistics Department website, analyse which age group of people are more commonly infected with COVID-19.

Still remember getting data from reliable sources?

Census and Statistics Department:

[https://www.censtatd.gov.hk/hkstat/hkif/index\\_tc.jsp](https://www.censtatd.gov.hk/hkstat/hkif/index_tc.jsp)

	Age Group	0-14	15-24	25-34	35-44	45-54	55-64	65+
Male	Confirmed	24	165	105	111	63	88	43
	Total number	450 700	350 600	454 000	465 200	497 000	589 400	616 100
	Percentage	0.005%	0.047%	0.023%	0.024%	0.013%	0.015%	0.007%
Female	Confirmed	17	128	114	76	56	76	46
	Total number	424 200	348 500	594 800	712 600	665 100	633 300	705 900
	Percentage	0.004%	0.037%	0.019%	0.011%	0.008%	0.012%	0.007%
All	Confirmed	41	293	219	187	119	164	89
	Total number	874 900	699 100	1 048 800	1 177 800	1 162 100	1 222 700	1 322 000
	Percentage	0.005%	0.042%	0.021%	0.016%	0.010%	0.013%	0.007%

From the results of spreadsheet, we know that the percentages of number of confirmed cases to the total population in the age groups range from 0.004% to 0.047%. Among the age groups, the 15-24 age group has the highest percentage of confirmed cases for both male and female cases, while the 0-14 age group has the lowest.

## Activity 2: Comparing the cases in different age groups

- c) [Extended exploration] The percentage of confirmed cases in the 15-24 age group is relatively high. Is it because the people in this age group have weaker immunity to the coronavirus? Or do we have to consider whether the cases are local or imported? What extra data do we need to know? Try to explore and illustrate the conclusions by tables/charts.



END