

# 從數據認識 2019冠狀病毒病疫情

---

高中數學科必修部分：統計的應用

教育局課程發展處

# 認識本地疫情



我們若希望對本地新型冠狀病毒病疫情有更深入的認識，探討各年齡組別感染病毒的情況，我們該如何著手？

在開始尋找相關數據和作出分析前，提醒同學們

- 應從可靠來源尋找數據（例如相關政府部門網頁）
- 抱持小心、嚴謹的態度分析數據，避免主觀、失實的結論

# 認識本地疫情



我們若希望對本地新型冠狀病毒病疫情有更深入的認識，探討各年齡組別感染病毒的情況，我們該如何著手？

可靠的數據的來源：

- 例如政府部門的數據，政府的「資料一線通」網頁提供衛生署的2019冠狀病毒病數據（除數據準確可靠外，亦提供CSV檔，可利用試算表軟件開啟分析）。

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

## 活動一：利用資訊科技分析確診者年齡

先從上頁「資料一線通: 2019冠狀病毒病的數據」網址下載最新「本港疑似/確診2019冠狀病毒的個案詳情」的數據。在這活動同學們將使用試算表分析數據並完成下表。

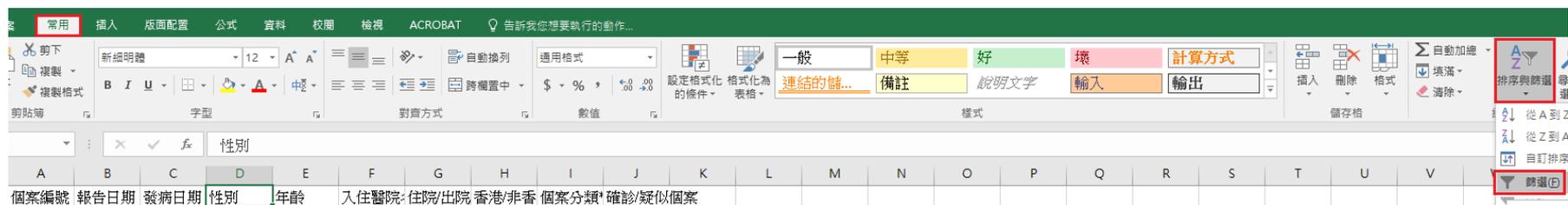
確診個案		男性	女性	整體
人數				
年齡	分佈域			
	平均數			
	眾數			
	中位數			
	上四分位數			
	下四分位數			

以下步驟示範如何整理有關男性確診者的數據。

步驟1：從「資料一線通: 2019冠狀病毒病的數據」的網址下載最新數據。



步驟2：從「常用」→ 選「排序與篩選」→ 選「篩選」



步驟3：先按下「性別」旁的箭號，只選「男」並按「確定」，然後按下「確診/疑似個案」旁的箭號，只選「確診」並按「確定」，可得出只包括男性確診者的資料。工作表下方顯示有「599筆記錄」，即男性確診人數為599。(截至2020年6月16日的數據)

	A	B	C	D	E
1	個案編號	報告日	發病日	性別	年齡
2	1	23/01/2020	21/01/2020	男	39
3	2	23/01/2020	18/01/2020	男	56
6	5	24/01/2020	23/01/2020	男	63
7	6	26/01/2020	21/01/2020	男	47
9	8	26/01/2020	25/01/2020	男	64
11	10	29/01/2020	25/01/2020	男	72
13	12	30/01/2020	22/01/2020	男	75
14	13	31/01/2020	29/01/2020	男	39
15	14	01/02/2020	23/01/2020	男	80
18	17	04/02/2020	22/01/2020	男	60

此圖顯示 Excel 篩選器中的「性別」欄位。選項包括「全選」、「女」和「男」。其中「男」選項已被勾選，並用紅框標出。下方「確定」按鈕亦被紅框標出。

此圖顯示 Excel 篩選器中的「香港/非」欄位下的「個案分類」子欄位。選項包括「全選」、「疑似」和「確診」。其中「確診」選項已被勾選，並用紅框標出。下方「確定」按鈕亦被紅框標出。

工作表1  
從 1113 中找出 599 筆記錄

步驟4：用滑鼠點選第一項年齡資料（即 E2 儲存格），然後同時按下三個鍵「Shift」、「Ctrl」及下箭咀，即可選取所有男性確診者的年齡數據，按下「Ctrl」及「C」把資料複製。

	A	B	C	D	E
1	個案編號	報告日	發病日	性別	年齡
2	1	23/01/2020	21/01/2020	男	39
3	2	23/01/2020	18/01/2020	男	56

步驟5：按下「工作表1」旁的「加號」以建立新的工作表。

1110	13/06/2020	04/06/2020	男	33
1112	15/06/2020	無病徵	男	10

工作表1 

步驟6：點選「工作表2」的 A1 儲存格，然後按下「Ctrl」及「V」把全部男性確診者的年齡數據複製到「工作表2」的欄A（如右圖所示）。

然後，我們會利用不同函數以求得下表各數值。

確診個案		男性
人數		599
年齡	分佈域	
	平均數	
	眾數	
	中位數	
	上四分位數	
	下四分位數	

The screenshot shows an Excel spreadsheet with three columns labeled A, B, and C. Column A contains a list of 35 age values. The values are: 39, 56, 63, 47, 64, 72, 75, 39, 80, 60, 25, 56, 58, 42, 24, 70, 68, 22, 23, 69, 52, 59, 71, 75, 66, 37, 51, 43, 37, 70, 54, 69, 45, 58, 68. The spreadsheet has a tab labeled '工作表2' at the bottom right, which is highlighted with a red box.

	A	B	C
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		

步驟7a：在欄C輸入相關資料，在欄D相應儲存格輸入不同公式。

✓ fx		=max(A1:A599)
C	D	
最大值	=max(A1:A599)	
最小值		
分佈域		
平均數		
眾數		
中位數		
上四分位數		
下四分位數		

✓ fx		=min(A1:A599)
C	D	
最大值	93	
最小值	=min(A1:A599)	
分佈域		
平均數		
眾數		
中位數		
上四分位數		
下四分位數		

✓ fx		=D1-D2
C	D	
最大值	93	
最小值	0	
分佈域	=D1-D2	
平均數		
眾數		
中位數		
上四分位數		
下四分位數		

由於這活動涉及大量數據，因此建議運用資訊科技計算各統計量，同學們可集中分析結果。

步驟7b：在欄C輸入相關資料，在欄D相應儲存格輸入不同公式。

✓ f <sub>x</sub>	=average(A1:A599)	
	C	D
最大值		93
最小值		0
分佈域		93
平均數	=average(A1:A599)	
眾數		
中位數		
上四分位數		
下四分位數		

✓ f <sub>x</sub>	=mode(A1:A599)	
	C	D
最大值		93
最小值		0
分佈域		93
平均數		37.13522538
眾數	=mode(A1:A599)	
中位數		
上四分位數		
下四分位數		

✓ f <sub>x</sub>	=median(A1:A599)	
	C	D
最大值		93
最小值		0
分佈域		93
平均數		37.13522538
眾數		20
中位數	=median(A1:A599)	
上四分位數		
下四分位數		

步驟7c：在欄C 輸入相關資料，在欄D 相應儲存格輸入不同公式。

✓ f <sub>x</sub>	=quartile.inc(A1:A599,3)
C	D
最大值	93
最小值	0
分佈域	93
平均數	37.13522538
眾數	20
中位數	35
上四分位數	=quartile.inc(A1:A599,3)
下四分位數	

✓ f <sub>x</sub>	=quartile.inc(A1:A599,1)
C	D
最大值	93
最小值	0
分佈域	93
平均數	37.13522538
眾數	20
中位數	35
上四分位數	51
下四分位數	=quartile.inc(A1:A599,1)

C	D
最大值	93
最小值	0
分佈域	93
平均數	37.13522538
眾數	20
中位數	35
上四分位數	51
下四分位數	22

## 活動一：利用資訊科技分析確診者年齡

a) 試參考以上步驟，找出女性及整體的相應數據，完成下表。

確診個案		男性	女性	整體
人數		599		
年齡	分佈域	93		
	平均數	37.1		
	眾數	20		
	中位數	35		
	上四分位數	51		
	下四分位數	22		

b) 試以箱型圖分別表示男性、女性和整體確診個案的年齡。並比較三者的分佈。

參考答案如下（以截至2020年6月16日的數據計算而得）

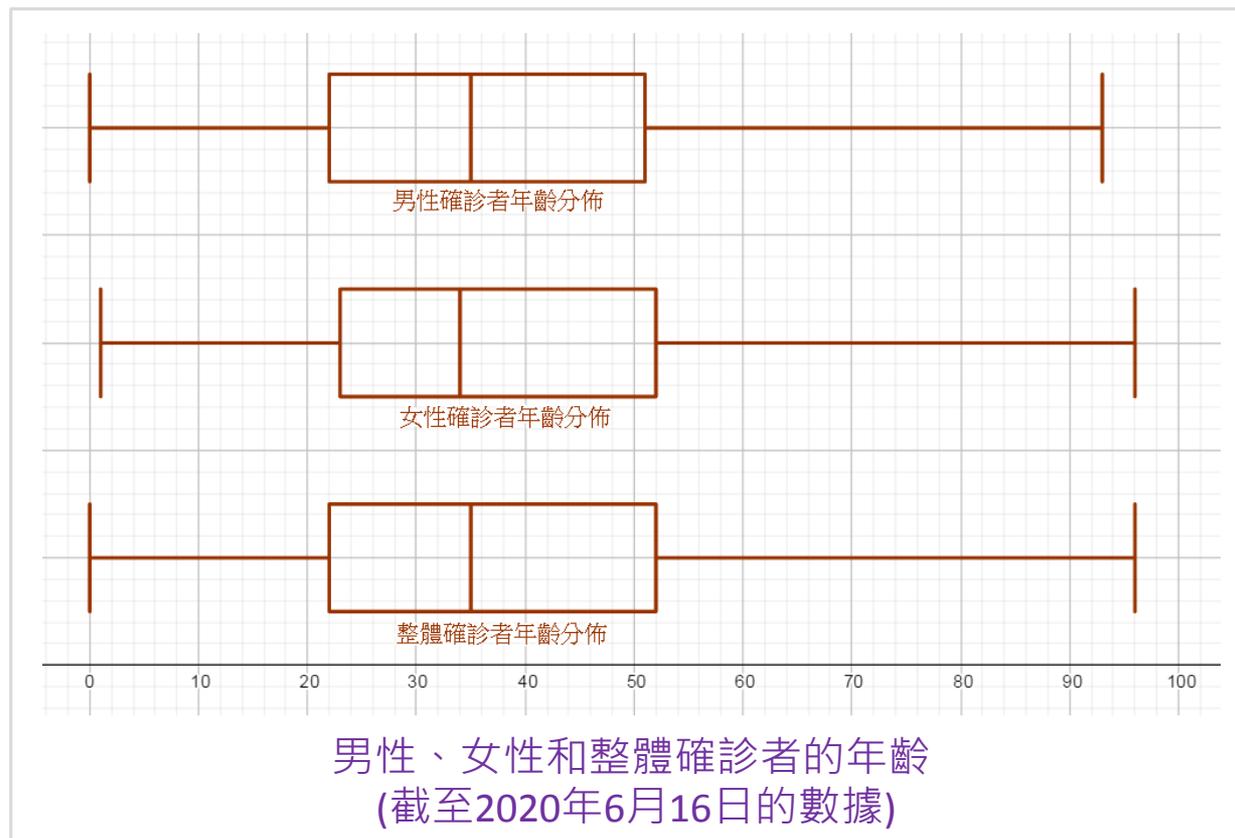
確診個案		男性	女性	整體
人數		599	513	1112
年齡	分佈域	93	95	96
	平均數*	37.1	38.2	37.6
	眾數	20	20	20
	中位數	35	34	35
	上四分位數	51	52	52
	下四分位數	22	23	22

\*準確至三位有效數字

我們可以利用 GeoGebra 軟件，輸入以下指令以繪畫所需的箱型圖。



BoxPlot 指令為「 $\text{BoxPlot}(50, 5, 0, 22, 35, 51, 93)$ 」，當中7個數字的意思依次序為：箱型圖位置的高低，箱型圖的闊度，最小值，下四分位數，中位數，上四分位數和最大值。



觀察箱型圖，我們發現男性確診者、女性確診者和整體確診者的年齡分佈非常相似。

## 活動二：比較不同年齡組別的情況

### 觀察與猜想

觀察上表和箱型圖，有同學發現不論是男性、女性或整體，有25%的數據集中於二十多歲至三十多歲之間，而年齡眾數均為二十歲。這些結果是否顯示二十多歲的年齡組別較其他年齡組別更普遍地感染2019冠狀病毒病？

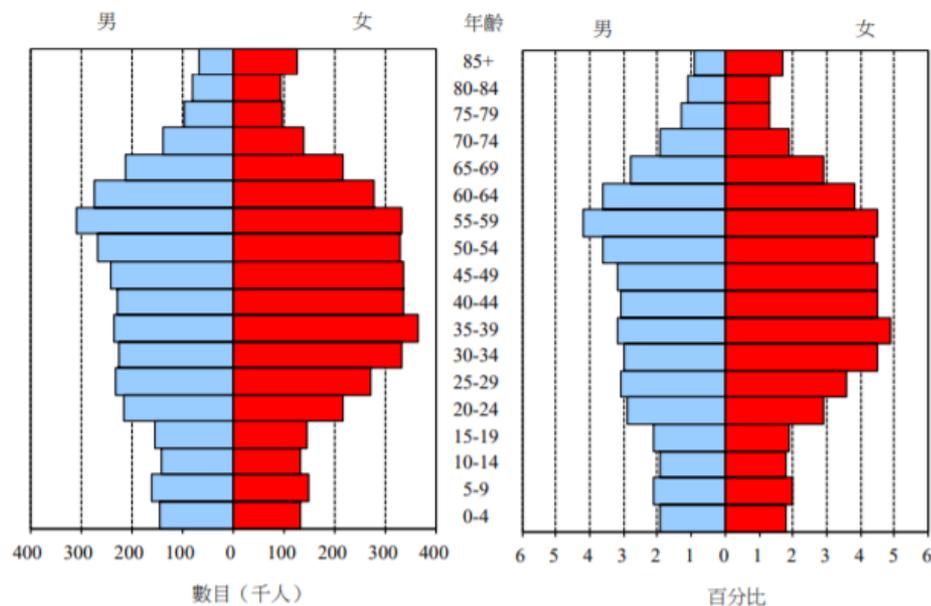


## 活動二：比較不同年齡組別的情況

### 分析

要知道猜想是否正確，我們要留意由於各年齡組別的人數本身已有差異，因此我們還需要各年齡組別總人數的數據，才可以進一步了解哪個年齡組別的人士較普遍感染2019冠狀病毒病。

2018 年年中按年齡及性別劃分的香港人口



圖表來源：「政府統計處：統計與生活」  
[https://www.censtatd.gov.hk/hkstat/sub/sc460\\_tc.jsp?productCode=B8XX0025](https://www.censtatd.gov.hk/hkstat/sub/sc460_tc.jsp?productCode=B8XX0025)

## 活動二：比較不同年齡組別的情況

- a) 利用活動一「本港疑似/確診2019冠狀病毒的個案詳情」的數據，完成下表。

年齡組別	0-14	15-24	25-34	35-44	45-54	55-64	65+	總人數
男性								
女性								
整體								

## 活動二：比較不同年齡組別的情況

本港2019冠狀病毒病確診個案的年齡分佈  
(截至2020年6月16日的數據)

年齡組別	0-14	15-24	25-34	35-44	45-54	55-64	65+	總人數
男性	24	165	105	111	63	88	43	599
女性	17	128	114	76	56	76	46	513
整體	41	293	219	187	119	164	89	1112

你懂得利用試算表求以上數據嗎？

## 活動二：比較不同年齡組別的情況

- b) 試從政府統計處網頁找出「按年齡組別劃分的人口」數據，分析哪個年齡組別的人士較普遍感染2019新型冠狀病毒病。

還記得我們要從可靠來源獲取數據嗎？

政府統計處：

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

	年齡組別	0-14	15-24	25-34	35-44	45-54	55-64	65+
男性	確診人數	24	165	105	111	63	88	43
	總人數	450 700	350 600	454 000	465 200	497 000	589 400	616 100
	百分比	0.005%	0.047%	0.023%	0.024%	0.013%	0.015%	0.007%
女性	確診人數	17	128	114	76	56	76	46
	總人數	424 200	348 500	594 800	712 600	665 100	633 300	705 900
	百分比	0.004%	0.037%	0.019%	0.011%	0.008%	0.012%	0.007%
整體	確診人數	41	293	219	187	119	164	89
	總人數	874 900	699 100	1 048 800	1 177 800	1 162 100	1 222 700	1 322 000
	百分比	0.005%	0.042%	0.021%	0.016%	0.010%	0.013%	0.007%

從試算表的計算結果得知，不同年齡組別的確診人數佔該組別總人數的百分比介乎 0.004% 至 0.047% 不等。其中，男性和女性確診百分比最高的年齡組別皆為15-24歲，而確診百分比最低的年齡組別皆為 0-14歲。

## 活動二：比較不同年齡組別的情況

- c) [延伸探究] 年齡組別15-24歲的確診人數比率較高，是否由於這年齡組別的人士對新型冠狀病毒的抵抗力較弱？還是須先考慮個案是本地感染或外地傳入？需要哪些數據才可進一步了解？試進行探究並以圖表說明所得結論。



完