

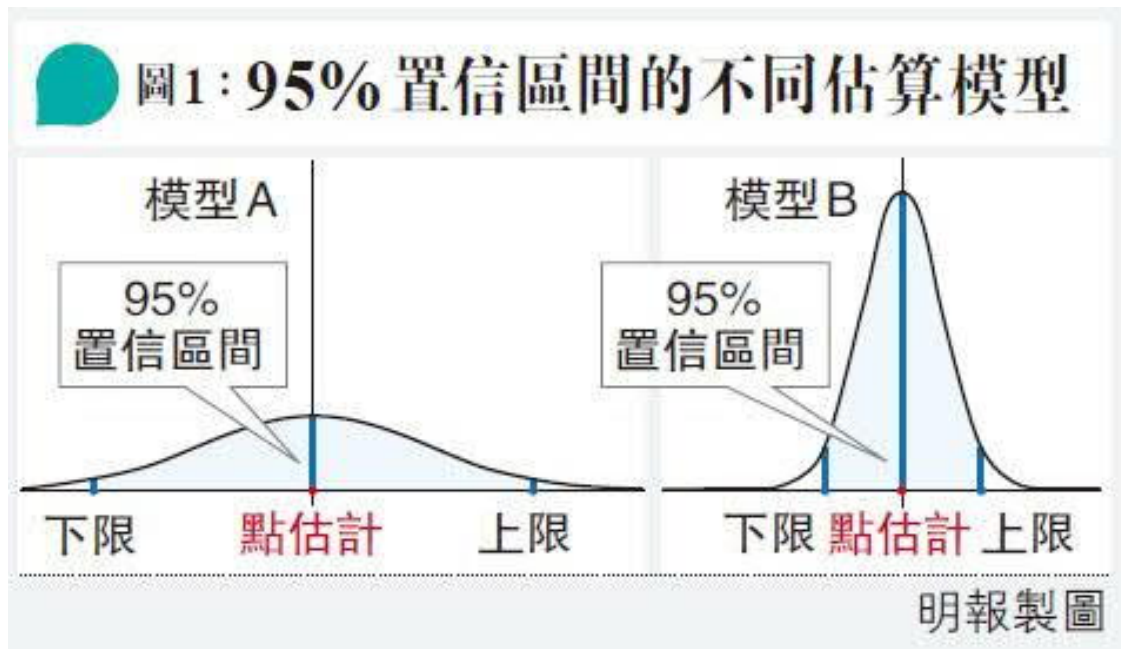
估算染疫人數 需用正確統計方法 (文：尹國聖、陳齊浩) (09:00)



【明報文章】本港疫情嚴峻，每日新增新冠病毒感染人數持續在數以萬計的水平。雖然港府極力提高檢測能力，但社區中仍然存在相當數量未被發現的病毒感染者。近日多家大學的研究團隊，發表對本地實際感染人數的估算，香港大學醫學院於 3 月 14 日估計本港實際感染人數為 358 萬，惟多數新聞報道並未強調該研究估計的 95% 置信區間為 230 萬至 460 萬人。同日，香港中文大學醫學院賽馬會公共衛生及基層醫療學院助理教授郭健安估計，本港第五波最終約有 500 萬人感染。

若置信區間過大 未能反映實況

從統計學角度看，任何統計模型都基於某些假設，如果模型假設與數據不相符，自然會導致結論的偏差。累計感染人數的預測屬於點估計（point estimate），這裏的感染人數是指所有感染新冠病毒的人（包含感染過病毒但已康復者）。我們如果未能結合置信區間（confidence interval）來分析，點估計的意義其實不大。在給定 95% 置信水平下，95% 置信區間為對於未知量的區間估計（該區間約有 95% 概率包含未知統計量）。如果統計模型得出感染人數的置信區間過大，代表我們對感染人數的估計有較高不確定性，而未能反映出感染人數的真實情況。



舉例說明：假設我們採用兩個不同模型來估計感染人數，基於「大樣本原理」，兩個估計量（estimator）都近似服從正態分佈（normal distribution），其中一個正態分佈（圖1模型A）方差較大，另外一個（圖1模型B）方差較小。顯然，模型A的95%置信區間，比模型B的要寬許多，因此具有更高的不確定性。置信區間的大小，可以直接反映估計量與實際感染數字的誤差。

預測 358 萬人感染 存相當程度誤差

港大醫學院估計的感染人數為 358 萬，其 95% 置信區間是 230 萬至 460 萬。以現時香港總人口約 741 萬來算，我們可以求出模型估計的感染比例（感染率）為「 $358 / 741 \approx 0.48$ 」，其 95% 置信區間的下限和上限，分別為「 $230 / 741 \approx 0.31$ 」和「 $460 / 741 \approx 0.62$ 」。基於 95% 置信區間的臨界值為 1.96，我們通過圖 2 公式可以推算出有效樣本量（effective sample size）約為 40 人，這相當於我們按不同地域隨機選取 40 幢樓宇，每幢樓宇隨機抽取一人檢測（以確保抽樣個體的獨立性）。所以，港大醫學院估計的感染人數，相當於僅僅基於樣本量為 40 的隨機抽樣調查結果。如此小的樣本量，有多高的可信度？

附表列出近日的圍封強檢樓宇名單，每一幢強制檢測樓宇的人數都遠遠多過 40，這些應該是香港病毒感染最高危的樓宇，其平均初步陽性比率約為 11.6%，遠遠低於港大醫學院模型推算的 48% 感染率，這反映了港大醫學院 3 月 14 日模型預測的 358 萬感染人數，存在相當程度的誤差，更何況中大醫學院郭健安所估計的 500 萬人感染？當市民在新聞看到聳人聽聞的預測感染人數時，會更加劇他們對疫情的擔憂，導致很多輕症患者動輒求醫，增加醫療體系負擔。因此，過大的置信區間，使感染人數的預測失去意義，結果會本末倒置，引起不必要的社會恐慌。

作者尹國聖是香港大學統計及精算學系系主任、教授，陳齊浩是香港大學統計及精算學系研究生

(本網發表的時事文章若提出批評，旨在指出相關制度、政策或措施存在錯誤或缺點，目的是促使矯正或消除這些錯誤或缺點，循合法途徑予以改善，絕無意圖煽動他人對政府或其他社群產生憎恨、不滿或敵意)

圖2：感染率的標準差

$$\frac{\text{上限}-\text{下限}}{2 \times \text{臨界值}} = \sqrt{\frac{\text{感染率}(1-\text{感染率})}{\text{樣本量}}}$$

明報製圖

表：近期強制檢測樓宇

明報製圖

樓宇	初步陽性個案	檢測人數	圍封時間	初步陽性比率
何文田愛民邨信民樓	138	1,100	3月12至13日	12.5%
屯門翠寧花園第三座	41	704	3月13至14日	5.8%
屯門友愛邨愛禮樓	180	1,438	3月13至14日	12.5%
黃大仙竹園南邨榮園樓	253	1,560	3月14至15日	16.2%
屯門蝴蝶邨蝶舞樓	281	1,260	3月14至15日	22.3%
深水埗富昌邨富凱樓及富怡樓	54	527	3月17至18日	10.2%
沙田隆亨邨賞心樓	120	1,320	3月17至18日	9.1%
鴨脷洲邨利怡樓	61	940	3月17至18日	6.5%
黃大仙彩雲（一）邨 白鳳樓及玉麟樓	296	1,890	3月17至18日	15.7%
屯門蝴蝶邨蝶聚樓	134	1,180	3月17至18日	11.4%
屯門湖景邨湖暉樓	136	1,350	3月17至18日	10.1%
屯門寶田邨第五座	57	827	3月17至18日	6.9%
平均初步陽性比率				11.6%

文章連結:

<https://news.mingpao.com/ins/%e6%96%87%e6%91%98/article/20220321/s00022/1647792085510/%e4%bc%b0%e7%ae%97%e6%9f%93%e7%96%ab%e4%ba%ba%e6%95%b8-%e9%9c%80%e7%94%a8%e6%ad%a3%e7%a2%ba%e7%b5%b1%e8%a8%88%e6%96%b9%e6%b3%95%ef%bc%88%e6%96%87-%e5%b0%b9%e5%9c%8b%e8%81%96-%e9%99%b3%e9%bd%8a%e6%b5%a9%ef%bc%89>