

數學百子櫃系列 (二十四)

2017/18 中學生統計 創意寫作比賽 作品集



教育局
課程發展處數學教育組

Published by
Mathematics Education Section, Curriculum Development Institute,
Education Bureau, Government of the Hong Kong Special Administrative Region.
香港特別行政區政府教育局課程發展處數學教育組出版

數學百子櫃系列(二十四) 2017/18 中學生統計創意寫作比賽 作品集

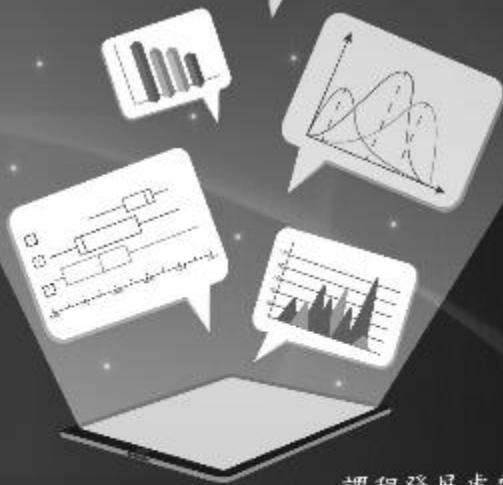
教育局數學教育組



政府物流服務署印

數學百子櫃系列 (二十四)

2017/18
中學生統計
創意寫作比賽
作品集



教育局
課程發展處數學教育組

Published by
Mathematics Education Section, Curriculum Development Institute,
Education Bureau, Government of the Hong Kong Special Administrative Region.
香港特別行政區政府教育局課程發展處數學教育組出版

版權

©2018 本書版權屬香港特別行政區政府教育局所有。本書任何部分之文字及圖片等，如未獲版權持有人之書面同意，不得用任何方式抄襲、節錄或翻印作商業用途，亦不得以任何方式透過互聯網發放。

ISBN 978-988-8370-80-1

編者的話

為配合香港數學教育的發展，並向教師提供更多的參考資料，課程發展處數學教育組於 2007 年開始邀請大學學者及資深教師撰寫專文，以及蒐集及整理講座資料，輯錄成《數學百子櫃系列》。本書《2017/18 中學生統計創意寫作比賽作品集》，是這個系列的第二十四冊。本書輯錄的文章，大部分是「2017/18 中學生統計創意寫作比賽」的優勝作品，由參賽的中學生撰寫。

本書所輯錄的參賽作品嘗試透過統計創意寫作，以簡潔的語言輕鬆地介紹統計的知識。

本書共有 15 篇文章，第 1 至 11 篇為「2017/18 中學生統計創意寫作比賽」的冠軍、亞軍、季軍和優異作品。其餘 4 篇則為邀請作品，分別由政府統計處的統計師，香港大學統計及精算學系的教授和數學教育組的課程發展主任撰寫，供讀者們閱覽。本書的文章充滿創意，期望讀者閱後能獲得啟發、不僅增加統計的知識，還能善用統計決策、解難。

此書得以順利出版，實有賴這次比賽的籌備委員會成員所付出的努力。在此，謹向撰寫作品的得獎隊伍、政府統計處的統計師、香港大學精算及統計學系的教授和數學教育組同工致以衷心的感謝。最後，更要多謝這次比賽的籌備委員會主席楊良河

博士和總評審主任張家俊博士。兩位鼎力協助，審訂本書的內容，讓學生能夠閱讀更多有趣的文章，增加他們學習統計的興趣。

如對本書有任何意見或建議，歡迎以郵寄、電話、傳真或電郵方式聯絡教育局課程發展處數學教育組：

九龍油麻地彌敦道 405 號九龍政府合署 4 樓

教育局課程發展處

總課程發展主任(數學)收

(傳真: 3426 9265 電郵: ccdoma@edb.gov.hk)

教育局課程發展處

數學教育組

前言

香港統計學會一直致力向社會各界推廣對統計的認知。除了每年與教育局合辦「中學生統計習作比賽」(SPC)，以鼓勵同學透過團隊合作形式學習正確運用統計數據及增進對社會的認識外，我們於 2009 年再與教育局合作創辦「中學生統計創意寫作比賽」(SCC)，旨在鼓勵學生透過創意的手法，以及科學和客觀的精神，用文字表達日常生活所應用的統計概念或利用統計概念創作一個故事。

回顧過去的參賽作品，喜見同學們對統計概念有更深入的認識及掌握如何正確地運用統計。近年，得獎作品的質素亦有所提升。本年度的比賽專題是「離散」。在簡介會中，張家俊博士介紹相關概念，並提示創作故事的不同方向，參加者能獲得一定的啟發。由於增加了初級組別，本屆參賽作品增至約 70 份，當中不乏精彩之作。文章取材創新，趣味盎然；同學們活學活用各種統計和概率的知識，分析有條有理，見解獨到，言之有物。中學生能有這樣的水平，實在難能可貴，值得欣喜和嘉許。本書輯錄了今屆所有的得獎作品，藉此嘉許得獎同學所付出的努力。希望同學能夠從創作或閱讀這些得獎作品中得到啟發，對統計的知識及其運用有更深入和正確的理解。

我們藉此機會感謝籌備委員會和評審委員會全體成員對評審的幫助和支持。他們的不遺餘力無疑是有助提高學生對統計的認知和興趣。最後，感謝香港大學統計及精算學系贊助今屆比賽的最佳專題寫作獎，和理大香港專上學院贊助今屆比賽的最佳文章演繹獎。

籌委會主席 楊良河博士

總評審主任 張家俊博士

2018年10月15日

目錄

編者的話	3
前言	5
目錄	7
冠軍作品: 超市抽獎大作戰	8
亞軍作品: 放假先嚟打風?	26
季軍作品: NBA MVP 寶座鹿死誰手— 誰能脫穎而出?.....	38
優異作品: 氣定神閒, 矢無虛發	52
優異作品: 真相, 盡在拉斯莫方程之中	74
優異作品: 如何挑選強積金	87
優異作品: To Trust, or Not to Trust – Gini Coefficient.....	98
優異作品: Assassination – ES.Drump.....	116
優異作品: 常見的罕有	150
優異作品: 報章「掩眼法」	155
優異作品: 「平均」中位數	163
邀請作品: 港式女士品茶	167
邀請作品: 背叛, 還是合作	175
邀請作品: 區別相關性與因果關係	179
邀請作品: The Use of Neighbours in Improving Prediction	183

冠軍作品:

超市抽獎大作戰

學校名稱：香港神託會培基書院

學生姓名：鄭傲臨、王伊澄

指導老師：黃智君老師

引言

在兩間超市的抽獎活動中，贏取三個大獎的概率是多少呢？本文以真人真事改編而成，筆者將會透過策略分析及概率計算去研究以上的問題，並剖析數學老師如何運用「抽獎必殺技」令自己的中獎機會增加超過一千倍！



圖片來源：黃智君老師的臉書專頁

伊澄在午膳時間看社交網站，並驚訝地說：「黃老師在超市 A 連續兩次抽中大獎！」傲臨回應說：「嘩！抽到一次大獎的機會已經非常低，連中兩次實在不可思議！我一定要找他問問！」傲臨如此積極是因為超市 B 也有抽獎活動，而獎品正是他渴望擁有的最新型號智能電話，所以他想看看黃老師有沒有「抽獎必殺技」！可是，他們到放學時已把此事忘得一乾二淨！

 **Dennis Wong** is 🍀 feeling lucky. ...
 22 January · Hong Kong · 🧑

當你用策略抽中兩次百佳大獎，以為智慧比運氣重要之後，惠康突然通知你抽中一部iPhone X!

真係抽一次就中！究竟是自己運氣太好還是活動太少人參加呢？

無獎數學題：在兩間超市的抽獎活動中，贏取三個大獎的概率是多少呢？



 Sheryl Wong and 33 others 12 comments

圖片來源：黃智君老師及惠康超級市場的臉書專頁

直到一個星期後，他們發現黃老師竟然再次中獎，更抽中了徹臨夢寐以求的智能電話！徹臨嫉妒地說：「其實，我也參加了這

個抽獎活動，為甚麼黃老師一再中獎，自己卻是『抽獎絕緣體』！」伊澄眼見傲臨深深不忿，決意與他一起查出真相：「黃老師作為數學老師，一定有方法增加自己的中獎機會。不如我們先研究一下超市 B 抽獎活動的條款及細則，以便推敲中獎背後的原理吧！」

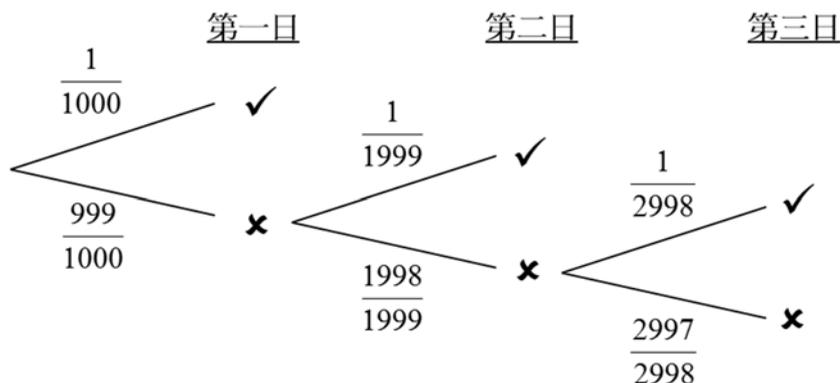
超市 B 抽獎活動的條款及細則

1. 在共 5 周的推廣期內，顧客憑單一發票購物滿\$88 或以上，可獲 1 次抽獎機會，並需即日起到抽獎網站登記。
2. 每日均設一個大獎，並將於推廣期內每日下一個工作天由電腦隨機抽出 1 名得獎者。除每日大獎得獎者外，其餘未被抽中的參加者均會自動累積到下一天的抽獎名單。

作為數學高材生的伊澄胸有成竹地說：「會不會是我們學過的概率有關呢？我們嘗試用樹形圖作分析吧！」傲臨臉有難色地說：「因為推廣期共有 35 日，所以樹形圖應該極其複雜吧！」伊澄拿起紙筆及計算機，冷靜地回應：「那麼，我們先將推廣期縮短為 3 日，並假設每日有 1000 人參加抽獎，作簡化的分析吧！」

研究模型(一)的樹形圖及概率計算(抽獎期簡化為 3 日)

假設每日有 1000 人參加抽獎



✓：參加者於當日中獎

✗：參加者未能於當日中獎，並自動累積到下一天的抽獎名單

第 n 日 參加	新參加 人數	已中獎的 總人數	未中獎的 總人數 N	中獎概率 P_n
1	1000	0	1000	$P_1 = \frac{1}{1000} + \left(1 - \frac{1}{1000}\right) \times P_2 \approx 0.18328\%$
2	1000	1	$1000 \times 2 - 1$ $= 1999$	$P_2 = \frac{1}{1999} + \left(1 - \frac{1}{1999}\right) \times P_3 \approx 0.08336\%$
3	1000	2	$1000 \times 3 - 2$ $= 2998$	$P_3 = \frac{1}{2998} \approx 0.03336\%$

在完成樹形圖及概率計算後，伊澄自豪地說：「原來如此！只要顧客愈早參加，中獎機會便愈高！」傲臨不滿地表示：「廢話！愈早參加，累積的抽獎次數當然愈多，中獎機會當然愈高！第一日參加的中獎機會不就是第三日參加的 3 倍嗎？」伊澄臉帶奸笑地回應：「傲臨！你的想法未免太簡單、太天真吧！不如我們先仔細研究一下中獎概率的計算！」

研究模型(一)的解說(抽獎期化為 3 日)

第三日(即最後一日)參加的中獎概率

$$\begin{aligned} &= \frac{1}{2998} \\ &\approx 0.03336\% \end{aligned}$$

第二日參加的中獎概率

$$\begin{aligned} &= P(\text{在第二日中獎}) + P(\text{在第二日未能中獎，並在第三日中獎}) \\ &= \left(\frac{1}{1999}\right) + \left(1 - \frac{1}{1999}\right) \times \left(\frac{1}{2998}\right) \\ &\approx 0.05003\% + 0.03334\% \\ &\approx 0.08336\% \end{aligned}$$

第一日參加的中獎概率

$$\begin{aligned} &= P(\text{在第一日中獎}) + P(\text{在第一日未能中獎，並在第二日中獎}) \\ &\quad + P(\text{在第一日及第二日未能中獎，並在第三日中獎}) \\ &= \frac{1}{1000} + \left(1 - \frac{1}{1000}\right) \times \left(\frac{1}{1999}\right) + \left(1 - \frac{1}{1000}\right) \times \left(1 - \frac{1}{1999}\right) \times \left(\frac{1}{2998}\right) \\ &\approx 0.1\% + 0.04997\% + 0.03331\% \\ &\approx 0.18328\% \end{aligned}$$

伊澄繼續解釋：「如果顧客在第三日參加抽獎，累積參加人數已有 3000 人，減去第一及第二天的得獎者共 2 人，中獎概率只有 $\frac{1}{2998}$ ，即 0.03336%，機會可謂微乎其微！如果顧客在第一

日參加抽獎，累積的抽獎次數當然較多！除此之外，因為活動初期的累積參加人數較少，中獎概率便會大增！第一日參加的中獎概率是第三日參加的 5.49 倍(即 $0.18328\% \div 0.03336\%$)！」傲臨恍然大悟地表示：「如果抽獎期是 35 日，第一日參加的中獎概率豈不是更高！」伊澄為了尋根究底，決定繼續研究，並回家利用電腦的試算表(Excel)作輔助(見附件一)！

研究模型(二)的假設及重要發現(抽獎期共 35 日)

- 假設每日新參加的人數形成等差數列 (arithmetic sequence)，而首項及等差同樣是 1000，整個抽獎期的總參加人數是 35000。
- 假設顧客在第 n 日參加抽獎，
未中獎的總人數 $N = 1000 + 999(n-1)$
$$= 999n + 1$$

中獎概率 $P_n = \frac{1}{N} + \left(1 - \frac{1}{N}\right) \times P_{n+1}$
- 第一日參加的中獎概率是第三十五日參加的 145 倍 (即 $0.41415\% \div 0.00286\%$)
- 如果只考慮每日大獎的價值(\$13,920)及參加抽獎的購物成本(\$88)，並假設黃老師在抽獎期的第一日參加，

$$\begin{aligned}
& \text{抽獎的期望值} = P_1 \times 13920 + (1 - P_1) \times (-88) \\
& = 0.41415\% \times 13920 + (1 - 0.41415\%) \times (-88) \\
& = -29.99 \\
& < 0
\end{aligned}$$

研究模型(二)的概率計算

第 n 日 參加	新參加 人數	已中獎的 總人數	未中獎的 總人數 N	中獎概率 P_n
1	1000	0	1000	$P_1 = \frac{1}{1000} + \left(1 - \frac{1}{1000}\right) \times P_2 \approx 0.41415\%$
2	1000	1	$\frac{1000 \times 2 - 1}{= 1999}$	$P_2 = \frac{1}{1999} + \left(1 - \frac{1}{1999}\right) \times P_3 \approx 0.31447\%$
3	1000	2	$\frac{1000 \times 3 - 2}{= 2998}$	$P_3 = \frac{1}{2998} + \left(1 - \frac{1}{2998}\right) \times P_4 \approx 0.26457\%$
n	1000	$n - 1$	$\frac{1000n - (n - 1)}{= 999n + 1}$	$P_n = \frac{1}{N} + \left(1 - \frac{1}{N}\right) \times P_{n+1}$
34	1000	33	$\frac{1000 \times 34 - 33}{= 33967}$	$P_{34} = \frac{1}{33967} + \left(1 - \frac{1}{33967}\right) \times P_{35} \approx 0.00580\%$
35	1000	34	$\frac{1000 \times 35 - 34}{= 34966}$	$P_{35} = \frac{1}{34966} \approx 0.00286\%$

經過廢寢忘餐的計算後，伊澄自豪地向傲臨展示自己的研究成果，但傲臨卻臉帶不屑地回應：「即使黃老師在抽獎期的第一日參加，中獎概率只有 0.4%，而期望值更是負數！以黃老師的聰明才智，絕不會參加這個『不划算的賭博』！」伊澄覺得傲臨一矢中的，所以決定一起去找黃老師問個清楚！

看過兩位同學的計算後，黃老師大讚他們的探究精神，並回應說：「我的確是在第一天便參加抽獎，但中獎概率絕對比 0.4% 高很多！」傲臨謙虛地說：「願聞其詳！」「首先，我必須澄清『以小博大』的賭博心態絕不可取，但抽獎與賭博的性質並不相同！我購買\$88 的貨品全是必需的日用品，而不是刻意消費去創造抽獎機會！因為我只是善用超市購物所衍生的額外價值去參加抽獎，除時間成本外，並無實質的成本，所以期望值絕不是負數吧！」黃老師為學生的「賭博論」反駁。



黃老師繼續解釋：「雖然你們的概率分析做得不錯，但關於參加人數的假設不太合理！在抽獎活動的初期，參加人數未必如你們想像那麼多！你們試找出原因吧！」傲臨搶著回答：「因為超市的宣傳在活動初期未到位，所以未能吸引太多顧客參加抽

獎！」伊澄恍然大悟地說：「原來如此！那麼，『每日有 1000 人參加抽獎』的假設絕不合理！隨著宣傳的效果增加，每日新參加的人數亦會相應增加！」黃老師滿意地回應：「伊澄真聰明，能夠舉一反三！其實，每日新參加的人數還會受到很多其他因素影響！例如，週末的顧客較多，參加抽獎的人數亦會相應增加！在概率研究中，與其用『等差數列』作假設，何不用『等比數列』以突顯新參加人數的快速增長呢？」黃老師在伊澄的試算表略作修改（見附件二）以輔助概率的計算！

研究模型(三)的假設及重要發現(抽獎期共 35 日)

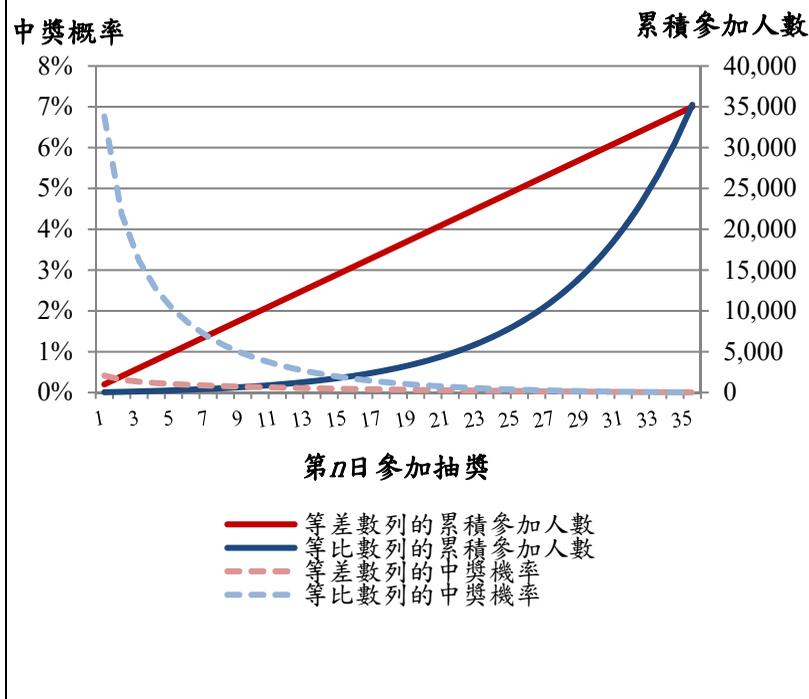
- 假設每日新參加的人數形成等比數列 (geometric sequence)，而首項及等比分別是 40 及 1.15，即每日新參加的人數增加 15%，整個抽獎期的總參加人數是 35247。
- 假設顧客在第 n 日參加抽獎，
未中獎的總人數
$$N = \frac{40(1.15^n - 1)}{1.15 - 1} - (n - 1)$$
- 第一日參加的中獎概率是第三十五日參加的 2381 倍（即 $6.76329\% \div 0.00284\%$ ）。

研究模型(三)的概率計算

第 n 日 參加	新參加 人數	已中獎的 總人數	未中獎的 總人數 N	中獎概率 P_n
1	40	0	40	$P_1 = \frac{1}{40} + \left(1 - \frac{1}{40}\right) \times P_2 \approx 6.76329\%$
2	40×1.15^1 ≈ 46	1	85	$P_2 = \frac{1}{85} + \left(1 - \frac{1}{85}\right) \times P_3 \approx 4.37261\%$
3	40×1.15^2 ≈ 53	2	137	$P_3 = \frac{1}{137} + \left(1 - \frac{1}{137}\right) \times P_4 \approx 3.23419\%$
n	$40 \times 1.15^{n-1}$	$n-1$	$\frac{40 \times (1.15^n - 1)}{1.15 - 1}$ $-(n-1)$	$P_n = \frac{1}{N} + \left(1 - \frac{1}{N}\right) \times P_{n+1}$
34	40×1.15^{33} ≈ 4028	33	30582	$P_{34} = \frac{1}{30582} + \left(1 - \frac{1}{30582}\right) \times P_{35} \approx 0.00611\%$
35	40×1.15^{34} ≈ 4635	34	35213	$P_{35} = \frac{1}{35213} \approx 0.00284\%$

聽過黃老師的分析後，徹臨若有所失地表示：「2381 倍那麼高！因為我在最後一日才參加抽獎，難怪我的中獎機會那麼低！」黃老師繼續解釋：「在我們的假設中，等差數列和等比數列的研究模型同樣地有大約 35000 人參加抽獎，但中獎的概率可謂差天共地！除了概率計算外，我們更可以利用圖表了解累積參加人數的增長！除了最後一日外，等比數列的累積參加人數一直遠低於等差數列，所以其中獎概率亦會較高！」

中獎概率與累積參加人數的比較



雖然伊澄能理解黃老師的分析，但仍按捺不住地問：「其實，超市 B 的 6.76% 中獎率亦不算太高，但為何老師可以在超市 A 連中兩個大獎，那豈不是超級幸運？」黃老師捧腹大笑地用全新的智能電話向同學展示超市 A 的抽獎活動宣傳，並解釋說：「超市 A 的抽獎是靠策略，而不是靠運氣，中獎可說是十拿九穩！」



圖片來源：百佳超級市場的臉書專頁

傲臨大惑不解地表示：「如何可以確保自己成為每日第 8 名或第 88 名參加者呢？」黃老師滿有自信地說：「非常容易！只要第 1 至第 8 名參加者也是自己便可以！」伊澄搶著回應：「沒可能！超市 A 在全港有超過 250 間分店，而且每日的顧客人數以萬計！」黃老師臉帶奸笑地回應：「其實，全港只有 3 間分店在早上 7 點開始營業！在兩天的清晨，我只是分別買了 8 盒朱古力及 8 盒曲奇餅禮盒，而且每盒以獨立的賬單結算，然後立即到抽獎網站登記！在不違反抽獎活動規則下，輕鬆地贏走兩套名貴廚具！」伊澄激動地回應：「老師在兩間超市都用了相同的『抽獎必殺技』，就是『早起的鳥兒有蟲吃』！」「這策略非常管用！讀書如是，做人如是！」黃老師深表認同。



最後，做臨慨嘆：「自己既不善於概率分析，又不懂運用策略！看來今生今世也無緣中獎吧！」黃老師雙眼發光地回應：「年輕人！不要太小看自己吧！何不將剛才的討論及研究寫成故事去參加『中學生統計創意寫作比賽』！只要你們願意努力，相信比賽過程中的得著比兩間超市的三個大獎更寶貴吧！」

(~2500 字)

重要聲明：以上的假設及數字全為學生的猜想，而超市抽獎活動的條款及細則亦略作簡化以突顯本文的故事性。

附件一：研究模型(二)的參加人數及中獎概率計算

假設每日新參加的人數形成等差數列，而首項及等差同樣是1000。

第 n 日 參加	新參加 人數	累積參加 人數	已中獎的 總人數	未中獎的 總人數 N	中獎機率 P_n
1	1000	1000	0	1000	0.41415%
2	1000	2000	1	1999	0.31447%
3	1000	3000	2	2998	0.26457%
4	1000	4000	3	3997	0.23130%
5	1000	5000	4	4996	0.20633%
6	1000	6000	5	5995	0.18635%
7	1000	7000	6	6994	0.16970%
8	1000	8000	7	7993	0.15542%
9	1000	9000	8	8992	0.14293%
10	1000	10000	9	9991	0.13182%
11	1000	11000	10	10990	0.12183%
12	1000	12000	11	11989	0.11274%
13	1000	13000	12	12988	0.10440%
14	1000	14000	13	13987	0.09671%
15	1000	15000	14	14986	0.08957%
16	1000	16000	15	15985	0.08290%
17	1000	17000	16	16984	0.07665%
18	1000	18000	17	17983	0.07077%

第 n 日 參加	新參加 人數	累積參加 人數	已中獎的 總人數	未中獎的 總人數 N	中獎機率 P_n
19	1000	19000	18	18982	0.06521%
20	1000	20000	19	19981	0.05995%
21	1000	21000	20	20980	0.05494%
22	1000	22000	21	21979	0.05018%
23	1000	23000	22	22978	0.04563%
24	1000	24000	23	23977	0.04128%
25	1000	25000	24	24976	0.03711%
26	1000	26000	25	25975	0.03311%
27	1000	27000	26	26974	0.02926%
28	1000	28000	27	27973	0.02555%
29	1000	29000	28	28972	0.02198%
30	1000	30000	29	29971	0.01853%
31	1000	31000	30	30970	0.01519%
32	1000	32000	31	31969	0.01196%
33	1000	33000	32	32968	0.00884%
34	1000	34000	33	33967	0.00580%
35	1000	35000	34	34966	0.00286%

附件二：研究模型(三)的參加人數及中獎概率計算

假設每日新參加的人數形成等比數列，而首項及等比分別是 40 及 1.15。

第 n 日 參加	新參加 人數	累積參加 人數	已中獎的 總人數	未中獎的 總人數 N	中獎機率 P_n
1	40	40	0	40	6.76329%
2	46	86	1	85	4.37261%
3	53	139	2	137	3.23419%
4	61	200	3	197	2.52215%
5	70	270	4	266	2.02414%
6	80	350	5	345	1.65400%
7	93	443	6	437	1.36823%
8	106	549	7	542	1.14184%
9	122	671	8	663	0.95913%
10	141	812	9	803	0.80962%
11	162	974	10	964	0.68597%
12	186	1160	11	1149	0.58283%
13	214	1374	12	1362	0.49624%
14	246	1620	13	1607	0.42313%
15	283	1903	14	1889	0.36114%
16	325	2229	15	2214	0.30837%
17	374	2603	16	2587	0.26331%
18	430	3033	17	3016	0.22475%

第 n 日 參加	新參加 人數	累積參加 人數	已中獎的 總人數	未中獎的 總人數 N	中獎機率 P_n
19	495	3528	18	3510	0.19166%
20	569	4098	19	4079	0.16322%
21	655	4752	20	4732	0.13873%
22	753	5505	21	5484	0.11763%
23	866	6371	22	6349	0.09941%
24	996	7367	23	7344	0.08368%
25	1145	8512	24	8488	0.07007%
26	1317	9828	25	9803	0.05829%
27	1514	11343	26	11317	0.04810%
28	1741	13084	27	13057	0.03926%
29	2003	15087	28	15059	0.03161%
30	2303	17390	29	17361	0.02497%
31	2648	20038	30	20008	0.01921%
32	3046	23084	31	23053	0.01421%
33	3503	26587	32	26555	0.00988%
34	4028	30615	33	30582	0.00611%
35	4632	35247	34	35213	0.00284%

亞軍作品：

放假先嚟打風？

學校名稱：香港培正中學

學生姓名：莫嘉浩 潘逸朗 黃啓元

指導教師：梁偉雄老師

摘要：香港常受颱風吹襲，坊間當有人說香港在週末或假期才會發出八號烈風或暴風信號。這個專題習作嘗試探討此類“坊間傳言”，發現天文台在考慮發出八號信號時，雖然沒有完全根據錄得的數據來決定的，但上述的“傳言”亦不真確。



引言

眾所周知，香港是一個常受颱風吹襲的地方，但到底香港的颱風有甚麼特點呢？有人說香港在週末或假期時才有八號烈風或暴風信號發出，又有人說信號只會在晚上下班後才發出等等問題。是不是真有其事，抑或只是湊巧？我們為大家拆解。

我們收集了近十年來颱風吹襲香港的資料，希望探討這些問題。



發出八號信號的指標

根據香港天文台的官方網站，發出八號信號的參考範圍是由八個涵蓋全港並接近海平面的參考測風站組成的網絡。當參考網絡中半數或以上的測風站錄得或預料錄得的持續風速達到每小時 63 至 117 公里，且風勢可能持續時，則會發出八號信號。

當熱帶氣旋距離香港較遠時，可能它的外圍雨帶會橫過本港並帶來狂風，導致短暫時間有四個或以上的參考測風站錄得風速超過八號信號指定的風速下限。如果預料在雨帶掠過後風勢不會持續，天文台不會發出八號信號。

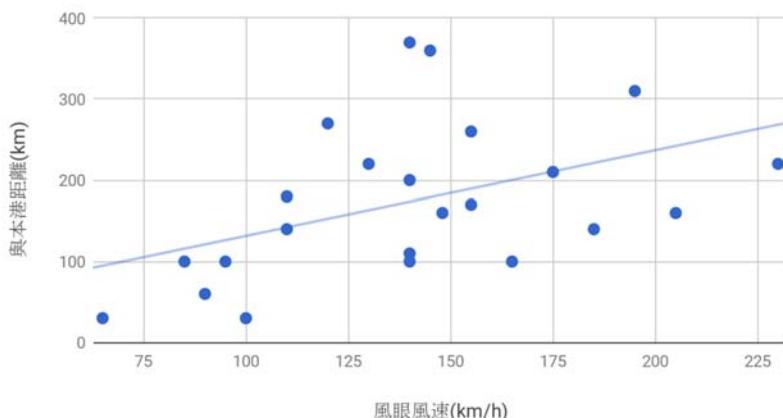
揀選這些測風站是基於它們處於較為空曠的位置及地理上的分布，當中包括自然山脈分隔的考慮。這個參考測風站網絡應可概括地反映全港的風勢。



改發八號信號的時候...

颱風距離香港有多遠？風速是多少？

2007-2017年改發八號信號時颱風的距離和風速



(以上的風速是指由香港天文台所錄得的十分鐘平均風速)

由上圖可見，我們隱約可以看見距離和風速有一個綫性的關係，但有很多例外。經過計算，颱風的距離和風速的相關係數是 $r = 0.4543$ ，可見有正比的關係。

如圖中和相關度可見，它的範圍是很廣的，可以有很多例外：例如同樣風速是 140km/h 的兩個颱風，一個在 100km 時改發信號，但另一個在 370km 外已經發出八號烈風或暴風信號了。所以，我們發現原來天文台不只是完全參考颱風的數據來決定是否要發出八號的颱風信號。

公眾假期才會刮颱風？

這個研究的主要目的是研究會否只在週末或假期時才會發出八號信號，但到底是否在 17 天公眾假期時發出特別多八號信號呢？

根據 2007-2017 年的八號信號數據，我們發現天文台並沒有在公眾假期發出八號信號，證明公眾假期才會打風的說法並不準確。

星期幾發出颱風的次數最多？

發放信號的時間分佈平不平均？

發現公眾假期才會打風的說法並不準確後，我們想找出是否在週末時才有八號或以上的信號發出，所以我們就將搜集到的資料分類。

2007-2017年發出八號信號的日期



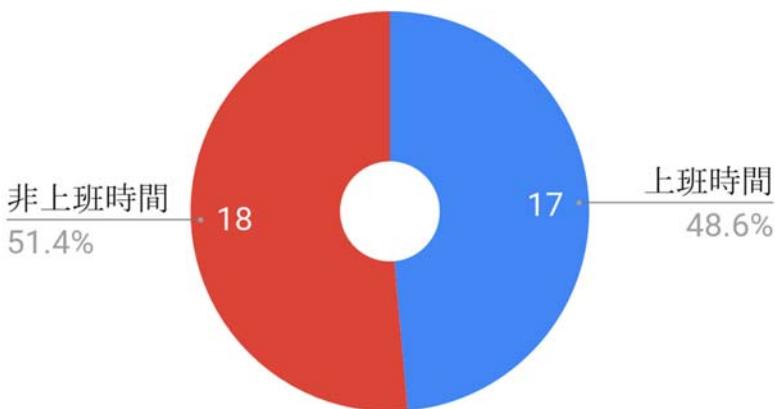
上圖中顯示了 38 日有達致八號或以上信號等級的颱風吹襲日

期。可見在星期的層次上，除了星期二特別多，星期四、六特別少外，其他日子的分佈都和平均數 5.4 接近。所以我們得出週末和放颱風假並沒有正向關係，反而平日有颱風的比率還高於週末的比率。

甚麼時段發出八號的信號

發出八號或以上的信號的時間，是否都在上班時間？假設星期一至五上班時間為上午九時至晚上六時，星期六的為上午九時至下午一時。

2007-2017年發出八號或以上信號的時段



35 段有颱風的時段中，在上班時間的有 17 個，不在上班時間的有 18 個。一星期內上班時間有 49 小時，非上班時間則有 119 小時。相比上班時間佔一星期的 30%，48.6% 颱風在上班時間吹襲，可說非常划算。

可是，是不是在上班時間的發出的信號都有足夠長時間，令當日成為真正的全日假期呢？

發出信號的維持時間



圖中可見，上班時間內時間較短(0-8 小時)有 13 次，比上班時間內足夠放全日的(8-10 小時)的 4 次為多。

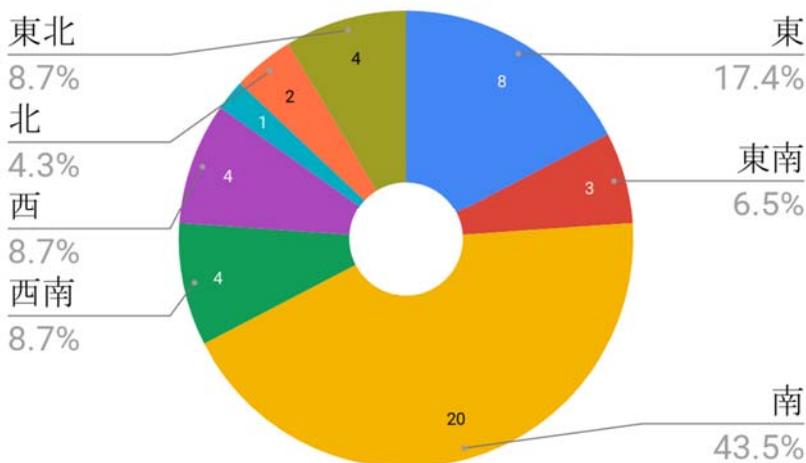
颱風登陸後缺乏水氣發展，風勢會很快減弱。於香港東面登陸的颱風缺乏水氣發展，風勢減弱快於同時在香港或香港以西登陸，未上岸的颱風。由於近年較多於香港或香港以東登陸的颱風，所以近年發出八號或以上信號的時段長度較短。

改發八號或以上信號的時候...

颱風在香港的哪個方向？

颱風會從不同方向吹襲香港，颱風會不會從一個特定方向吹襲香港呢？

2007-2017年發出八號或以上信號時颱風吹襲的方向



(偏南作南計算，偏東作東計算，如此類推)

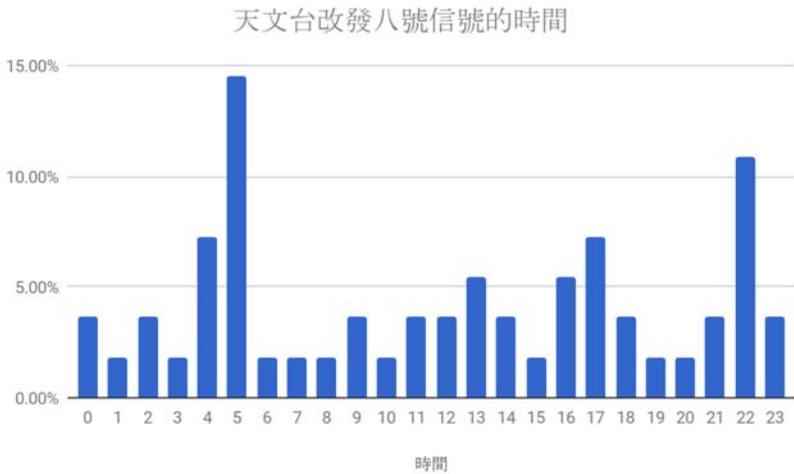
從圖中可見，大部份颱風都是從南方、西方或東方吹來的(共 39 次)，極少有從北方來的風(共 7 次)。

香港只有北方不是對着南海，而是對着內陸。而南海有非常多颱風生成或吹來，所以香港北方的風較為少，向南海處較多颱風。從上圖所見，很多時颱風都由東方或南方的南海吹襲香港。

近年多了颱風於香港東面登陸，少了在香港以西登陸的颱風，所以從香港以西吹襲的颱風便較少。

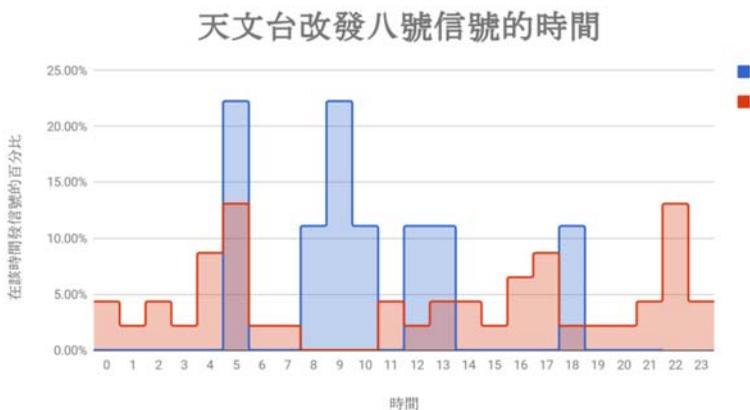
天文台多在清晨五點改發八號信號？

(資料為 1980-2017 年的數據)



從上圖所見，天文台較常在早上五時改發八號信號 (8 次)，而在早上六時至十時較少改發八號信號 (1-2 次)。早上六時至十時是上班時間，所以很多人都認為天文台預測到那天的風力將增大，為了保障市民安全而在早上六時前發出信號。這樣做可避免在上班時間改發八號信號，令正在前往上班，上學地點的市民要突然改回家，做成混亂。

上班日子和公眾假期發出八號信號的時間差別



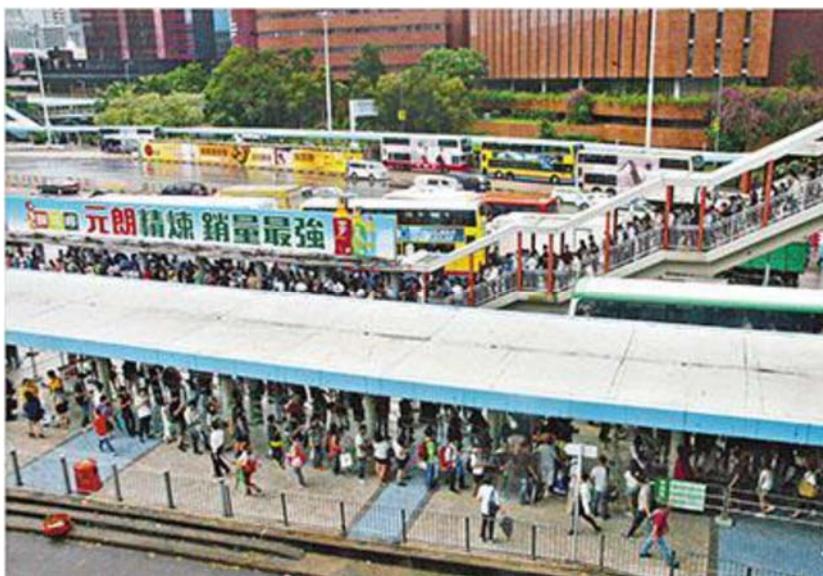
*藍色是公眾假期，紅色是上班日（包括星期六）

但當我們把資料分開成平日和周末統計，發現天文台不論在週末或平日都是集中在早上五時改發八號信號的。但週末是大部分人放假的日子，天文台沒有必要在這段時間，為了保障安全，避免混亂等原因，特意提早改發八號信號。所以我們認為天文台是主要根據天氣情況而作出考慮而非受到其他因素影響的。例如天文台可能預測到在早上五時日出後會導致風向或風力改變，而決定改發八號信號。

總結：

經過這次研究，我們得出以下的結論：

- 天文台在考慮發出八號信號時，沒有完全根據錄得的數據來決定的。
- 星期六日發出八號烈風或暴風信號的機會不是特別的高，而是平均分佈的。
- 普遍八號信號維持的時間較短。
- 颱風通常都是從香港南面吹襲本港的。
- 天文台通常在清晨 5 點改發八號信號。
- 公眾假期改發八號信號的時間與上班日的時間是有所不同的。



(字數：2440 字)

參考資料

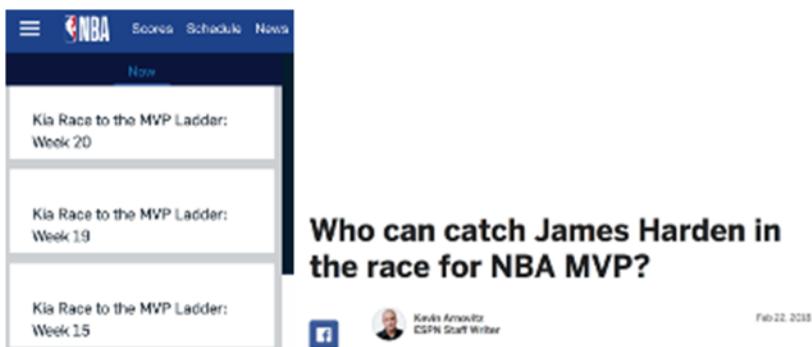
1. 天文台自 1946 年發出熱帶氣旋警告訊號的記錄:
http://www.hko.gov.hk/wxinfo/climat/warndb/warndb1_c.shtml
2. 香港熱帶氣旋追擊站:
<http://www.hkcoc.com/record.php>
3. 天文台天氣報告資料庫:
<https://www.weather.org.hk/wxreport.html>
4. 地下天文台《為甚麼近年多在晚間打風?》網誌
<http://www.weather.org.hk/blog/?p=1957>

引言

NBA(National Basketball Association，國家籃球協會)是不少同學喜歡的籃球賽事，當子軒看到有關本年度 NBA 最有價值球員(MVP，Most Valuable Player)的新聞時，立即與自己的好朋友雪程討論今年度的 MVP 熱門候選球員，看看誰較大機會脫穎而出。他們繼而搜集相關的數據，並運用統計學作分析。



一天，子軒和雪程看到了關於本年度 NBA MVP 候選人的新聞。



(新聞標題截圖)

子軒：從 NBA 及 ESPN 的分析，本年度 NBA MVP 熱門有：James Harden，LeBron James，Kevin Durant，Stephen Curry 及 Anthony Davis。雪程，你覺得今年的 MVP 寶座會花落誰家呢？



James Harden



LeBron James



Kevin Durant



Stephen Curry



Anthony Davis

雪程：我覺得 James Harden 會拿到 MVP ，
你認為呢？



子軒：我認為 LeBron James 會拿到 MVP 。



雪程：哈哈！原來我們的看法都很相近，大家認為能拿到 MVP
的球員都有「James」這個名字。

子軒：你真是風趣！

雪程：坊間在談論 MVP 候選人時，往往只比較球員們得分的高低。但作為 MVP 球員，他亦應該對球隊有「穩定」的貢獻。不如我們試試用標準差(Standard Deviation, SD)來看看誰表現較平穩。

子軒：好！但甚麼是標準差呢？

雪程：標準差是用作量度一組數據的離散程度。簡單來說，一個較大的標準差，代表大部分數據和其平均值之間的差異較大；一個較小的標準差，代表這組數據較接近其平均值。

子軒：我明白了。如果球員得分的標準差較大，代表他的表現較不穩定；球員得分的標準差較小，代表他的表現比較平穩，對球隊絕對有益處。

雪程：雖然 NBA 網上的數據庫並沒有展示各球員得分的標準差，但我們可以下載五名大熱 MVP 候選球員每場的數據資料，再以 Excel 自行計算其**得分標準差**。標準差區分為母體標準差以及樣本標準差：

樣本標準差

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

母體標準差

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - \mu)^2}$$

(μ 為平均值)

因為我們下載了所有場次的數據，在計算標準差時，會選用母體標準差。

子軒：對呀！我們是次不是作抽樣分析，而是分析所有比賽的數據，所以應該使用母體標準差。今天是 2018 年 3 月 3 日，那麼我們就只統計由開季至今天的賽事吧！另外，我們也可以計算五名球員的助攻數據，讓我們可以多一個角度分析球員的表現及貢獻。

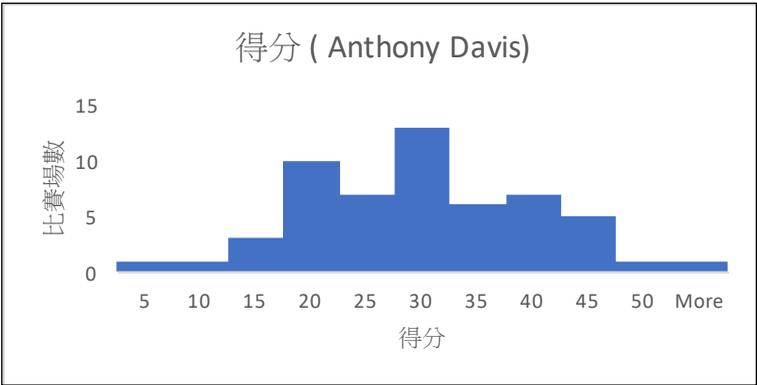
雪程：直至 2018 年 3 月 3 日，五名球員的表現如下：

	得分	標準差 (得分)	助攻	標準差 (助攻)
<u>LeBron James</u>	26.72	7.63	8.97	3.49
James Harden	31.26	9.67	8.85	2.99
Anthony Davis	28.13	10.36	2.38	1.74
Kevin Durant	26.05	6.79	5.42	2.47
Stephen Curry	26.69	7.95	6.35	2.23

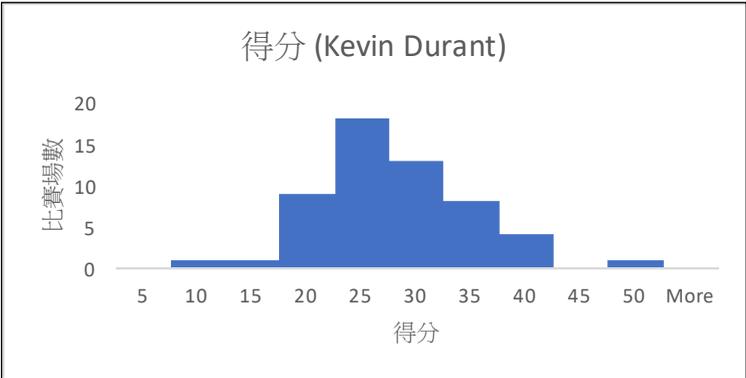
(五位球員的得分及助攻概要：平均值、標準差)

從以上的分析，可見 Anthony Davis 的得分標準差是五人當中最大的，為 10.36，即代表他每場得分的差異比較大；反之；Kevin Durant 的得分標準差是最小的，為 6.79，代表他每場得分相差較少，表現較穩定。

子軒：我們也可以用直方圖來看看 Anthony Davis 和 Kevin Durant 的得分分佈：



(直方圖：Anthony Davis 的得分)



(直方圖：Kevin Durant 的得分)

雪程，你所用的標準差推測的結果正確無誤。從兩者的直方圖所看，Kevin Durant 的得分較 Anthony Davis 分佈得比較集中，Kevin Durant 大部分比賽都獲得 20 分至 45 分，表現穩定。相反，Anthony Davis 的得分直方圖則比較分散，而且在 20 及 30 分都呈現兩個高位，反映其標準差較高。

子軒：James Harden 的平均得分是最高的，有 31.26 分。他得分是最高的，標準差相對較大。相反，如果得分較小，其標準差亦有機會較小。因此我們很難單靠標準差去量度球員的穩定性，那麼應該怎樣公平地比較各球員的穩定值？

雪程：我有辦法。我們可以利用變異係數(Coefficient of Variation, CV)來公平地比較兩組或以上的數據的離散程度。比較數據離散程度時，不能只比較標準差的大小，需要以一種相對的測度值作為比較的標準，所以需同時考慮平均值。

$$CV = \frac{\sigma}{\mu} \quad (\sigma \text{ 為標準差；}\mu \text{ 為平均值})$$

當變異係數的數值愈接近 0 的時候，則代表那一組數據的離散程度較低。相反地，如果變異係數的數值遠離 0 的話，那一組的數據則是分散得較嚴重。

子軒：我明白了！我們立刻開始計算球員們得分及助攻的變異係數吧！

	得分	標準差 (得分)	變異係數 (得分)	助攻	標準差 (助攻)	變異係數 (助攻)
LeBron James	26.72	7.63	0.29	8.97	3.49	0.39
James Harden	31.26	9.67	0.31	8.85	2.99	0.34
Anthony Davis	28.13	10.36	0.37	2.38	1.74	0.73
Kevin Durant	26.05	6.79	0.26	5.42	2.47	0.46
Stephen Curry	26.69	7.95	0.30	6.35	2.23	0.35

(五位球員的得分及助攻概要：平均值、標準差、變異係數)

雪程：你看看 James Harden 的平均得分比 LeBron James 高，他一定能拿到今次的 MVP。

子軒：你不要那麼開心啊！還記得我們剛剛所提及的標準差和變異係數均是量度球員的穩定性嗎？穩定的表現是獲得 MVP 的重要條件之一呀！

雪程：我真是大意了。五名球員的得分變異係數大約為 0.3，當中 Anthony Davis 為最高的 0.37，顯示考慮平均值及標準差後，他的表現最不穩定。



子軒：那除去 Anthony Davis 後，讓我們集中看看 LeBron James 和 James Harden 吧！James Harden 的得分比 LeBron James 多

4.5 分，而且兩者的得分變異係數十分相似，所以 James Harden 在得分方面略勝一籌！

雪程：在助攻方面，LeBron James 和 James Harden 的助攻平均值都非常接近，大家不相伯仲，但 James Harden 的助攻變異係數稍稍比 LeBron James 低。總括而言，James Harden 在得分及助攻方面的比 LeBron James 優勝。子軒，你同意我的推測嗎？

子軒：哈哈！雪程你果然是「神算子」。我現在亦都同意 James Harden 會是本年度的 MVP。

雪程：既然我們都支持 James Harden 拿到 MVP，不如我們看看他以往的表現。

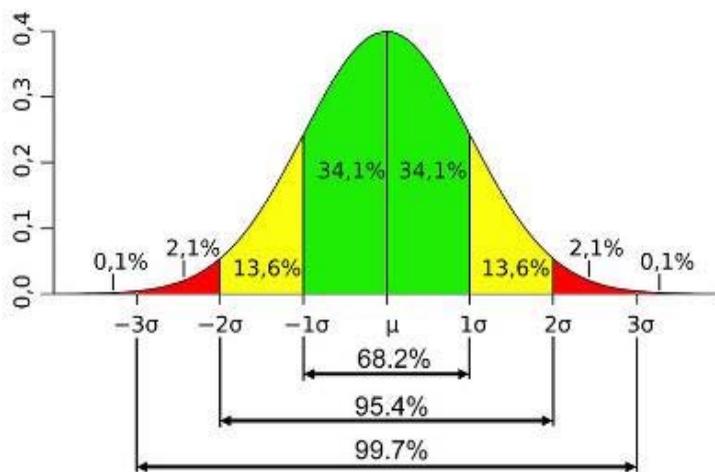
	<u>賽數</u>	<u>得分</u>	<u>標準差 (得分)</u>	<u>變異係數 (得分)</u>
2017-2018 (直至 3/3/2018)	54	31.26	9.67	0.31
2016-2017	81	29.09	8.78	0.30
2015-2016	82	28.98	8.51	0.29
2014-2015	81	27.37	9.72	0.36
2013-2014	73	25.36	8.64	0.34

(James Harden 各賽季的表現概要)

子軒：原來 James Harden 的得分一直都有進步，由 2013-2014 賽季的 25.36 分，到今年 2017-2018 賽季的 31.26，進步了兩成

多，十分不錯！再者，他的變異係數由 2013-2014 球季的 0.34，進步到今年的 0.31。可見他的穩定性有所提升。

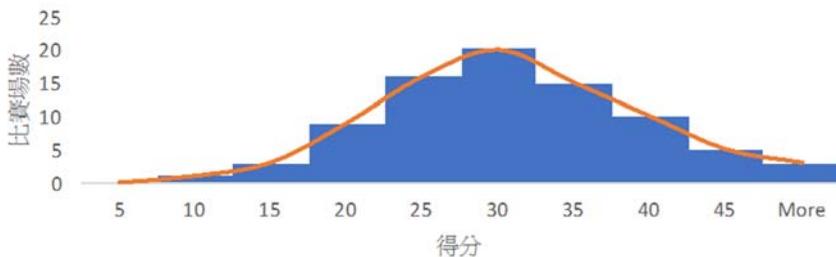
雪程：標準差表現在正態分佈圖上，有其一定的比例，大約為「68%，95%，99.7%」，也就是 68% 數值分佈在距離「平均值」有 1 個「標準差」之內的範圍；約 95% 數值分佈在距離「平均值」有 2 個「標準差」之內的範圍；以及約 99.7% 數值分佈在距離「平均值」有 3 個「標準差」之內的範圍。



(正態分佈圖)

	賽數	得分(μ)	標準差(σ)	$Pr(\mu-\sigma < X \leq \mu+\sigma)$	$Pr(\mu-2\sigma < X \leq \mu+2\sigma)$
2017-2018	54	31.26	9.67	0.74	0.93
2016-2017	81	29.09	8.78	0.60	0.96
2015-2016	82	28.98	8.51	0.67	0.96
2014-2015	81	27.37	9.72	0.68	0.96
2013-2014	73	25.36	8.64	0.60	0.96

James Harden 的得分分佈 (2015 - 2016)



(James Harden 2015-2016 年度得分分佈)

子軒：從上表及直方圖可見，James Harden 的得分數據非常接近一個正態分佈，尤其是 2015 至 2016 年的賽季。因為在大部分的賽季內，數據分佈在距離「平均值」分別有 1 個及 2 個「標準差」之內的比例十分接近 68% 及 95%。這正是標準差和正態分佈密切的關聯，亦幫助我們使用「標準差」去闡述數據。

總結：

經過反覆討論，通過各種的數據搜集和分析後，子軒及雪程都一致認為 James Harden 會是 MVP 的大熱人選。這不但基於他驚人的得分及助攻能力，更歸功於他強大的「穩定性」。



(2452 字)

參考資料：

1. NBA: Kia Race to the MVP Ladder
<http://www.nba.com/mvp-ladder#/>
2. ESPN: Who can catch James Harden in the race for NBA MVP?
http://www.espn.com/nba/story/_/id/22467731/2018-nba-mvp-race-steph-curry-lebron-james-catch-james-harden
3. 標準差 (Standard Deviation, SD)
<https://zh.wikipedia.org/wiki/標準差>
<http://slashlook.com/archive2014/20140506.html>
4. 球員數據
<http://stats.nba.com>

5. 變異係數 (Coefficient of Variation, CV)

<https://baike.baidu.com/item/變異係數>

<http://wiki.mbalib.com/wiki/變異係數>

6. 圖片

<https://clutchpoints.com/nba-all-rookie-teams-announced/>

<https://detroitssportsnation.com/general/rgriffin/2017-18-nba-season-preview/10-2017/96808/>

優異作品:

氣定神閒，矢無虛發

學校名稱：香港培正中學

學生姓名：林靜嵐、湯迦立

指導教師：梁偉雄老師、王敏瑤老師

摘要：射箭是一項歷史悠久的運動，在比賽中箭手要有穩定的發揮，才能取得高分。我們會從以下方向去探討如何令箭手獲取高分且有穩定的表現：

- 射手的拉距磅數
- 穩定性與分數的關係
- 室外與室內比賽的區別



背景：

小迦剛剛開始射箭，有一點成績，想有更加好的表現。小立是一名射箭運動員，他在香港有一定的水準，即將代表香港到海外比賽。他們現正和教練談論增加弓的磅數的問題和甚麼因素會影響表現。



教練：小迦，你在剛剛的比賽中升了初級組，要射 30 米的靶，要開始加磅啦。

小迦：為甚麼要加磅呢？拉開越重的弓是不是會有更好的成績呢？

小立：應該是吧！我用比較重磅數的弓的成績比用輕磅數的弓的成績好。

教練：我作了世界排名首 100 的選手在近兩年參加所有室外遠距離的比賽的平均箭值和選手的拉距磅數的關係比較。

小迦：為甚麼用平均箭值呢？

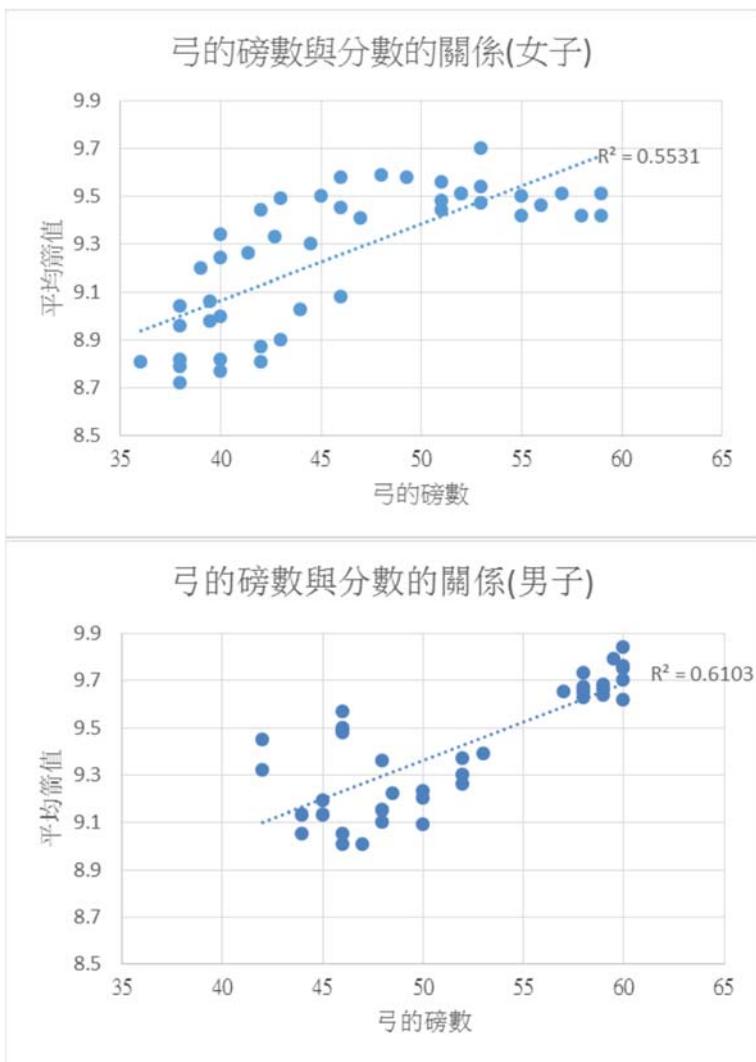
教練：因為平均箭值能真正代表射手的表現，如果選用某一次或兩次的比賽，有機會該比賽是某射手失準或突然很高分，並不是穩定的表現。而且遠距離比賽箭到靶需時較長，比近距離需要較高的技術。而且遠距離的比賽較容易受外來因素影響，例如環境、天氣、雨量等等。

小立：每支箭的最高分是 10 分，平均箭值的最高分也就是 10 分。把在近兩年參加所有的比賽的箭的分數除以射出箭的數目，就是平均箭值。

教練：讓我們看看兩者的關係圖吧。



世界排名首 100 的射手這兩年比賽的平均箭值和該射手的拉距磅數（附錄表一）



圖中藍色虛線為迴歸直線

教練：從上圖所見，弓的磅數和平均箭值有明顯的正向關係，因此加磅是十分重要，要有一定的磅數才會有更好的表現。

小迦：原來重磅些真的會影響成績的！

小立：迴歸直線旁的 R^2 是甚麼來的？

教練：那一個代表 x 軸影響 y 軸的程度。男子運動員的 R^2 值較高是因為男子運動員通常較強壯，比女子運動員較少因太大磅數而失準。因此磅數便更能影響到射手的成績。

小立：從男子的迴歸直線圖可見比較低磅數的人的分數大多不集中，相反高磅數的人則十分集中。

小迦：原來愈重磅真的愈好！

教練：我計算了男子組高磅數的人的分數(高分組)與低磅數的人的分數(低分組)的標準差，分別是 0.06233 及 0.2163。

高分組射手姓名	平均箭值	磅數
CHOI YONGHEE	9.73	58
ALBERTO BLAZQUEZ	9.67	58
RODOLFO GONZALEZ	9.65	58
SEBASTIEN PEINEAU	9.68	59
DOMAGOJ BUDEN	9.66	59
MARIO VAVRO	9.64	59
MIKE SCHLOESSER	9.84	60
REO WILDE	9.76	60
PIERRE JULIEN DELOCHE	9.75	60
MARTIN DAMSBO	9.7	60
平均：	9.708	
標準差：	0.06233	

低分組射手姓名	平均箭值	磅數
KU BONCHAN	9.57	46
KIM WOO JIN	9.49	46
LEE SEUNGYUN	9.48	46
IM DONG HYUN	9.45	42
JAYANTA TALUKDAR	9.32	42
MARCUS D'ALMEIDA	9.13	45
ALEXANDER KOZHIN	9.13	44
PABLO ACHA	9.05	46
ANDREAS GSTOETTNER	9.05	44
KACPER SIERAKOWSKI	9.01	46
平均：	9.268	
標準差：	0.2163	

教練：這二十人的能力都相當，最大分別則是他們得分和弓的磅數。上表可見磅數高的一組的平均箭值較高，而標準差遠低於低分組射手。相反磅數低的人的平均分較低，而標準差相對較大。可見高磅數能令射手的分數更高且更穩定。

教練：再深入看看世界首 50 名射手每箭的標準差(附錄表五)，他們的標準差與分數的關係能以皮耳森積差相關係數(r)來表達，數值為 -0.9527，相關性極高，代表他們的標準差越小，分數便越高。標準差能看到射手的穩定性。為了更好表達穩定性，我們試定義一個穩定值，如下

$$\text{穩定值} = \frac{\text{平均箭值}}{\text{標準差}}$$

由此，平均箭值愈高，標準差越小，穩定值便愈大。



姓名	分數	平均箭值	標準差	穩定值
STEPHAN HANSEN	712	9.89	0.3165	31.24
KIM JONGHO	711	9.88	0.333	29.65
PIERRE-JULIEN DELOCHE	710	9.86	0.3483	28.31
OH JINHYEK	670	9.31	0.8331	11.17
ZHANG DAN	668	9.28	0.7505	12.36
RYAN TYACK	667	9.26	0.7314	12.67
SUGIMOTO TOMOMI	660	9.17	0.8049	11.39
OIBEK SAIDIYEV	659	9.15	0.8502	10.77
KIKUCHI HIDEKI	658	9.14	0.9539	9.58

由上表可見，穩定值愈大，代表射手的水平高，且表現穩定。

小立：穩定性實在重要，除了磅數以外，還有甚麼會影響到我的穩定發揮呢？

教練：比賽場地的溫度和風速都會有影響的，還有你自己的心理質素！以 Mexico City 2017 Hyundai World Archery Championships(16/10/2017)為例 (附錄表二)

再看看男子和女子世界排名第一的選手前五次比賽的總分與標準差。

男子姓名	分數	平均箭值	標準差	穩定值
KIM WOOJIN :	695	9.65	0.5607	17.22
	679	9.43	0.6677	14.12
	689	9.57	0.5522	17.33
	683	9.49	0.6278	15.11
	695	9.65	0.4794	20.14
標準差：	7.15542			

女子姓名	分數	平均箭值	標準差	穩定值
CHANG HYE JIN	683	9.49	0.6919	13.71
	672	9.33	0.628	14.86
	683	9.49	0.7871	12.05
	673	9.35	0.7536	12.4
	666	9.25	0.7645	12.1
標準差：	7.4364			

教練：我們能看見兩人分數的標準差十分小，而且分數高，穩定值也很大，可見成為優秀的射手必定要高分並且穩定。

教練：還要留意，當天比賽的溫度為 14°C 至 24°C，相對很多國家溫度比較冷。當天的風速為大約 4-6mph，風速不高，對射手的影響較小。從上圖可見,每位射手的表現頗為穩定，而每位射手的差的平均為+0.05，可見整體分數比預期好。但我們來看看韓國隊的表現。

姓名	分數/720	該室外賽 平均箭值 ¹	室外賽 平均箭值	差 ²	磅數	國籍
IM DONG HYUN	674	9.36	9.45	-0.09	42	KOR
KIM WOJIN	672	9.33	9.49	-0.16	46	KOR
OH JIN HYEK	670	9.31	9.5	-0.19	46	KOR
KANG CHAE YOUNG	684	9.5	9.3	0.2	44.5	KOR
CHANG HYE JIN	683	9.49	9.34	0.15	40	KOR
CHOI MISUN	671	9.32	9.33	-0.01	42.7	KOR
KIM JONGHO	711	9.88	9.79	0.09	59.5	KOR
SONG YUN SOO	694	9.64	9.59	0.05	48	KOR
CHOI BOMIN	689	9.57	9.58	-0.01	49.3	KOR

教練：九個人中有五人都比平均表現差。當年十月，韓國的氣溫也是大約 12°C-20°C，相比墨西哥當時 14°C-24°C 的溫度相差不遠，可見氣溫不是導致他們失準的主要原因

小迦：對啊！兩國當天的溫度相差不是太大呢！

教練：可能是當天是正式比賽的第一天，經過很長的路程，尤其韓國到墨西哥十多小時的旅程，射手要適應時差，相反其他國家射手離墨西哥較近。坐了十多小時的飛機也會影響到心情和心理質素，這也是令射手不能發揮最好的表現的一個因素。所以小立你代表香港出去參加大型的國際賽事時要注意，調節好自己的心態和情緒啊！

小立：知道！剛才韓國隊不是受溫度影響，哪其實溫度會不會

影響到我的表現？

教練：當然會，就以 SALT LAKE CITY 2017 HYUNDAI ARCHERY WORLD CUP STAGE 3(20/6/2017)為例(附錄表三)

姓名	分數/720	該室外賽 平均箭值 ¹	室外賽 平均箭值	差 ²	磅數	國籍
SJEF VAN DEN BERG	670	9.3	9.36	-0.06	48	NED
BRADY ELLISON	669	9.29	9.36	-0.07	48	USA
OH JIN HYEK	661	9.18	9.5	-0.32	46	KOR
MARCUS D'ALMEIDA	645	8.96	9.13	-0.17	45	BRA
LEI CHIEN YING	642	8.92	8.96	-0.04	38	TPE
AIDA ROMAN	640	8.89	9.08	-0.19	46	MEX
PIERRE-JULIEN DELOCHE	691	9.6	9.75	-0.15	60	FRA
ALBERTO BLAZQUEZ	688	9.56	9.67	-0.11	58	ESP
MARTIN DAMSBO	685	9.51	9.7	-0.19	60	DEN

教練：當天的風速為大約 5-9mph，風速不高。當天的溫度為 22°C 至 30°C，相對很多國家溫度比較熱。對射手的影響較大。上表舉出了幾個該場賽事的數據，因為溫度的關係，有大部分射手的得分都比平常低。而當天有 76% 的人的分數比平常低，因此場地的環境也會影響到射手的表現。

小立：除此之外，我想問其實室外和室內有沒有不同？

教練：當然有！在室外比賽會受風向、風速、光照方向等影響，而室內就沒有這些影響，而且室內比賽只是 18 米。因近距離較容易，所以分數可能會突然增加，因此根據射手遠距離得分而

作的迴歸直線圖不能全套用在該射手的近距離上，又看看室內賽各射手的表現。

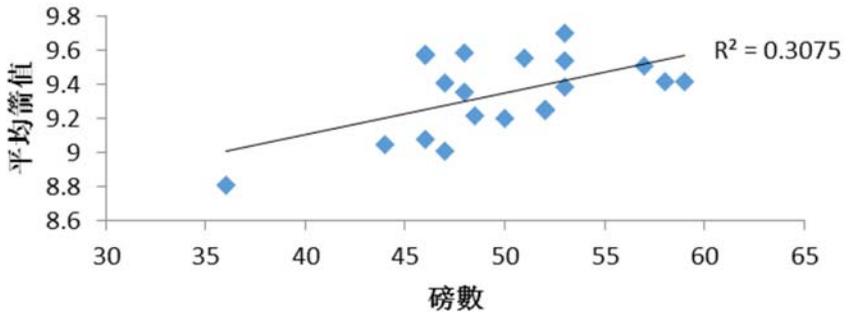
附錄表四列出了一場國際室內賽(LAS VEGAS 2018 INDOOR ARCHERY WORLD CUP STAGE 4)各射手的成績。

教練：從表四可見，所有射手在室內賽的表現都比室外賽高，證明室內賽比室外賽容易，各射手能有較穩定的表現。

教練：看看室內賽的迴歸線圖。



室內賽迴歸線圖



教練：我們也能從上圖可見，室內賽的磅數與分數也有正向的關係。

小迦：咦，為甚麼室內賽的 R^2 會比室外賽的低呢？

教練：因為室內賽不會受環境因素而影響，因此磅數比較低的

人在室外賽分數不高，但在室內賽也能表現優秀。而且室內賽的距離較近，不需要高磅數，只要射手有好的動作便能有好的成績。所以 R^2 反映出室內賽分數被磅數影響的程度不高。

教練：對於射手的表現，磅數有一定的影響，但同時比賽的距離，射手的心理質素和比賽的環境都有一定的影響。射箭講求的是一致度，每次比賽場地都不同，射手需不受環境的影響，才能有更好的發揮。

小迦：原來是這樣，今天真學到很多東西！

小立：幾個月後我便要出國參加比賽了，我要努力練習好自己的動作並且要加磅數，讓比賽時有更好更穩定的表現！

教練：小立，你不要太緊張，不少人會緊張自己的成績，擔心自己不能打敗對手甚至奪獎，令自己不能每一箭都輕鬆射出，表現每況愈下。也不要再在比賽前練習過多，以免受傷。出國前記住要檢查你的器材，若器材出現問題會影響發揮，甚至不能完成比賽。

小立：好的！

參考資料：

1. <https://worldarchery.org/>
2. <https://weatherspark.com/d/5674/10/15/AverageWeather-on-October-15-in-Mexico-City-Mexico#Sections-Wind>
3. <https://weatherspark.com/m/142033/10/Average-Weather-in-October-in-Seoul-South-Korea>
4. <https://www.timeanddate.com/weather/usa/salt-lake-city/historic?month=6&year=2017>
5. https://en.wikipedia.org/wiki/Target_archery

附錄

表一：世界排名首 100 的射手這兩年比賽的平均箭值和該射手的拉距磅數

射手姓名(女子)	平均箭值	磅數
SARA LOPEZ	9.7	53
SONG YUN SOO	9.59	48
SARAH PRIEELS	9.58	46
CHOI BOMIN	9.58	49.3
ANDREA MARCOS	9.56	51
LINDA OCHOA ANDERSON	9.54	53
KRISTINA HEIGENHAUSER	9.51	59
NATALIA AVDEEVA	9.51	57
MARIIA VINOGRADOVA	9.51	52
JYOTHI SUREKHA VENNAM	9.5	55
KIM YUNHEE	9.5	45
DANELLE WENTZEL	9.49	43
AMELIE SANCENOT	9.48	51
NORA VALDEZ	9.47	53
ALEJANDRA USQUIANO	9.46	56
JANINE MEISSNER	9.45	46
FERNANDA ZEPEDA	9.44	42
INGE VAN CASPEL	9.44	51
MARIYA SHKOLNA	9.42	59
ERIKA DAMSBO	9.42	58
DELLIE THREESYADINDA	9.42	55
STEPHANIE SALINAS	9.41	47
CHANG HYE YIN	9.34	40
CHOI MISUN	9.33	42.7

射手姓名(男子)	平均箭值	磅數
MIKE SCHLOSSER	9.84	60
KIM JONGHO	9.79	59.5
REO WILDE	9.76	60
PIERRE JULIEN DELOCHE	9.75	60
CHOI YONGHEE	9.73	58
MARTIN DAMSBO	9.7	60
SEBASTIEN PEINEAU	9.68	59
ALBERTO BLAZQUEZ	9.67	58
DOMAGOJ BUDEN	9.66	59
RODOLFO GONZALEZ	9.65	58
NICO WIENER	9.65	57
MARIO VAVRO	9.64	59
KIM TAEYOON	9.63	58
MADS HAUGSETH	9.62	60
KU BONCHAN	9.57	46
OH JIN HYEK	9.5	46
KIM WOO JIN	9.49	46
LEE SEUNGYUN	9.48	46
IM DONG HYUN	9.45	42
SJEF VAN DEN BERG	9.39	53
ZACH GARRETT	9.37	52
BRADY ELLISON	9.36	48
JAYANTA TALUKDAR	9.32	42
MAURO NESPOLI	9.3	52

KANG CHAE YOUNG	9.3	44.5
KI BO BAE	9.26	41.4
DEEPIKA KUMARI	9.24	40
TAN YA TING	9.2	39
AIDA ROMAN	9.08	46
ELENA RICHTER	9.06	39.5
KSENIA PEROVA	9.04	38
ALEJANDRA VALENCIA	9.03	44
AUDREY ADICEOM	9	40
LIN SHIH CHIA	8.98	39.5
LEI CHIEN YING	8.96	38
LUCILLA BOARI	8.9	43
YASEMIN ANAGOZ	8.87	42
ALEXANDRA MIRCA	8.82	38
ALEXANDRA LONGOVA	8.82	40
BRYONY PITMAN	8.81	36
MARIANA AVITIA	8.81	42
GULNAZ COSKUN	8.79	38
MONIKA SAREN	8.77	40
GUENDALINA SARTORI	8.72	38

CRISPIN DUENAS	9.26	52
DAVID PASQUALUCCI	9.23	50
TAYLOR WORTH	9.22	48.5
PATRICK HUSTON	9.2	50
KHAIRUL ANUAR MOHAMAD	9.19	45
MARCO GALIAZZO	9.15	48
MARCUS D'ALMEIDA	9.13	45
ALEXANDER KOZHIN	9.13	44
JAKE KAMINSKI	9.1	48
JUAN RENE SERRANO	9.09	50
PABLO ACHA	9.05	46
ANDREAS GSTOETTNER	9.05	44
MAXIMILIAN WECKMUELLER	9.01	47
KACPER SIERAKOWSKI	9.01	46

表二：Mexico City 2017 Hyundai World Archery Championships (16/10/2017)

姓名	分數 /720	該室外賽 平均箭值 ¹	室外賽 平均箭值	差 ²	磅數	國籍
SJEF VAN DEN BERG	676	9.39	9.36	+0.03	48	NED
IM DONG HYUN	674	9.36	9.45	-0.09	42	KOR
MAURO NESPOLI	672	9.33	9.3	+0.03	52	ITA
KIM WOJIN	672	9.33	9.49	-0.16	46	KOR
OH JIN HYEK	670	9.31	9.5	-0.19	46	KOR
RYAN TYACK	667	9.26	9.25	+0.01	52	AUS
DAVID PASQUALUCCI	665	9.24	9.23	+0.01	50	ITA
CRISPIN DUENAS	663	9.21	9.26	-0.05	52	CAN
BRADY ELLISON	661	9.18	9.36	-0.18	48	USA
MARCUS D'ALMEIDA	655	9.1	9.13	-0.03	45	BRA
KANG CHAE YOUNG	684	9.5	9.3	+0.2	44.5	KOR
CHANG HYE JIN	683	9.49	9.34	+0.15	40	KOR
TAN YA-TING	680	9.44	9.2	+0.24	39	TPE
CHOI MISUN	671	9.32	9.33	-0.01	42.7	KOR
ALEJANDRA VALENCIA	658	9.14	9.03	+0.11	44	MEX
MARIANA AVITIA	653	9.07	8.81	+0.26	42	MEX
ALEXANDRA LONGOVA	645	8.96	8.82	+0.14	40	SVK
ALEXANDRA MIRCA	643	8.93	8.82	+0.11	38	MDA
BRYONY PITMAN	635	8.82	8.81	+0.01	36	GBR
KIM JONGHO	711	9.88	9.79	+0.09	59.5	KOR
PIERRE-JULIEN DELOCHE	710	9.86	9.75	+0.11	60	FRA
SEBASTIEN PEINEAU	704	9.78	9.75	+0.03	60	FRA
MARIO VAVRO	702	9.75	9.64	+0.11	59	CRO
MARTIN DAMSBO	699	9.71	9.7	+0.01	60	DEN
NICO WIENER	695	9.65	9.62	+0.03	60	AUT
ALBERTO BLAZQUEZ	691	9.6	9.67	-0.07	58	ESP
MAXIMILIANO CABRAL FAVORETO	671	9.32	9.01	+0.31	47	BRA
LINDA OCHOA-ANDERSON	701	9.74	9.54	+0.2	53	MEX
SARAH PRIEELS	699	9.71	9.58	+0.13	46	BEL
SONG YUN SOO	694	9.64	9.59	+0.05	48	KOR
SARA LOPEZ	693	9.63	9.7	-0.07	53	COL
NATALIA AVDEEVA	691	9.6	9.51	+0.09	57	RUS
ALEJANDRA USQUIANO	691	9.6	9.46	+0.14	56	COL

ANDREA MARCOS	690	9.58	9.56	+0.02	51	ESP
CHOI BOMIN	689	9.57	9.58	-0.01	49.3	KOR
ERIKA DAMSBO	683	9.49	9.42	+0.07	58	DEN

表三：SALT LAKE CITY 2017 HYUNDAI ARCHERY WORLD CUP
STAGE 3(20/6/2017)

姓名	分數 /720	該室外賽 平均箭值 ¹	室外賽 平均箭值	差 ²	磅數	國籍
IM DONG HYUN	680	9.44	9.45	-0.01	42	KOR
KIM WOOJIN	679	9.43	9.49	-0.06	46	KOR
KIM JONGHO	673	9.35	9.79	-0.44	59.5	KOR
DAVID PASQUALUCCI	670	9.3	9.23	+0.07	50	ITA
SJEF VAN DEN BERG	670	9.3	9.36	-0.06	48	NED
BRADY ELLISON	669	9.29	9.36	-0.07	48	USA
MAURO NESPOLI	668	9.28	9.3	-0.02	52	ITA
OH JIN HYEK	661	9.18	9.5	-0.32	46	KOR
PATRICK HUSTON	661	9.18	9.2	-0.02	50	GBR
KHAIRUL ANUAR MOHAMAD	660	9.17	9.19	-0.02	45	MAS
MARCO GALIAZZO	657	9.13	9.15	-0.02	48	ITA
CRISPIN DUENAS	652	9.06	9.26	-0.2	52	CAN
MARCUS D'ALMEIDA	645	8.96	9.13	-0.17	45	BRA
CHANG HYE JIN	683	9.49	9.34	+0.15	40	KOR
KI BO BAE	673	9.35	9.26	+0.09	41.4	KOR
TAN YA-TING	666	9.25	9.2	+0.05	39	TPE
CHOI MISUN	662	9.19	9.33	-0.14	42.7	KOR
KANG CHAE YOUNG	661	9.18	9.3	-0.12	44.5	KOR
ALEJANDRA VALENCIA	659	9.15	9.03	+0.12	44	MEX
LIN SHIH CHIA	646	8.97	8.98	-0.01	39.5	TPE
LEI CHIEN YING	642	8.92	8.96	-0.04	38	TPE
AIDA ROMAN	640	8.89	9.08	-0.19	46	MEX
LUCILLA BOARI	640	8.89	8.9	-0.01	43	ITA
BRYONY PITMAN	638	8.86	8.81	+0.05	36	GBR
REO WILDE	704	9.78	9.76	+0.02	60	USA
KIM JONGHO	704	9.78	9.79	-0.01	59.5	KOR

MIKE SCHLOESSER	703	9.76	9.84	-0.08	60	NED
RODOLFO GONZALEZ	696	9.67	9.65	+0.02	58	MEX
DOMAGOJ BUDEN	693	9.63	9.66	-0.03	59	CRO
KIM TAEYOON	691	9.6	9.63	-0.03	58	KOR
PIERRE-JULIEN DELOCHE	691	9.6	9.75	-0.15	60	FRA
CHOI YONGHEE	690	9.58	9.73	-0.15	58	KOR
ALBERTO BLAZQUEZ	688	9.56	9.67	-0.11	58	ESP
MARTIN DAMSBO	685	9.51	9.7	-0.19	60	DEN
SEBASTIEN PEINEAU	684	9.5	9.75	-0.25	60	FRA
ANDREA MARCOS	694	9.64	9.56	+0.08	51	ESP
SARA LOPEZ	692	9.61	9.7	-0.09	53	COL
SONG YUN SOO	690	9.58	9.59	-0.01	48	KOR
LINDA OCHOA-ANDERSON	689	9.57	9.54	+0.03	53	MEX
SARAH PRIEELS	687	9.54	9.58	-0.04	46	BEL
KIM YUNHEE	686	9.53	9.5	+0.03	45	KOR
FERNANDA ZEPEDA	679	9.43	9.44	-0.01	42	MEX
CHOI BOMIN	679	9.43	9.58	-0.15	49.3	KOR
AMELIE SANCENOT	668	9.28	9.48	-0.2	51	FRA
STEPHANIE SALINAS	661	9.18	9.41	-0.23	47	MEX

表四：LAS VEGAS 2018 INDOORARCHERY WORLD CUP STAGE 4

姓名	分數 /600	室內賽 平均箭值 ¹	室外賽 平均箭值	差 ²	磅數	國籍
ANDREAS GSTOETTNER	591	9.85	9.05	+0.8	44	AUT
RYAN TYACK	590	9.83	9.25	+0.58	52	AUS
TAYLOR WORTH	588	9.8	9.22	+0.58	48.5	AUS
PATRICK HUSTON	588	9.8	9.2	+0.6	50	GBR
BRADY ELLISON	587	9.78	9.36	+0.42	48	USA
CRISPIN DUENAS	586	9.77	9.26	+0.51	52	CAN
SJEF VAN DEN BERG	584	9.73	9.39	+0.34	53	NED
MAXIMILIAN WECKMUELLER	581	9.68	9.01	+0.67	47	GER
KU BONCHAN	577	9.62	9.57	+0.05	46	KOR
AIDA ROMAN	571	9.62	9.08	+0.54	46	MEX
BRYONY PITMAN	561	9.35	8.81	+0.54	36	GBR
SARAH PRIEELS	598	9.97	9.58	+0.39	46	BEL
SARA LOPEZ	598	9.97	9.7	+0.27	53	COL
LINDA OCHOA-ANDERSON	598	9.97	9.54	+0.43	53	USA
ANDREA MARCOS	598	9.97	9.56	+0.41	51	ESP
NATALIA AVDEEVA	597	9.95	9.51	+0.44	57	RUS
SONG YUN SOO	597	9.95	9.59	+0.36	48	KOR
MARIYA SHKOLNA	596	9.93	9.42	+0.51	59	POL
STEPHANIE SALINAS	595	9.92	9.41	+0.51	47	MEX
ERIKA DAMSBO	595	9.92	9.42	+0.5	58	DEN

表五：Mexico City 2017 Hyundai World Archery Championships
首 50 人每一箭的標準差

姓名	分數	標準差	穩定值
STEPHAN HANSEN	712	0.3165	31.24
KIM JONGHO	711	0.3330	29.65
PIERRE-JULIEN DELOCHE	710	0.3483	28.31
SERGIO PAGNI	710	0.3483	28.31
BRADEN GELLENTHIEN	709	0.3623	27.18
DOMAGOJ BUDEN	707	0.3873	25.35
LUC MARTIN	707	0.3873	25.35
ABHISHEK VERMA	707	0.3873	25.35
MIKE SCHLOESSER	706	0.4324	22.68
PATRICK ROUX	706	0.3985	24.61
LINDA OCHOA-ANDERSON	701	0.4438	21.94
SARAH PRIEELS	699	0.4875	19.91
SARAH SONNICHSEN	699	0.4577	21.21
JYOTHI SUREKHA VENNAM	698	0.5474	17.71
YESIM BOSTAN	697	0.5261	18.40
TOJA ELLISON	696	0.5035	19.20
SONG YUN SOO	694	0.5888	16.37
SO CHAEWON	694	0.5388	17.89
KRISTINA HEIGENHAUSER	693	0.5676	16.96
TRISHA DEB	693	0.5422	17.75
KANG CHAE YOUNG	684	0.6051	15.70
CHANG HYE JIN	683	0.6919	13.71
TAN YA-TING	680	0.6690	14.12

SJEF VAN DEN BERG	676	0.7609	12.34
IM DONG HYUN	674	0.7374	12.69
MAURO NESPOLI	672	0.7316	12.76
KIM WOJIN	672	0.6280	14.86
CHOI MISUN	671	0.7473	12.47
OH JINHYEK	670	0.8331	11.17
ZHANG DAN	668	0.7505	12.36
RYAN TYACK	667	0.7314	12.67
KRISTINE ESEBUA	666	0.7505	12.33
DAVID PASQUALUCCI	665	0.7960	11.60
ARSALAN BALDANOV	665	0.7218	12.80
FURUKAWA TAKAHARU	664	0.7358	12.53
LIN YU-HSUAN	664	0.7165	12.87
LISA UNRUH	664	0.6965	13.24
CRISPIN DUENAS	663	0.8548	10.77
FLORIAN KAHLLUND	663	0.8038	11.46
MACKENZIE BROWN	663	0.8711	10.57
WEI CHUN-HENG	662	1.0433	8.81
ARPAD BANDA	662	0.8824	10.42
ATANU DAS	661	0.8933	10.28
BRADY ELLISON	661	0.6986	13.14
SUGIMOTO TOMOMI	660	0.8049	11.39
OIBEK SAIDIYEV	659	0.8502	10.77
KIKUCHI HIDEKI	658	0.9539	9.58
MAXIMILIAN WECKMUELLER	658	0.8770	10.42
WANG WENXUAN	658	0.8443	10.82
PATRICK HUSTON	658	0.8608	10.62
相關係數：		r= -0.9527	

1. 這為該場比賽該射手的平均箭值
2. 這為該比賽與該射手近兩年室外賽的平均箭值的得分的差異，+為比賽表現比平時好

優異作品:

真相，盡在拉斯莫方程之中

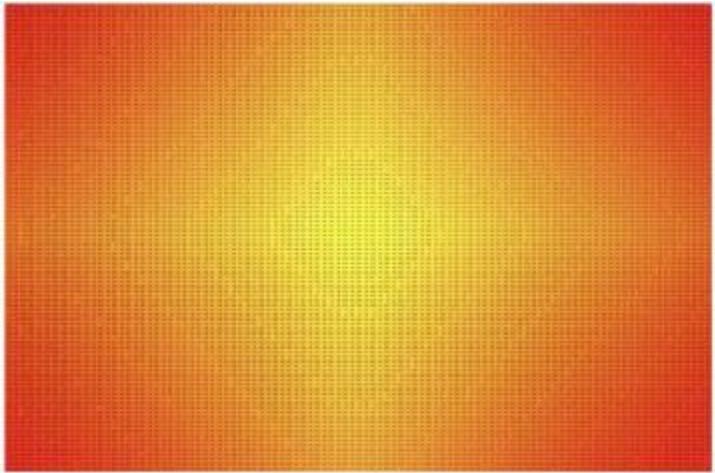
學校名稱：香港培正中學

學生姓名：曾樂勤、張燁熙、李皓澄

指導教師：梁偉雄老師

摘要：連環兇殺案，令人聞風喪膽；疫症肆虐，令人生活憂心忡忡，這時警察正急於逮捕犯人，醫生期望追查出疫症源頭。這份統計習作將會利用金姆羅斯莫於 1996 年發明的拉斯莫方程，嘗試應用數學統計的角度於犯罪學和尋找疫症源頭上。

真相，盡在拉斯莫方程之中



引言

恐怖的事情接連發生，一定引來人們的恐慌，人們一定急於知道真相，尋找來源；連環殺人犯四處殺人，令市民聞風喪膽，警察急於逮捕犯人；疫症肆虐，人民生於憂慮死於恐懼，政府忙於尋找疫症根源。不知能否利用數學的方法，揭曉真相呢？

這份報告將會利用金姆羅斯莫於 1996 年發明的拉斯莫方程，嘗試利用數學的角度於犯罪學和尋找疫症源頭上。

故事是這樣的……

熊貓: 嘉俊，你是否想修讀犯罪學？我常看到你沉迷於閱讀這個連環殺人犯的資料。

嘉俊: 他是彼得·薩特克利夫，英國著名的連環殺手，謀殺多人。我正在閱讀有關他的資料，並且從不同方向入手，例如犯人性格、殺人時間、地點、手法等，進行分析，研究連環殺人犯的犯案動機。

熊貓: 我明白殺人時間和手法等與分析犯人心理有關係，而心理則影響行為，但地點也有關嗎？殺人地點不是按難易度，或人流少等因素而決定嗎？

樂勤: 那就不對了，一向都有犯罪學家認為犯案地點和犯人心理與其慣性行為有關。在 1996 年，一位前美國警察，金姆拉斯姆建立出一條方程式—拉斯莫方程，顯示出犯人過往犯案地點、與犯案機會、犯人有可能居住範圍之間的關係。證明犯人藏身

之處可以從犯人過往的犯案行為中推演出來的，亦代表犯人藏身之處和犯案地點有關係！

熊貓: 竟然如此！噫？我們是否可以利用彼得的犯案數據，推演他的藏身之處？

樂勤: 對！而這裡是我們要計算的方程，利用犯人過往犯案地點、與犯案機會計算犯人有可能居住範圍

$$p_{i,j} = k \sum_{n=1}^{(\text{total crimes})} \left[\underbrace{\frac{\phi_{ij}}{(|X_i - x_n| + |Y_j - y_n|)^f}}_{1^{\text{st}} \text{ term}} + \underbrace{\frac{(1 - \phi_{ij})(B^{g-f})}{(2B - |X_i - x_n| - |Y_j - y_n|)^g}}_{2^{\text{nd}} \text{ term}} \right]$$

$$\text{where } \phi_{ij} = \begin{cases} 1, & \text{if } (|X_i - x_n| + |Y_j - y_n|) > B \\ 0, & \text{else} \end{cases}$$

我們假設犯人不因自己特殊的喜好或任何心理影響選擇犯案地點，而且犯人不會在自己的居住地點附近犯罪。B 是緩衝帶的半徑，f, g 是常數，這些 B, f, g 參數都是根據數據集的大小、特性來訂立，沒有一個特定的定義方法。例如緩衝帶的半徑因當年的情況而定，如果犯人利用車輛逃跑，會逃得更遠，半徑要設得較大，如果犯人只是徒步離開，不會走遠，半徑要設得較小。

熊貓: 但如何設 f 和 g 呢？

樂勤: f 和 g 是要之後控制和調整，令之後顯示的結果合理。

(尋找關於彼得的資料)

犯案日期	犯案地點		搜索範圍		參數	
	緯度	經度	緯度	經度		
30/10/1975	53.8178	-1.5428	54.2	-2.3	B	0.5
20/1/1976	53.8082	-1.5311	54.07	-2.165	f	0.25
1977/5/2	53.8334	-1.002	53.94	-2.03	g	0.25
23/4/2017	53.8109	-1.7633	53.81	-1.895		
26/6/1977	53.8179	-1.5325	53.68	-1.76		
1977/1/10	53.4324	-2.2506	53.55	-1.625		
21/1/1978	53.8001	-1.7721	53.42	-1.49		
31/1/1978	53.6544	-1.78	53.29	-1.355		
16/5/1978	53.4599	-2.2225	53.16	-1.22		
1979/4/4	53.7117	-1.8736	53.03	-1.085		
1979/2/9	53.79	-1.764	52.9	-0.95		
20/8/1980	53.8085	-1.6715				
17/11/1980	53.8228	-1.5781				

表 1

樂勤: 首先輸入你給我的經緯度,表 1 左面的是不同案發地點的經緯度,中間是簡化座標,用來控制搜索範圍,而右面是計算參數。我們把彼得犯案地點的經緯度輸入在地圖上,圖 1 是英國一部分的地圖,是我們要計算的範圍,黑點是犯案地點。

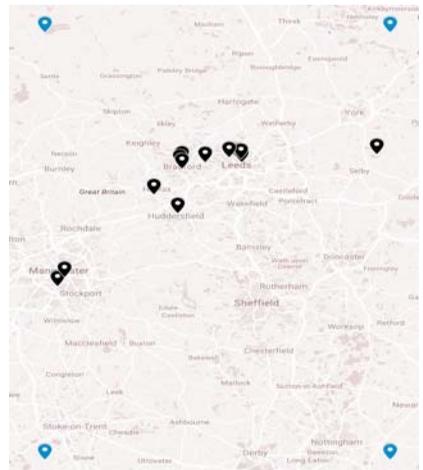


圖 1

表 2 左邊計算曼哈頓距離，即是每一個犯案地點對於 B 範圍裡每一點的距離，所以總共有 $13 \times 11 \times 11$ 個數據，而右邊就是利用拉斯莫方程，繼而算 B 範圍中每一點的犯人居住概率：

列	行	距離	曼哈頓距離											利用拉斯莫方程																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																							
			1	2	3	4	5	6	7	8	9	10	11	1	2	3	4	5	6	7	8	9	10	11																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
1	1	0.0000	0.1176	0.2353	0.3529	0.4706	0.5882	0.7059	0.8235	0.9412	1.0588	1.1765	1.2941	1.4118	1.5294	1.6471	1.7647	1.8824	1.9999	2.1176	2.2353	2.3529	2.4706	2.5882	2.7059	2.8235	2.9412	3.0588	3.1765	3.2941	3.4118	3.5294	3.6471	3.7647	3.8824	3.9999	4.1176	4.2353	4.3529	4.4706	4.5882	4.7059	4.8235	4.9412	5.0588	5.1765	5.2941	5.4118	5.5294	5.6471	5.7647	5.8824	5.9999	6.1176	6.2353	6.3529	6.4706	6.5882	6.7059	6.8235	6.9412	7.0588	7.1765	7.2941	7.4118	7.5294	7.6471	7.7647	7.8824	7.9999	8.1176	8.2353	8.3529	8.4706	8.5882	8.7059	8.8235	8.9412	9.0588	9.1765	9.2941	9.4118	9.5294	9.6471	9.7647	9.8824	9.9999	10.1176	10.2353	10.3529	10.4706	10.5882	10.7059	10.8235	10.9412	11.0588	11.1765	11.2941	11.4118	11.5294	11.6471	11.7647	11.8824	11.9999	12.1176	12.2353	12.3529	12.4706	12.5882	12.7059	12.8235	12.9412	13.0588	13.1765	13.2941	13.4118	13.5294	13.6471	13.7647	13.8824	13.9999	14.1176	14.2353	14.3529	14.4706	14.5882	14.7059	14.8235	14.9412	15.0588	15.1765	15.2941	15.4118	15.5294	15.6471	15.7647	15.8824	15.9999	16.1176	16.2353	16.3529	16.4706	16.5882	16.7059	16.8235	16.9412	17.0588	17.1765	17.2941	17.4118	17.5294	17.6471	17.7647	17.8824	17.9999	18.1176	18.2353	18.3529	18.4706	18.5882	18.7059	18.8235	18.9412	19.0588	19.1765	19.2941	19.4118	19.5294	19.6471	19.7647	19.8824	19.9999	20.1176	20.2353	20.3529	20.4706	20.5882	20.7059	20.8235	20.9412	21.0588	21.1765	21.2941	21.4118	21.5294	21.6471	21.7647	21.8824	21.9999	22.1176	22.2353	22.3529	22.4706	22.5882	22.7059	22.8235	22.9412	23.0588	23.1765	23.2941	23.4118	23.5294	23.6471	23.7647	23.8824	23.9999	24.1176	24.2353	24.3529	24.4706	24.5882	24.7059	24.8235	24.9412	25.0588	25.1765	25.2941	25.4118	25.5294	25.6471	25.7647	25.8824	25.9999	26.1176	26.2353	26.3529	26.4706	26.5882	26.7059	26.8235	26.9412	27.0588	27.1765	27.2941	27.4118	27.5294	27.6471	27.7647	27.8824	27.9999	28.1176	28.2353	28.3529	28.4706	28.5882	28.7059	28.8235	28.9412	29.0588	29.1765	29.2941	29.4118	29.5294	29.6471	29.7647	29.8824	29.9999	30.1176	30.2353	30.3529	30.4706	30.5882	30.7059	30.8235	30.9412	31.0588	31.1765	31.2941	31.4118	31.5294	31.6471	31.7647	31.8824	31.9999	32.1176	32.2353	32.3529	32.4706	32.5882	32.7059	32.8235	32.9412	33.0588	33.1765	33.2941	33.4118	33.5294	33.6471	33.7647	33.8824	33.9999	34.1176	34.2353	34.3529	34.4706	34.5882	34.7059	34.8235	34.9412	35.0588	35.1765	35.2941	35.4118	35.5294	35.6471	35.7647	35.8824	35.9999	36.1176	36.2353	36.3529	36.4706	36.5882	36.7059	36.8235	36.9412	37.0588	37.1765	37.2941	37.4118	37.5294	37.6471	37.7647	37.8824	37.9999	38.1176	38.2353	38.3529	38.4706	38.5882	38.7059	38.8235	38.9412	39.0588	39.1765	39.2941	39.4118	39.5294	39.6471	39.7647	39.8824	39.9999	40.1176	40.2353	40.3529	40.4706	40.5882	40.7059	40.8235	40.9412	41.0588	41.1765	41.2941	41.4118	41.5294	41.6471	41.7647	41.8824	41.9999	42.1176	42.2353	42.3529	42.4706	42.5882	42.7059	42.8235	42.9412	43.0588	43.1765	43.2941	43.4118	43.5294	43.6471	43.7647	43.8824	43.9999	44.1176	44.2353	44.3529	44.4706	44.5882	44.7059	44.8235	44.9412	45.0588	45.1765	45.2941	45.4118	45.5294	45.6471	45.7647	45.8824	45.9999	46.1176	46.2353	46.3529	46.4706	46.5882	46.7059	46.8235	46.9412	47.0588	47.1765	47.2941	47.4118	47.5294	47.6471	47.7647	47.8824	47.9999	48.1176	48.2353	48.3529	48.4706	48.5882	48.7059	48.8235	48.9412	49.0588	49.1765	49.2941	49.4118	49.5294	49.6471	49.7647	49.8824	49.9999	50.1176	50.2353	50.3529	50.4706	50.5882	50.7059	50.8235	50.9412	51.0588	51.1765	51.2941	51.4118	51.5294	51.6471	51.7647	51.8824	51.9999	52.1176	52.2353	52.3529	52.4706	52.5882	52.7059	52.8235	52.9412	53.0588	53.1765	53.2941	53.4118	53.5294	53.6471	53.7647	53.8824	53.9999	54.1176	54.2353	54.3529	54.4706	54.5882	54.7059	54.8235	54.9412	55.0588	55.1765	55.2941	55.4118	55.5294	55.6471	55.7647	55.8824	55.9999	56.1176	56.2353	56.3529	56.4706	56.5882	56.7059	56.8235	56.9412	57.0588	57.1765	57.2941	57.4118	57.5294	57.6471	57.7647	57.8824	57.9999	58.1176	58.2353	58.3529	58.4706	58.5882	58.7059	58.8235	58.9412	59.0588	59.1765	59.2941	59.4118	59.5294	59.6471	59.7647	59.8824	59.9999	60.1176	60.2353	60.3529	60.4706	60.5882	60.7059	60.8235	60.9412	61.0588	61.1765	61.2941	61.4118	61.5294	61.6471	61.7647	61.8824	61.9999	62.1176	62.2353	62.3529	62.4706	62.5882	62.7059	62.8235	62.9412	63.0588	63.1765	63.2941	63.4118	63.5294	63.6471	63.7647	63.8824	63.9999	64.1176	64.2353	64.3529	64.4706	64.5882	64.7059	64.8235	64.9412	65.0588	65.1765	65.2941	65.4118	65.5294	65.6471	65.7647	65.8824	65.9999	66.1176	66.2353	66.3529	66.4706	66.5882	66.7059	66.8235	66.9412	67.0588	67.1765	67.2941	67.4118	67.5294	67.6471	67.7647	67.8824	67.9999	68.1176	68.2353	68.3529	68.4706	68.5882	68.7059	68.8235	68.9412	69.0588	69.1765	69.2941	69.4118	69.5294	69.6471	69.7647	69.8824	69.9999	70.1176	70.2353	70.3529	70.4706	70.5882	70.7059	70.8235	70.9412	71.0588	71.1765	71.2941	71.4118	71.5294	71.6471	71.7647	71.8824	71.9999	72.1176	72.2353	72.3529	72.4706	72.5882	72.7059	72.8235	72.9412	73.0588	73.1765	73.2941	73.4118	73.5294	73.6471	73.7647	73.8824	73.9999	74.1176	74.2353	74.3529	74.4706	74.5882	74.7059	74.8235	74.9412	75.0588	75.1765	75.2941	75.4118	75.5294	75.6471	75.7647	75.8824	75.9999	76.1176	76.2353	76.3529	76.4706	76.5882	76.7059	76.8235	76.9412	77.0588	77.1765	77.2941	77.4118	77.5294	77.6471	77.7647	77.8824	77.9999	78.1176	78.2353	78.3529	78.4706	78.5882	78.7059	78.8235	78.9412	79.0588	79.1765	79.2941	79.4118	79.5294	79.6471	79.7647	79.8824	79.9999	80.1176	80.2353	80.3529	80.4706	80.5882	80.7059	80.8235	80.9412	81.0588	81.1765	81.2941	81.4118	81.5294	81.6471	81.7647	81.8824	81.9999	82.1176	82.2353	82.3529	82.4706	82.5882	82.7059	82.8235	82.9412	83.0588	83.1765	83.2941	83.4118	83.5294	83.6471	83.7647	83.8824	83.9999	84.1176	84.2353	84.3529	84.4706	84.5882	84.7059	84.8235	84.9412	85.0588	85.1765	85.2941	85.4118	85.5294	85.6471	85.7647	85.8824	85.9999	86.1176	86.2353	86.3529	86.4706	86.5882	86.7059	86.8235	86.9412	87.0588	87.1765	87.2941	87.4118	87.5294	87.6471	87.7647	87.8824	87.9999	88.1176	88.2353	88.3529	88.4706	88.5882	88.7059	88.8235	88.9412	89.0588	89.1765	89.2941	89.4118	89.5294	89.6471	89.7647	89.8824	89.9999	90.1176	90.2353	90.3529	90.4706	90.5882	90.7059	90.8235	90.9412	91.0588	91.1765	91.2941	91.4118	91.5294	91.6471	91.7647	91.8824	91.9999	92.1176	92.2353	92.3529	92.4706	92.5882	92.7059	92.8235	92.9412	93.0588	93.1765	93.2941	93.4118	93.5294	93.6471	93.7647	93.8824	93.9999	94.1176	94.2353	94.3529	94.4706	94.5882	94.7059	94.8235	94.9412	95.0588	95.1765	95.2941	95.4118	95.5294	95.6471	95.7647	95.8824	95.9999	96.1176	96.2353	96.3529	96.4706	96.5882	96.7059	96.8235	96.9412	97.0588	97.1765	97.2941	97.4118	97.5294	97.6471	97.7647	97.8824	97.9999	98.1176	98.2353	98.3529	98.4706	98.5882	98.7059	98.8235	98.9412	99.0588	99.1765	99.2941	99.4118	99.5294	99.6471	99.7647	99.8824	99.9999	100.1176	100.2353	100.3529	100.4706	100.5882	100.7059	100.8235	100.9412	101.0588	101.1765	101.2941	101.4118	101.5294	101.6471	101.7647	101.8824	101.9999	102.1176	102.2353	102.3529	102.4706	102.5882	102.7059	102.8235	102.9412	103.0588	103.1765	103.2941	103.4118	103.5294	103.6471	103.7647	103.8824	103.9999	104.1176	104.2353	104.3529	104.4706	104.5882	104.7059	104.8235	104.9412	105.0588	105.1765	105.2941	105.4118	105.5294	105.6471	105.7647	105.8824	105.9999	106.1176	106.2353	106.3529	106.4706	106.5882	106.7059	106.8235	106.9412	107.0588	107.1765	107.2941	107.4118	107.5294	107.6471	107.7647	107.8824	107.9999	108.1176	108.2353	108.3529	108.4706	108.5882	108.7059	108.8235	108.9412	109.0588	109.1765	109.2941	109.4118	109.5294	109.6471	109.7647	109.8824	109.9999	110.1176	110.2353	110.3529	110.4706	110.5882	110.7059	110.8235	110.9412	111.0588	111.1765	111.2941	111.4118	111.5294	111.6471	111.7647	111.8824	111.9999	112.1176	112.2353	112.3529	112.4706	112.5882	112.7059	112.8235	112.9412	113.0588	113.1765	113.2941	113.4118	113.5294	113.6471	113.7647	113.8824	113.9999	114.1176	114.2353	114.3529	114.4706	114.5882	114.7059	114.8235	114.9412	115.0588	115.1765	115.2941	115.4118	115.5294	115.6471	115.7647	115.8824	115.9999	116.1176	116.2353	116.3529	116.4706	116.5882	116.7059	116.8235	116.9412	117

而菱形的中間是曾犯案的地點，菱形外圍夾住最深色的地方是彼得最可能藏身的地方。

熊貓: 原來警察可以用這個方程來縮窄搜索範圍，更快捷把殺手繩之於法。不知拉斯莫方程能否有利於其他方面呢?

嘉俊: 拉斯莫方程是利用不同地方發生同樣事件，推測出一單事情的源頭.....如果用於尋找疫症的源頭呢? 例如我有幾個爆發疫症的地點，能否利用拉斯莫方程來找到病毒源頭，就像找到犯人的居住範圍?

樂勤: 可以呀! 病毒沒有心理質素或個人喜好去選擇傳播的地點，就像拉斯莫方程假設犯人不因自己特殊的喜好或任何心理影響選擇犯案地點，大概共通處歸納如下表，可見切合拉斯莫方程的假設，應用於兩情境有異曲同工之妙! 不如我們把拉斯莫方程應用於當年沙士，先找出當年沙士的爆發疫症地點，再推測沙士的源頭。

疫症傳播與連環殺人案共通處

連環殺人犯	病毒
盡快找到殺人犯居住地點	盡快找到疫症的源頭
不考慮個人因素	沒有選擇性偏好傳播
殺人犯在藏身處一定範圍外犯案	帶菌者在住所一定範圍距離生活

嘉俊: 但如何運用拉斯莫方程中緩衝區的概念呀?

樂勤: 由於人們大部分時間都是點對點來往，例如上班下班、返放學，大部分活動時間都是在工作或上學地點，病毒最大機會傳播的地點是在你的工作或上學地點。不論是地圖上那一點，正常情況下都和自己居住的範圍有一段距離，所以較難在自己的居住地方附近散播病毒，能夠應用緩衝區這個概念。

但如果我們要利用拉斯莫方程找到疫症的源頭，只能採用疫症初期爆發的地方，不能夠是後期爆發的地區。因為在疫症中後期，當疫症已經有一定廣泛程度時，社區中傳播會比起初期的容易。當傳播模式不再是點對點，而是大規模社區交叉感染時，我們一開始的假設便不再合理了！而緩衝區這個概念亦不能被運用了！

嘉俊: 但如何設緩衝帶的半徑？

樂勤: 我們用人們點對點來往的距離作半徑。

嘉俊: 原來如此，那我們快點研究吧！

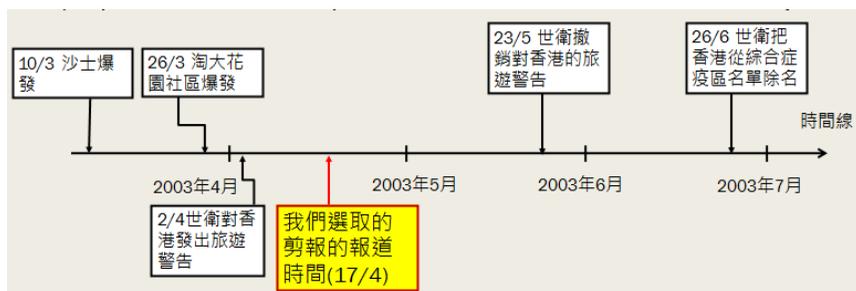
(搜集資料)

熊貓: 我找到沙士爆發初期的新聞報導(2003年4月17日)。

新增證實有居民染疫大廈			
大埔	太湖花園 5 座、翠屏花園 B 座、富亨邨亨昌樓、富亨邨亨耀樓、廣福邨廣禮樓、富善邨善美樓、大元邨泰寧樓	深水埗	美孚新邨蘭秀街 34A 號、李鄭屋邨孝慈樓
沙田	花園城雅芳苑、新城市廣場 3 期 3 座、馬鞍山新港城 F 座、穗禾苑豐裕閣	元朗	雍翠豪園 2 座、天頌苑頌水閣、壽富街 60 號好順福大廈
屯門	寶怡花園 2 座、雅都花園 2 座、翠寧花園 6 座、大興邨興盛樓	東區	英皇道 710 號吉祥大廈前座
中西區	薄扶林道 89 號寶翠園 5 座	黃大仙	竹園南邨趣園樓
		觀塘	藍田康雅苑 C 座
		西貢	將軍澳東港城 2 座

圖 2

嘉俊：我們就由新聞中指出的好幾個地方開始研究，我根據沙士專家委員會報告(Report of the SARS Expert Committee)，整理出以下沙士事件的大概發展時序。



樂勤：為免緩衝帶的半徑太短，不應包括所有地點，所以先從以上的染疫大廈中選取其中五點。就選用感染人數達 329 人的淘大花園，及中西區的寶翠園、黃大仙的趣園樓、深水埗的孝慈樓，觀塘的康雅苑吧！我同樣在表 5 輸入疫症地點的經緯度、搜索範圍和參數，但由於範圍太小，所以這次搜索範圍要分成 100×100 格，來增加準確度。再計算曼哈頓距離，計算 B 範圍中每一點的作為源頭疫症的概率 (表 6A 及 6B)。

	座標		地圖大小		係數	
	緯度	經度	緯度	經度		
淘大花園	22.325	114.217054	22.34591	114.1238	B	0.05
寶翠園	22.28504	114.133334	22.34537	114.125	f	0.1
李慈樓	22.337819	114.161534	22.34484	114.1262	g	0.1
康雅苑	22.304883	114.238223	22.3443	114.1274		
趣園樓	22.344182	114.194353	22.34376	114.1286		
			22.34322	114.1298		
			22.34268	114.131		
			22.34214	114.1322		
			22.3416	114.1334		
			22.34106	114.1346		
			22.34052	114.1358		
			22.33998	114.137		
			22.33944	114.1382		
			22.3389	114.1394		
			22.33836	114.1406		
			22.33782	114.1418		
			22.33728	114.143		
			22.33674	114.1442		
			22.3362	114.1454		
			22.33566	114.1466		
			22.33512	114.1478		
			22.33459	114.149		
			22.33405	114.1502		
			22.33351	114.1514		
			22.33297	114.1526		
			22.33243	114.1538		
			22.33189	114.155		
			22.33135	114.1562		
			22.33081	114.1574		
			22.33027	114.1586		
			22.32973	114.1598		
			22.32919	114.161		
			22.32865	114.1622		
			22.32811	114.1634		
			22.32757	114.1646		
			22.32703	114.1658		
			22.32649	114.167		
			22.32595	114.1682		
			22.32541	114.1694		
			22.32488	114.1706		
			22.32434	114.1718		
			22.3238	114.173		
			22.32326	114.1742		
			22.32272	114.1754		
			22.32218	114.1766		
			22.32164	114.1778		

表 5

分別計算每一格網座標與五點爆發地點得曼哈頓距離
(Manhattan Distance)

103	曼哈頓距離								
104	行	列	緯度	經度	淘大花園	寶翠園	孝慈樓	康雅苑	趣園樓
105	1	1	22.345914	114.123762	0.11421	0.07045	0.04593	0.15549	0.07233
106	1	2	22.345914	114.1249624	0.11301	0.06925	0.04473	0.15429	0.07113
107	1	3	22.345914	114.1261628	0.11181	0.06805	0.04353	0.15309	0.06993
108	1	4	22.345914	114.1273633	0.11060	0.06684	0.04233	0.15188	0.06873
109	1	5	22.345914	114.1285637	0.10940	0.06564	0.04113	0.15068	0.06753
110	1	6	22.345914	114.1297641	0.10820	0.06444	0.03992	0.14948	0.06633
111	1	7	22.345914	114.1309645	0.10700	0.06324	0.03872	0.14828	0.06513
112	1	8	22.345914	114.132165	0.10580	0.06204	0.03752	0.14708	0.06393
113	1	9	22.345914	114.1333654	0.10460	0.06091	0.03632	0.14588	0.06273
114	1	10	22.345914	114.1345658	0.10340	0.06211	0.03512	0.14468	0.06153
∴					∴				∴
10104	100	100	22.292508	114.242604	0.05804	0.11674	0.12632	0.01676	0.09992

表 6A

利用拉斯莫方程計算每格網座標對於每爆發地點是源頭的機會，在最右一列是總和 (Sum Up)

103	源頭的機會									Sum up
104	行	列	緯度	經度	淘大花園	寶翠園	孝慈樓	康雅苑	趣園樓	
105	1	1	22.345914	114.123762	1.24231	1.30381	1.33876	1.20457	1.30038	1.27134
106	1	2	22.345914	114.1249624	1.24363	1.30605	1.33582	1.20550	1.30255	1.27309
107	1	3	22.345914	114.1261628	1.24496	1.30834	1.33295	1.20644	1.30477	1.27486
108	1	4	22.345914	114.1273633	1.24630	1.31067	1.33015	1.20739	1.30703	1.27667
109	1	5	22.345914	114.1285637	1.24766	1.31305	1.32742	1.20835	1.30934	1.27850
110	1	6	22.345914	114.1297641	1.24904	1.31547	1.32474	1.20932	1.31169	1.28036
111	1	7	22.345914	114.1309645	1.25043	1.31795	1.32212	1.21029	1.31409	1.28226
112	1	8	22.345914	114.132165	1.25184	1.32048	1.31956	1.21128	1.31653	1.28419
113	1	9	22.345914	114.1333654	1.25327	1.32292	1.31705	1.21227	1.31903	1.28615
114	1	10	22.345914	114.1345658	1.25472	1.32034	1.31459	1.21327	1.32158	1.28815
∴					∴					∴
10104	100	100	22.292508	114.242604	1.32931	1.23959	1.22985	1.28224	1.25903	1.29417

表 6B

然後重新排列成 100×100 格，塗上顏色（如表 7），再對照經緯度，把地圖放上去，再與地圖重疊（如表 8）。調整參數 f, g 使顯示的結果合理，例如海上的機會最低（藍色），市中心和地鐵路線機會最高（紅色），其他地方機會偏低（黃色）新排列成 100×100 格，塗顏色，調整參數 f, g ，使顯示的結果合理，如表 7：

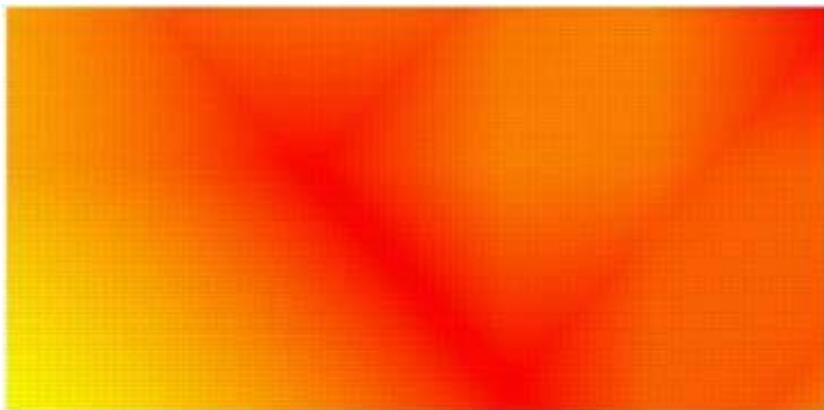


表 7



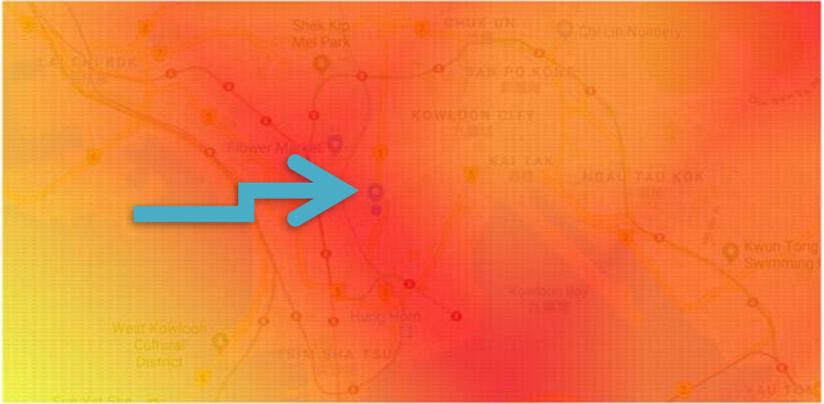


表 8

樂勤: 根據沙士報告，京華國際酒店是首宗沙士爆發的地方，即現時的香港九龍維景酒店，而在計算結果中，代表該酒店的藍色小屋正位於紅色區域中間，即是最有機會是沙士源頭的地方。

熊貓: 這代表你做對了！

嘉俊: 我們能透過拉斯莫方程，重新推測疫症的源頭，這次應用十分成功！如果沙士爆發時，拉斯莫方程能更早被應用，可能可以更快封鎖一些病毒源頭的地方，減少疫症傳播，可以避免更多人受感染而死亡。

總結

我們透過了解拉斯莫方程的假設以及背後原理，運用拉斯莫方程於尋找殺人犯，之後根據其方程的假設，再配合病毒傳播時的特質，成功應用拉斯莫方程於尋找沙士源頭上。證明拉斯莫方程可應用於探案尋兇以及疫症溯源的問題上。

參考資料：

1. 維基百科:

https://en.wikipedia.org/wiki/Rossmo%27s_formula

https://en.wikipedia.org/wiki/Peter_Sutcliffe

<https://zh.wikipedia.org/wiki/%E5%9A%B4%E9%87%8D%E6%80%A5>

<https://zh.wikipedia.org/wiki/%E6%80%A7%E5%91%BC%E5%90%B8%E7%B3%BB%E7%B5%B1>

<https://zh.wikipedia.org/wiki/%E7%B6%9C%E5%90%88%E7%97%87>

https://en.wikipedia.org/wiki/Taxicab_geometry

2. 沙士專家委員會報告(Report of the SARS Expert Committee)

http://www.sars-expertcom.gov.hk/tc_chi/reports/reports.html

3. 香港網絡大典 2003 年沙士疫潮

[http://evchk.wikia.com/wiki/2003%E5%B9%B4%E6%B2%99%E5%A3%](http://evchk.wikia.com/wiki/2003%E5%B9%B4%E6%B2%99%E5%A3%AB%E7%96%AB%E6%BD%AE)

[AB%E7%96%AB%E6%BD%AE](http://evchk.wikia.com/wiki/2003%E5%B9%B4%E6%B2%99%E5%A3%AB%E7%96%AB%E6%BD%AE)

優異作品：

如何挑選強積金

學校名稱：順德聯誼總會李兆基中學

學生姓名：李志江、劉嘉熙

指導教師：許俊江老師



如何挑選強積金

強制性公積金計劃（強積金）推行多年，僱員和僱主必須供款，僱員可於計劃內分配不同投資百分比到不同基金，本文嘗試運用中學數學課程的知識，幫助僱員分析不同基金的過往表現，作未來表現的參考。

引言

強積金是近年較具爭議的話題之一，根據制度，僱員和僱主須按最低及最高收入水平，每一個月分別向強積金受託人，注入有關僱員收入的 5%或以上作為供款，自僱人士亦須以個人收入最小 5%的作為強制性供款。但同時要繳交的基金管理費較其他國家高，於是有些批評認為強積金回報率偏低，加上行政費昂貴，根本達不到保障效果，甚至有市民建議政府應全面檢討甚至取消強積金。可是，大家都知道政府是不會取消強積金的，那麼，如何從眾多強積金中選擇一個適合我們的基金便成為了大家的煩惱。

本文的創作主要參考楊良河博士所著的「數學的應用投資組合及市場效率」書冊，其內容深入淺出地利用不同數學和統計概念作投資分析。我們嘗試班門弄斧，向各位「打工仔」介紹如何運用一些中學數學課程的知識，助你作出明智選擇。

我們選了「信安強積金計劃 800 系列」作為例子¹，共有 15 個基金選擇（沒有包括核心累積基金及 65 歲後基金，因它們在 2017 年推出，沒有足夠歷史數據供我們分析）。以下列出基金名稱及其過往五年的歷年回報，見表一：

¹ 信安強積金計劃 800 系列

https://iframe.principal.com.hk/allweb/pdf/S800_MPR.pdf

表一：2013至2017年基金的歷年回報

歷年回報(%)

	2013	2014	2015	2016	2017
中國股票基金	5.37	7.78	-8.61	0.51	39.48
香港股票基金	9.37	2.56	-5.14	0.51	41.52
恆指基金	4.39	3.91	-5.27	2.35	38.49
亞洲股票基金	1.49	-0.25	-9.35	4.28	41.32
美國股票基金	30.32	10.75	-1.39	4.56	18.91
國際股票基金	24.65	0.46	-3.61	0.96	22.13
環球增長基金	11.16	-0.45	-3.08	1.18	22.83
長線增值基金	6.93	-0.66	-3.12	1.23	17.7
平穩回報基金	2.86	-1.03	-3.17	1.01	12.64
國際債券基金	-3.53	-1.65	-4.35	0.93	6.48
香港債券基金	-2.61	3.07	2.08	-1.09	2.29
長線保證基金	-0.45	1.75	-0.98	-0.53	3.66
資本保證基金	0	0.01	0.01	0.01	0.03
港元儲蓄基金	0.41	-0.16	-0.4	-0.17	0.46
張積金保守基金	0.01	0.18	0.02	0.01	0.01

很多「打工仔」每當有新工作的時候，要重新選擇投資不同的基金，既環球又亞洲、既股票又債券、既增值又保本、……，種類繁多，眼花撩亂，不知道該如何挑選。

回報

當然，回報是「打工仔」重要的考慮因素之一。這時候可應用**算術平均數 (arithmetic mean)** 去計算平均每年回報率，作為回報指標。

相信大家對算術平均數都耳熟能詳。設有 x_1 、 x_2 、...、 x_n 共 n 個數，則

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n},$$

其中 \bar{x} 為算術平均數，是量度數據的**集中趨勢 (Central tendency)**的其中一種方法。我們計算過往五年的歷年回報的算術平均數作為平均每年回報率，即回報指標，再按高至低排列，見表二：

表二：基金過往五年的平均每年回報率

	歷年回報(%)					平均每年 回報率(%)
	2013	2014	2015	2016	2017	(回報指標)
美國股票基金	30.32	10.75	-1.39	4.56	18.91	12.63
香港股票基金	9.37	2.56	-5.14	0.51	41.52	9.76
國際股票基金	24.65	0.46	-3.61	0.96	22.13	8.92
中國股票基金	5.37	7.78	-8.61	0.51	39.48	8.91

	歷年回報(%)					平均每年 回報率(%)
	2013	2014	2015	2016	2017	(回報指標)
恆指基金	4.39	3.91	-5.27	2.35	38.49	8.77
亞洲股票基金	1.49	-0.25	-9.35	4.28	41.32	7.50
環球增長基金	11.16	-0.45	-3.08	1.18	22.83	6.33
長線增值基金	6.93	-0.66	-3.12	1.23	17.7	4.42
平穩回報基金	2.86	-1.03	-3.17	1.01	12.64	2.46
香港債券基金	-2.61	3.07	2.08	-1.09	2.29	0.75
長線保證基金	-0.45	1.75	-0.98	-0.53	3.66	0.69
張積金保守基金	0.01	0.18	0.02	0.01	0.01	0.05
港元儲蓄基金	0.41	-0.16	-0.4	-0.17	0.46	0.03
資本保證基金	0	0.01	0.01	0.01	0.03	0.01
國際債券基金	-3.53	-1.65	-4.35	0.93	6.48	-0.42

若只考慮過往五年的平均每年回報率，美國股票基金最好，每年約有 12.6% 的回報。

風險

若只關注基金的回報，挑選回報最高的基金，其缺點是價格變動大，在某些時期回報率甚低，萬一「打工仔」不幸需要於該段時間提取強積金，恐怕得不償失。因此，基金的風險亦是重要考慮之一。我們可利用標準差 (standard deviation)，去計算出數據的離散值 (dispersion)，即數據與數據之間相差的多少，

下列為標準差的公式：

$$\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}}$$

，其中 σ 為標準差。我們

計算過往五年的歷年回報的標準差作為風險指標，再按高至低排列，見表三：

表三：基金的歷年回報的標準差

	歷年回報(%)					平均每年 回報率(%)	標準差 (%)
	2013	2014	2015	2016	2017	(回報指標)	(風險指標)
亞洲股票 基金	1.49	-0.25	-9.35	4.28	41.32	7.50	17.52
香港股票 基金	9.37	2.56	-5.14	0.51	41.52	9.76	16.54
中國股票 基金	5.37	7.78	-8.61	0.51	39.48	8.91	16.28
恆指 基金	4.39	3.91	-5.27	2.35	38.49	8.77	15.26
國際股票 基金	24.65	0.46	-3.61	0.96	22.13	8.92	11.95
美國股票 基金	30.32	10.75	-1.39	4.56	18.91	12.63	11.11
環球增長 基金	11.16	-0.45	-3.08	1.18	22.83	6.33	9.56
長線增值 基金	6.93	-0.66	-3.12	1.23	17.7	4.42	7.42
平穩回報 基金	2.86	-1.03	-3.17	1.01	12.64	2.46	5.47
國際債券 基金	-3.53	-1.65	-4.35	0.93	6.48	-0.42	3.90
香港債券 基金	-2.61	3.07	2.08	-1.09	2.29	0.75	2.20
長線保證 基金	-0.45	1.75	-0.98	-0.53	3.66	0.69	1.76

	歷年回報(%)					平均每年 回報率(%)	標準差 (%)
	2013	2014	2015	2016	2017	(回報指標)	(風險指標)
港元儲蓄 基金	0.41	-0.16	-0.4	-0.17	0.46	0.03	0.34
張積金 保守基金	0.01	0.18	0.02	0.01	0.01	0.05	0.07
資本保證 基金	0	0.01	0.01	0.01	0.03	0.01	0.01

風險與回報

以數據的集中趨勢作為基金的回報指標，以數據的離散值作為基金的風險指標，正體現了我們分析數據時兩種概念雖不同、但同樣重要的方向。

選擇基金要權行風險與回報，一般來說，風險與回報有**正向關係 (positive relationship)**，即回報越高，風險越大，反之亦然。但如何衡量兩者？在此，我們嘗試應用**聯變 (Joint Variation)**的概念。

若 z 隨 x 正變且隨 y 反變，則 $z = k \left(\frac{x}{y} \right)$ ，意思即 z 與 x 成正比

(保持 y 不變)，且 z 與 y 成反比 (保持 x 不變)。

我們將回報指標 (x) 除以風險指標 (y)，計算出每單位風險的回報率 (z)，並按高至低排列。見表四：

表四：每單位風險的回報率

	平均每年 回報(%)	標準差(%)	平均每年回報率／標準差
	(回報指標)	(風險指標)	(每單位風險的回報率)
資本保證基金	0.01	0.01	1.22
美國股票基金	12.63	11.11	1.14
國際股票基金	8.92	11.95	0.75
強積金保守基金	0.05	0.07	0.69
環球增長基金	6.33	9.56	0.66
長線增值基金	4.42	7.42	0.59
香港股票基金	9.76	16.54	0.59
恆指基金	8.77	15.26	0.57
中國股票基金	8.91	16.28	0.55
平穩回報基金	2.46	5.47	0.45
亞洲股票基金	7.50	17.52	0.43
長線保證基金	0.69	1.76	0.39
香港債券基金	0.75	2.20	0.34
港元儲蓄基金	0.03	0.34	0.08
國際債券基金	-0.42	3.90	-0.11

若以每單位風險的回報率來看，資本保證基金最好，但其回報率甚低，可能受基數太小影響（風險指標只有 0.01%），另它排

名第一。排名第二的是美國股票基金，它亦是回報指標最高的。它可謂進取人士的最佳選擇。

組合

常言道，「不要把所有的雞蛋放在一個籃子裡」，「打工仔」其實可以持有多種不同的基金以分散風險，此視為一個組合。實際上組合如何能分散風險呢？見表五：

表五：組合的表現

	歷年回報					平均每年 回報率(%)	標準差 (%)	平均每年回 報／標準差	資 金 分 佈 比 例
	2013	2014	2015	2016	2017	(回報指標)	(風 險 指 標)	(每單位風險 的回報率)	
美國股票 基金	30.3	10.8	-1.4	4.56	18.9	12.63	11.11	1.14	0.28
國際股票 基金	24.7	0.46	-3.6	0.96	22.1	8.92	11.95	0.75	0.30
環球增長 基金	11.2	-0.5	-3.1	1.18	22.8	6.33	9.56	0.66	0.42
組合	20.6	3.22	-2.7	2.23	21.3	8.93	10.01	0.89	1.00

若我們按 0.28 : 0.30 : 0.42 的比例分別投資於美國股票基金、國際股票基金及環球增長基金，按基金的歷年回報計算出組合的歷年回報，會發現這個組合與單單選擇國際股票基金所得的回報率相若（8.93%和 8.92%），但風險指標較低（低近 2 個百分點）。

雖然「信安強積金計劃 800 系列」只有 15 個（實為 17 個）基金選擇，但透過不同的資金比例分佈，可得出成千上萬個不同的基金組合。「打工仔」可以先衡量自己承擔風險的能力，再訂一個合適的回報目標，再以不同的資金比例分佈去製造一個回報相若，但風險較低的組合。

結語

當然，以上例子只是根據過往五年的數據作計算，若有更長時序的（如十年）或時間間距更短的（如每個月）回報率就更加理想。強積金乃是一項長遠投資，時間以數十年計，「路遙知馬力」，在五年時間裡，某些基金波動不大，但隨著經濟週期的各個時期，我們可更加清楚知道基金的波動性。

另外，「股票的價格是可升可跌的」、「過往的表現並非未來表現的指標」，此兩句說話雖然曾經被人嘲諷，但卻是鐵一般的事實，本文與大家分享的方法只能作為一個參考，「打工仔」亦需要衡量其他因素，例如距離退休的年期、個人經濟狀況、基金的收費等等。

但是，數學與統計的確提供了簡單的方法，讓我們去觀察和比較大量數據，不容置疑。

（字數：1778）

參考資料：

楊良河博士(2009)。〈數學百子櫃系列(六)數學的應用：投資組合及市場效率〉。教育局課程發展處數學教育組。

優異作品:

To Trust, or Not to Trust – Gini Coefficient

School Name: Shau Kei Wan Government Secondary School

Name of Student: CHAN Ngo-kiu, Baycle

Supervising Teacher: Ms. WONG Pui-shan

Abstract

The Gini coefficient is commonly used to measure the income gap in a country. It measures the dispersion of all income levels in the population. This index ranges from 0 to 1 represented perfect equality to maximal inequality respectively. With a score of 0.539 in 2017, Hong Kong is one of the countries in the world with the highest income gap. With a country like Ethiopia has the Gini coefficient at 0.33 is a narrower income gap than Hong Kong. However, a conclusion cannot be drawn that Ethiopia is having a fairer economic condition than Hong Kong as the notion is counter-intuitive. The absolute income level using GDP is introduced to factor in the relative living standard between the two countries. The economic development of Ethiopia is far behind than that of Hong Kong resulted in substantially lower absolute income levels, hence the living standard, of Ethiopia. The Gini coefficient is a good measure of the income gap in a country, but for a comparison

between countries, other factors would need to be taken into account. Interpretation of statistical results is required to put into the appropriate context to derive a correct inference.

Calvin, a research student in the Engineering Department in the Hong Kong University was reading a newspaper on his tablet in a campus cafe. His classmate, Dawit, from Ethiopia, saw him.

Dawit hailed, ‘Hello, Calvin.’

Calvin said, ‘Hi, Dawit. How are you doing?’

‘Not bad. You?’



‘Alright. You see. I am reading this article about income gap in Hong Kong. I think it is outrageous. The Census and Statistics Department released the household income distribution report showing “Hong Kong as the world’s second-most unequal city in terms of income.” The Gini coefficient increased from 0.537 in 2011 to a record high of 0.539 last year...’

Dawit interrupted, ‘Well. But...’

Calvin continued, ‘What is more? The income of the richest ten percent of households is 44 times of that of the poorest 10 percent who earned only HK\$2,560.’

Dawit questioned, ‘How is the problem? Compared with my home country, Hong Kong is doing a lot better. What is your complaint?’

‘Can’t you see the Hong Kong’s Gini coefficient stays on a high level and keep rising? Let me check the index of your country.’

After a few swipes and taps on the screen of his tablet, Calvin exclaimed, ‘Wow! Come and see it yourself.’

Dawit was looking over the tablet as Calvin pointed to a figure shown on the webpage and said, ‘The number for Ethiopia is 0.33. It is low indeed. It ranks 113th place in the low end out of 156 countries while Hong Kong sits high up at the 9th place in the table.’

Dawit rebutted, ‘But this is not true ...’

Calvin interrupted, ‘No “but”. The numbers would not tell lies. It is clear that Ethiopia is doing better than Hong Kong.’

‘It is really absurd.’

At this time, Sophie, a common friend of them, approached.

‘Hi, guys. What’s up? Look like you two are in an argument. What is it all about?’

Dawit said, ‘We are discussing Gini index. Calvin insists Ethiopians are doing better Hong Kong people. But I don’t think this is right.’

Sophie said, ‘Well. I come at the right time as I have just learned Gini index in my econometrics course. Maybe I can try to explain how to apply it to explain income distribution and what else we should look at to make a comparison between nations.’

‘Before we look at the Gini index, I want to introduce you to the concept of dispersion.’

‘You all know rolling of dice, right? I can use it as an example. Let me pull the tables and charts from my tablets to show you the

details.’

‘This table displays the possibilities of occurrence for the sum of two dice. 2 and 12 have only one possibility because there is one possible combination only: 1+1 is 2 and 6 + 6 is 12. 7 has 6 combinations: 1 + 5, 5 + 1, 2 + 5, 5 +2, 3 +4 and 4 + 3. So 7 has a highest chance or probability of occurrence, in other words, it has the highest occurrence frequency.

Sum of two dice	2	3	4	5	6	7	8	9	10	11	12	Total
Frequency	1	2	3	4	5	6	5	4	3	2	1	36
Probability	0.027778	0.055556	0.083333	0.111111	0.138889	0.166667	0.138889	0.111111	0.083333	0.055556	0.027778	1

‘Plotting the numbers on a graph, we have a bar chart showing the distribution of the sum of two dice against the frequency of the occurrence. This bar chart shows what we call a discrete probability distribution made up of discrete random variables in statistics. The random variable provides a measure of outcomes of random events.’



‘The mean of a discrete random variable is the weighted mean of the values. The formula is $\mu_x = x_1p_1 + x_2p_2 + \dots + x_{12}p_{12} = \sum_{i=1}^{12} x_i p_i$. In other words, multiply each given value by the probability of getting that value, then add everything up.’

‘The variance of the discrete random variable is $\sigma^2 = \sum_{i=1}^{12} (x_i - \mu)^2 p_i$.’

‘It is a little complicated for working out the variance of continuous random variables as calculus is required, but the principle is the same. The formula is $\sigma^2 = \int_a^b (x - \mu)^2 f(x) dx$

‘We can also use a curve to represent the dispersion giving you a more intuitive feel of the dispersion.’



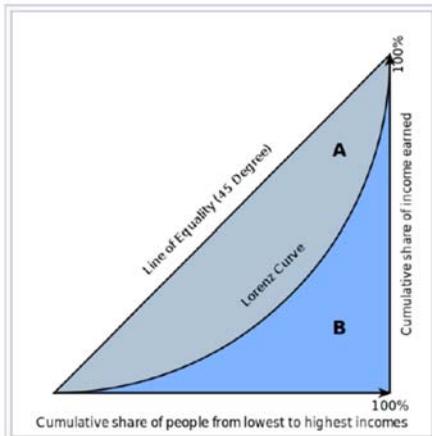
‘I think you guys have a good understanding of dispersion in statistics now. Now then, I will move to talk about Gini coefficient.’

‘Gini coefficient is a measure of statistical dispersion to compare the income or wealth of a country’s population and is a common adopted index for evaluation of inequality. It is also handy to use it to make comparison across nations.’

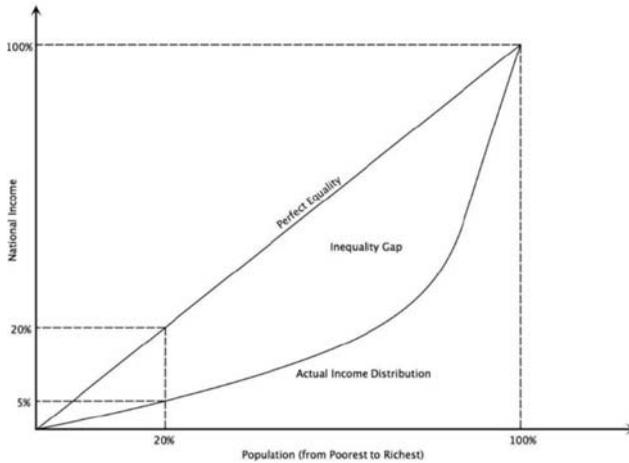
‘Income levels in a country ranges from lowest to highest based on their economic development. Different income levels are considered as values of a frequency distribution, like what I explained earlier. The Gini coefficient ranges from 0 to 1. A value of 0 represents perfect equality; in other words, every citizen of the nation is earning the same income. A value of 1 or 100% points to maximal inequality; meaning one person or a small group of citizens has all the income and everybody else has nothing. Of course, in reality, these two scenarios scarcely happen.’

‘The Italian statistician, Corrado Gini has developed the measure of Gini coefficient in 1912 and used it as the study of the inequality of a distribution. The Gini coefficient is based mathematically on Lorenz curve which is a graphical representation of the distribution of income or wealth developed by Max Lorenz in 1906. On this

chart, the Lorenz curve presents the proportion of the income earned by any given percentage of the population. The line of equality sloping at 45 degrees indicates a perfectly equal income distribution.'



'This chart gives you an example of how to read it. The 20% of the population earns 5% of the income compared with 20% income on the Line of Equality. The area A enclosed by the Lorenz Curve and the Line of Equality is the income gap. Gini coefficient is calculated by Area A divided by sum of Areas A and B. The more the Lorenz Curve shifts to the right, the larger is the Gini coefficient since the numerator grows faster than the denominator.'

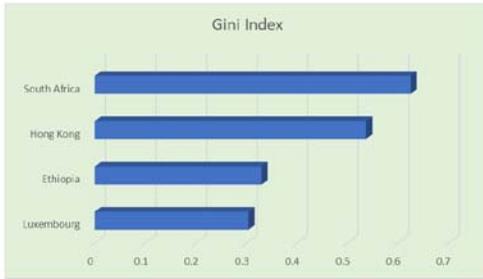


‘So far, do you have any questions, Calvin and Dawit?’

Dawit said, ‘Very well. You explained the Gini coefficient clearly. But it didn’t explain the case of Hong Kong and Ethiopia.’

Sophie replied, ‘Bear with me. I will come to this next.’

‘I am going to show you the Gini coefficients of 4 countries. They are South Africa, Hong Kong, Ethiopia and Luxembourg in the descending order.’

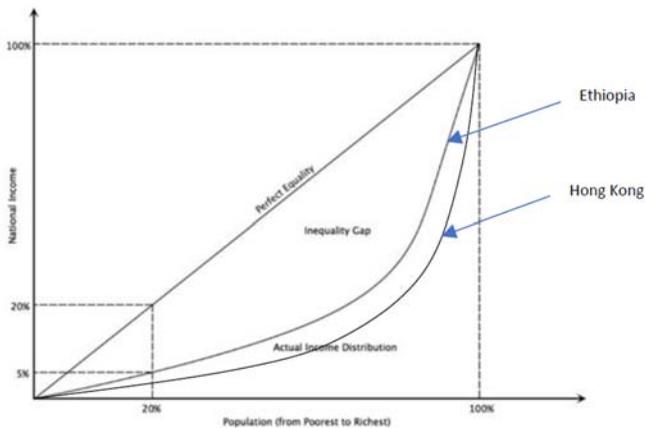


Country	Gini Index
South Africa	0.625
Hong Kong	0.537
Ethiopia	0.33
Luxembourg	0.304

‘The chart speaks louder than words. In term of income distribution within the country, Luxembourg and Ethiopia are doing the best among the 4 countries. Calvin is perfectly correct in this regard. But what not to forget about the use of the Gini coefficient is to single out the factor of income distribution but ignoring other factors such as economic development and absolute income levels.’

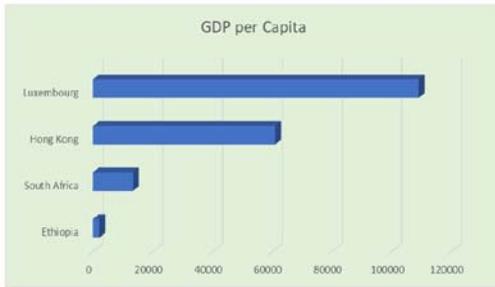
‘As the Gini coefficient is a normalised index to allow values measured on different scale to a notionally common scale. So we can use the normalised values of Gini coefficient to compare those values in datasets of different countries. It allows the index to be compared across countries regardless the absolute income levels.’

‘That is why Calvin said Ethiopia is doing better than Hong Kong. In more precise terms, what he meant is the income gap in Ethiopia is less severe than Hong Kong. The Lorenz Curve of Hong Kong is on the right side of Ethiopia.’



‘We look at the absolute income levels in the two countries. Here I introduce the concept of GDP or Gross Domestic Product which is the surrogate measure of income. I believe you have heard on TV or read in the newspaper of GDP which measures how much a country produces or earns in a given year.’

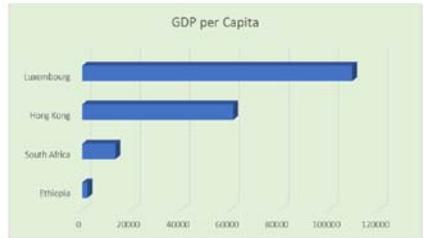
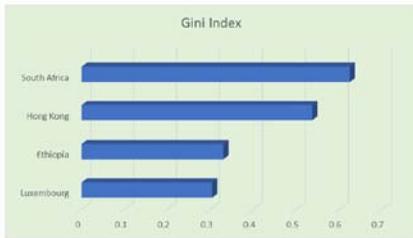
‘Using GDP per capita is more appropriate due to the difference in population. The GDP per capita is GDP of the country divided by its population. The chart and the table show the GDP per capita for the 4 countries mentioned in the ranking of the GINI coefficient in descending order.’



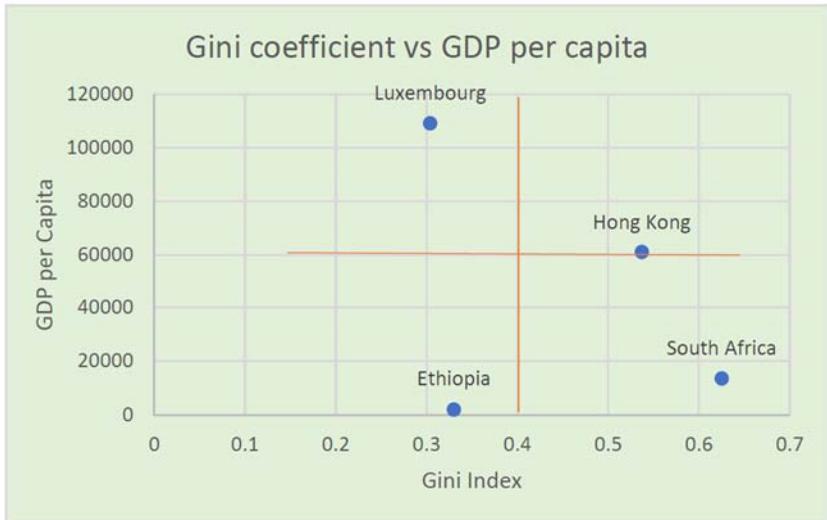
Country	GDP per Capita (USD)
Luxembourg	109,100
Hong Kong	61,000
South Africa	13,400
Ethiopia	2,100

‘You can see the startling differences in the numbers. Luxembourg has the highest GDP, Hong Kong comes second, South Africa the next and Ethiopia lowest. Luxembourg earns 51 times more than Ethiopia. Ethiopia earns less than 4 percent of the income of Hong Kong.’

‘If we put the two charts of Gini coefficient and GDP side by side, the trends of Gini coefficient and GDP go into different directions. Interestingly, Luxembourg has the lowest GINI coefficient, but the highest absolute income while South Africa has the highest GINI coefficient with a second lowest income. Hong Kong and Ethiopia have mixed results. Hong Kong has a relatively high in Gini coefficient and income and Ethiopia has a fairly low Gini coefficient, but the lowest income.’



‘How to make sense of these two sets of data? Gini coefficients of the 4 countries are plotted against GDP per capita to come up with this chart. I have drawn one horizontal line and one vertical line to divide the graph into 4 quadrants. Upper left quadrant where Luxembourg situated represents countries with low Gini coefficient and high GDP. Upper right quadrant where Hong Kong is located indicates countries with high Gini coefficient and high GDP. The other two are lower right and lower left quadrants. The former containing Ethiopia which has low Gini coefficient and low GDP and the latter has high Gini coefficient and low GDP.’



‘This chart displays there is no fixed relationship between Gini coefficient and GDP. Though Hong Kong has much wider income gap, the absolute income of Hong Kong is high. To put it into economic context, Hong Kong is a developed country with a high GDP level. But its economic development does not benefit evenly among the entire population. People of higher income has taken a larger share of economic growth reflecting in the increment in their income while those of lower income have gained a small share on average.’

‘On the surface, Ethiopia enjoys a low-income gap, but looking at the income level is an appallingly low level of USD 2,100. It can be interpreted as Ethiopians are equally poor because the overall economic growth is low and does not benefit many people.’

‘True that Calvin pointed out the low Gini coefficient in Ethiopia is a good thing. However, it might not be appropriate to compare with Hong Kong as the income levels of the two countries have a huge discrepancy. The reason why is the difference in the stage of economic development. Ethiopia is still a developing country at the low end of economic development phase. Hongkongers, on average, earn 29 times as Ethiopians. So even the low-income groups in Hong Kong may earn more than the high-income group in Ethiopians. From this economic perspective, Hong Kong enjoys a much higher living standard than Ethiopians. Dawit, you are equally right in proclaiming Hong Kong is in a significantly stronger economic status, despite the increasing income gap.’

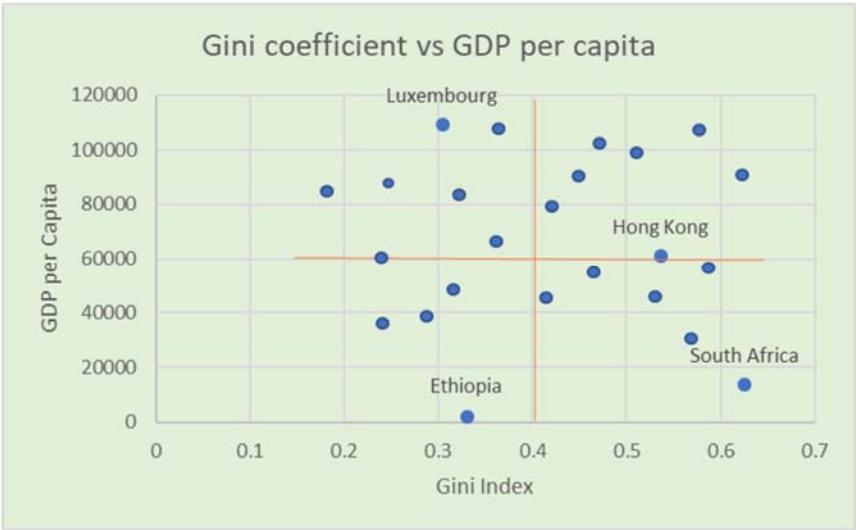
‘Calvin, Dawit, you are both right depending on how we phrase the question and frame the model of investigation. To sum it up, simply to investigate the income gaps between countries, the Gini coefficient is sufficient. When also considering the economic strength and living standard, other indicators like GDP should be included.’

‘Calvin, Dawit, are you happy with my explanation?’

Dawit exclaimed, ‘Thank you very much, Sophie. Your lecture is superb and has enlightened me. I have learned a good lesson not only about Gini coefficient but how to interpret the result in the economic context.’

Calvin said, ‘Sophie, thank you. I now have a full picture and have learned not to rely on a single factor when looking at an issue.’

Sophie continued, ‘If we plot the chart with all the countries in the world, we get something like this. We may then be able to do some analysis and research and draw economic conclusions,’



‘To extend the analysis to include more economic and social factors, multivariate analysis can be used to study more complex situations

in the real world. Not limited to economics, statistics can be used in demography, geology, ecology and any other fields with big data sets to draw inference for understanding trends and application to the real world.’

‘Alright. Really enjoyed our sharing notes on some interesting subjects. Got to go. See you later.’

(No. of words: 2270)

References

1. The Gini Coefficient figures of Hong Kong is retrieved from <http://www.scmp.com/news/hong-kong/economy/article/2097715/what-hope-poorest-hong-kong-wealth-gap-hits-record-high>
2. The Gini Coefficient and GDP figures of the 4 countries retrieved from <https://www.cia.gov/library/publications/the-world-factbook/rankorder/2172rank.html>
3. The Source of Lorenz curve https://en.wikipedia.org/wiki/Lorenz_curve

4. The article from South China Morning Post
<http://www.scmp.com/news/hong-kong/economy/article/2097715/what-hope-poorest-hong-kong-wealth-gap-hits-record-high>
5. Gini, C. (1912), *Variability and Mutability*, C. Cuppini, Bologna, 156 pages. Reprinted in *Memorie dimetodologica statistica* (Ed. Pizetti E, Salvemini, T). Rome: Libreria Eredi Virgilio Veschi (1955).
6. Gini, C. (1936) *On the Measure of Concentration with Special Reference to Income and Statistics*, Colorado College Publication, General Series No. 208, 73-79.
7. Lorenz, M. O. (1905). "Methods of measuring the concentration of wealth." *Publications of the American Statistical Association*. *Publications of the American Statistical Association*, Vol. 9, No. 70. 9 (70): 209–219.

優異作品:

Assassination – ES.Drump

School Name: St. Francis Xavier's College

Name of Student: KWOK Ho-yin, CHOI Siu-hin,
CHAN Pui -wing

Supervising Teacher: Mr. Chan Pak Kei

Assassination - ES.Drump

Part 1

One day, Perry was watching the TV.



Perry whispered, ‘This ES.Drump is such a bigot. No matter what he has done, he is damaging our society. I have to kick him down from the seat of the President. I have to be the Savior of the world!’

Suddenly, a plan to kill the President appeared at Perry’s mind. He then found Tiffany, his girlfriend, for help.

←  Tiffany



TODAY

Hey, Tiffany   23:45 ✓✓

Perry, what's up?   23:46

Do you want to do something incredible but dangerous? 23:46 ✓✓

Of course, if i work with you 23:46

I am planning to assassinate ES.Drump. Are you interested in it? 23:47 ✓✓

Of course   but why 23:47

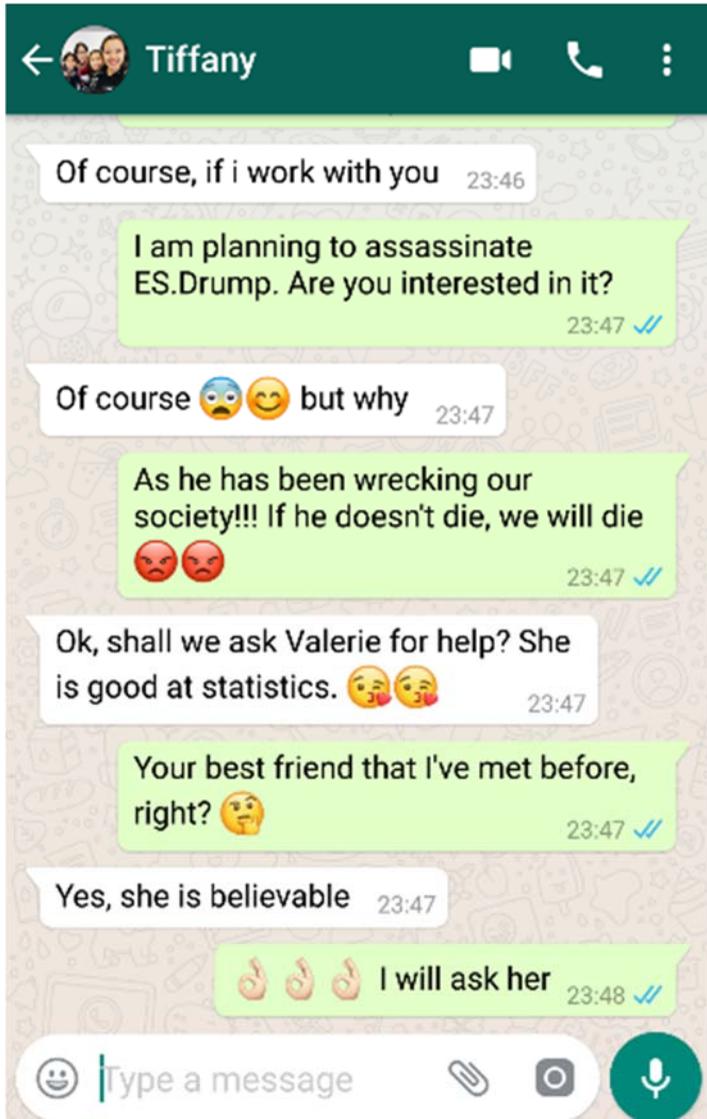
As he has been wrecking our society!!! If he doesn't die, we will die   23:47 ✓✓

Ok, shall we ask Valerie for help? She



Type a message





After that, Perry thought that Valerie is Tiffany's best friend. She should be with us. It would be better for him to ask her.



Perry thought that she was quite straightforward.



After finding the two helpers,
he was focusing on recruiting mercenaries with them.

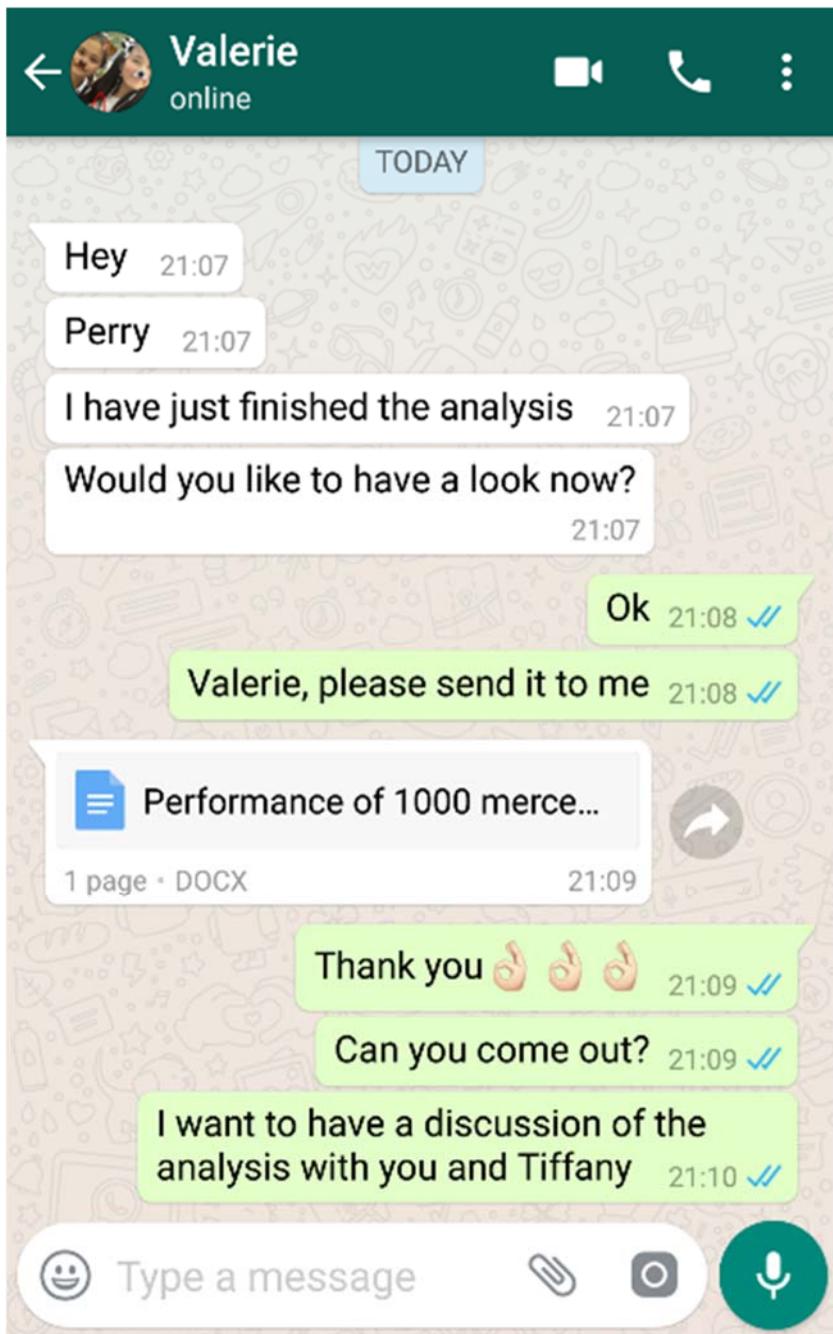


Part 2

After 2 months, Perry, Tiffany
and Valerie were working in a
house.

Suddenly, a message was sent
to Perry's phone.





Walking towards Perry, Valerie said, ‘Boss, what do you want to discuss?’

‘Wait,’ Perry replied, and turned his face to Tiffany. ‘My dear, come here. We have to discuss the performance of our soldiers who are going to assassinate ES.Drump,’ Perry asked.

‘Ok. I’m coming,’ replied Tiffany.

On the other side, when Valerie was walking towards Perry, she started to record the conversation between them using her phone.



Part 3

Valerie said, ‘Boss, in the test, assassigators were tested separately in the same room with the same test in different times. In 1000 appliers, only 60 of them could pass all the tests which are some basic requirements for our killers.’



Tiffany added, ‘The tests include unlocking handcuff, escaping the secret chamber and climbing up the wall.’

Perry said, ‘That seems great! Which country do they mostly come from?’

Tiffany replied, ‘There are 30 of them from Russia and 30 of them from China.’



‘Well, I think it’s good for our communication. In order to enhance the co-operation and communication, those recruited assassigators would be better if they come from the same country. So, we should analyse them in two groups, Russian and Chinese,’ said Perry.

Valerie said, ‘Ok, I’ve tidied up the data. Here is the data of unlocking handcuff.’

Perry whispered, ‘Yes. The assassigators need this skill to escape when they are caught.’

‘You’re right. Escaping from enemy at once is very important.’



Time for Unlocking Handcuff (s)	
Russian	Chinese
40 40 41 43 45	10 18 19 20 27
49 52 56 56 57	31 36 38 40 45
57 59 61 61 61	50 52 54 54 60
62 63 63 65 65	60 60 63 67 70
66 69 70 70 77	72 81 89 90 99
77 80 83 87 122	108 120 120 121 122

Valerie said, ‘I’ve analysed the data for this test. In my opinion, the Chinese performed much better.’

$$\text{Mean} = \frac{1}{30} \sum_{r=1}^{30} n_r \text{ where } n_r \text{ be } r^{\text{th}} \text{ datum}$$

$$\text{Median} = \frac{15\text{th datum} + 16\text{th datum}}{2}$$

Mode = the most frequent datum

Time for Unlocking Handcuff (s)		
	Russian	Chinese
Mean	63.23	63.2
Median	61	60
Mode	61	60

Tiffany replied, ‘Oh, yes. The mean time of Chinese is shorter than Russian. Chinese must be better than Russian. They can be our dawn of career.’

Perry doubted, ‘Wait... How come you use mean, mode, median to analyse these data? The median and mode is the same while mean is very similar. Comparing these data may not give us a valid conclusion. Moreover, team spirit is also an important factor. They should perform in similar result for having a sync mind.’



We would have a better analysis of the dispersion of the data. **Dispersion** is the extent to which a distribution is stretched or squeezed. Range will be a simpler way to find out dispersion in this case.’

Tiffany asked, ‘What is the meaning of range?’

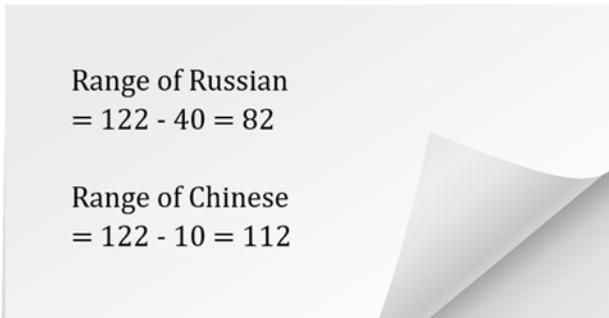
‘Range is the difference between the maxima and minima. The greater the range, it means the existing of extreme value or the

greater dispersion of data,’ replied Perry.

She continued, ‘Yes, but the maxima of both Chinese and Russian are 122s. For the minima, the Chinese’s is much smaller than the Russian’s.’

Range = The maximum data – the minimum data

Perry wrote some calculation on a piece of paper.



Range of Russian
= $122 - 40 = 82$

Range of Chinese
= $122 - 10 = 112$

Perry replied, ‘It is true. The range of Chinese is larger than that of Russian. But as you can see, the performance of Chinese contains 5 extremely large data (i.e. 108, 120, 120, 121, 122) and 4 extremely small data (i.e. 10, 18, 19, 20) while that of Russian just contains 1 extremely large datum (i.e. 122). Extreme value may affect the result of the measuring the dispersion.’

‘Also, the time for Russian to unlock handcuff is similar to each other but not as various as Chinese. So Russian is better in this task, right?’ disturbed Tiffany.

‘You are so smart, my beloved. Assassination requires team spirits but not individualism. If only 1 or 2 of them perform better, our mission is bounded to fail. Therefore, it is better for all of them to have a similar performance,’ continued Perry.

Part 4

‘OK. Here are the data of escaping secret chamber. This task is to check their analytic skills and calmness when they have to find the way to get inside the President’s House,’ said Valerie.



Time for escaping secret chamber (min)										
Russian					Chinese					
20.00	29.50	29.95	32.15	32.40	20.50	34.00	35.50	35.75	36.00	
33.00	33.00	34.00	34.00	34.25	36.50	37.00	37.50	37.90	38.00	
35.00	35.50	36.00	36.50	38.00	38.25	38.80	39.00	39.00	39.00	
38.00	38.00	39.50	39.50	40.00	39.90	40.00	40.75	41.00	42.30	
40.00	40.00	40.00	42.50	43.00	44.00	47.00	49.00	49.20	50.50	
45.00	45.50	47.00	48.50	80.00	51.00	52.50	53.00	54.50	61.00	

‘Do they have to find hints on their own?’ asked Tiffany.

‘Of course. Oh yes! Here are some calculations for median, mode and range.’ Valerie added.

Time for escaping secret chamber (min)		
	Russian	Chinese
Median	38.00	39.00
Mode	40.00	39.00
Range	60.00	40.50

‘They got the similar mode and median. But the range of Chinese is much lower than that of Russian. So, Chinese performed better in this case, right?’ analysed Valerie in deadly earnest.

‘Emm ... well, actually Russian should perform better. We can’t just refer to the range as it will be affected by some extreme data easily. This time, we may use inter-quartile range,’ replied Perry.

‘What is inter-quartile range?’ asked Tiffany.

‘It is the difference between the datum of 25% and the datum of 75%. In this case, it will not be affected by the extreme value easily.’ explained Perry.

‘Why my method is all banned by him ... How can I’ whispered Valerie with nobody heard.

25% quartile is represented by 25th percentile
 50% quartile is represented by 50th percentile
 75% quartile is represented by 75th percentile
 Interquartile range = 75% quartile – 25% quartile

Time for escaping secret chamber (min)		
	Russian	Chinese
25% quartile	33.50	37.25
50% quartile	38.00	39.00
75% quartile	40.00	48.00
Inter-quartile range	6.5	11.25

‘Oh really! Everything changes!’ shouted Tiffany.

‘Yes. As we can see, no matter which quartile, Russian still performed better. Besides, for the inter-quartile range of Russian is smaller than that of Chinese. It also means that the performance of each mercenary is less various to each other compared to the Chinese,’ said Perry.

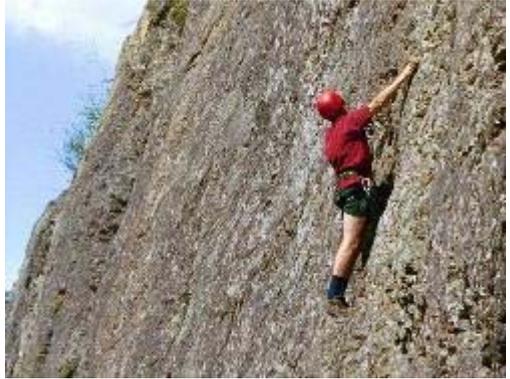
‘Well ... How can I change his mind’ whispered Valerie.



Part 5

Tiffany said, ‘This is the result of climbing up a wall with tying a rope on their own.’

Valerie added, ‘This test is to check their nimbleness and dexterousness because they have to climb up the enclosure of President’s House and get rid of the infra-red in some rooms.’



Time for Climbing up a 8-meter wall with tying a rope on their own (min)									
Russian					Chinese				
0.50	1.00	1.10	1.25	1.30	2.00	2.00	2.20	2.50	2.75
2.00	3.00	3.25	3.50	4.00	2.75	2.80	2.85	4.00	4.25
4.50	4.75	4.80	4.95	5.00	4.30	4.50	4.60	4.75	4.75
5.10	5.10	5.20	5.50	5.70	7.45	7.50	7.50	7.65	7.80
6.00	8.10	8.10	8.10	8.20	7.90	7.90	7.90	9.00	9.10
8.20	8.25	8.25	8.30	10.00	9.25	9.50	9.60	9.75	9.75

‘Well, this time should be in no doubt that Chinese perform the best,’ Valerie did some rough calculations.

Time for Climbing up a 8-meter wall with tying a rope on their own (min)		
	Russian	Chinese
Range	9.50	7.75
25% quartile	3.00	2.8125
50% quartile	5.00	4.75
75% quartile	8.10	7.90
Inter-quartile range	5.10	5.075

‘I did according to what you’ve said, boss. We should calculate the range first to estimate which side is better. It is clear that Chinese has a lower range. Besides, all the three quartiles and inter-quartile range show that Chinese mercenaries need less time to climb up the wall,’ Valerie added.

Tiffany surprised, ‘Oh, Russian performs relatively poor in this task.’

Valerie continued, ‘Actually, this task is much important than other two tasks. Usually, it’s impossible that all the killers are caught at the same time. Also, using analytic skills and



investigating the surroundings can be conducted by only a few people. They can work in small groups. Thus, these small groups can still find out the password when they are getting inside the

President's House. As a result, the Chinese assassins still should be recruited.'



Perry said, 'Wait. Wait. Wait. Valerie, you should remember that team spirit is the most essential part of a team. Without it, we are bounded to fail.

Through helping each other, they can perform better because of their friendship and spirit.'

Valerie disturbed, 'Yes. I understood but ...'

Perry said, 'I know what you want to say. But there are some fatal mistakes you've made, Valarie.'

Part 6

'Firstly, let's focus on task 3. I've got no idea that why you don't use another measurement of data when you get similar result of range and inter-quartile range,' questioned Perry.

Time for Climbing up 8-meter wall with typing a rope on their own (min)		
	Russian	Chinese
Mean	5.09167	5.95167
Median	5.00	4.75

‘There is about 1-minute discrepancy between the Russian and the Chinese. So Russian performed better, didn’t they?’ continued Perry.

‘But why don’t we choose Chinese who used less time to climb up the wall?’ asked Tiffany.

‘As I’ve said, they are not individual but a team. They should have a similar ability to strike for something. Bear in mind, **our mission is to slip into**



the President’s House and assassinate the President ES.Drump.’

Valerie smiled secretly.

‘But you take a look at the performance of Chinese. Their results can be divided in 4 main groups clearly. Or maybe we can use the standard deviation to explain the dispersion of data.’

‘What is standard deviation?’ asked Valerie.

‘It is a measurement for quantifying the amount of variance in a set of data. A lower standard deviation indicates each datum tend to be closer to the mean. In this case, we get the mean and standard deviation as follows.

x_1 represents the first datum
 x_n represents the n^{th} datum
 \bar{x} represents the mean of data
 n represents the no. of data
 σ represents standard deviation

$$\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}}$$

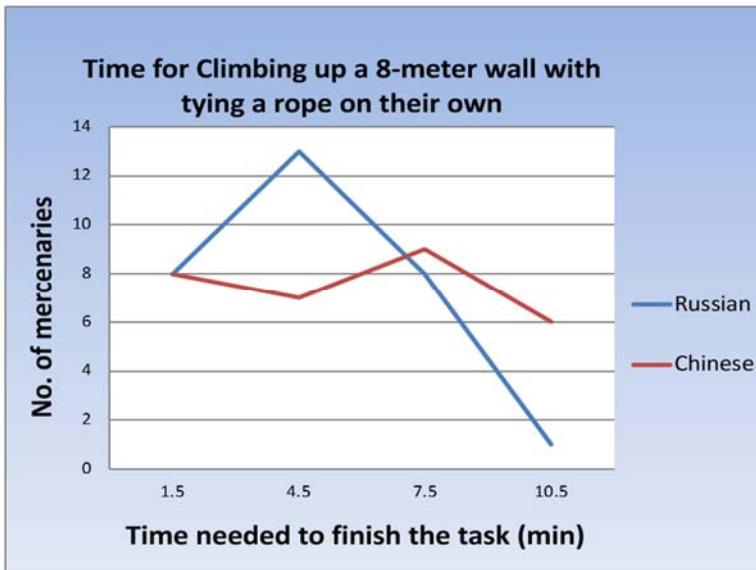
Time for Climbing up 8-meter wall with typing a rope on their own (min)		
	Russian	Chinese
Mean	5.09167	5.95167
Standard deviation	2.63116	2.72701

‘Since standard deviation of Russian is smaller, their consistency is higher than that of Chinese. Though the difference between data of two standard deviations is smaller, Russians still perform better and more reliable,’ analyzed Perry.

‘I don’t understand. Can you give me an example?’ Tiffany added.

‘Well, let me sketch a graph for you,’ said Perry.

Time for Climbing up a 8-meter wall with tying a rope on their own (min)											
Russian					Chinese						
0.50	1.00	1.10	1.25	1.30	2.00	2.00	2.20	2.50	2.75		
2.00	3.00	3.25	3.50	4.00	2.75	2.80	2.85	4.00	4.25		
4.50	4.75	4.80	4.95	5.00	4.30	4.50	4.60	4.75	4.75		
5.10	5.10	5.20	5.50	5.70	7.45	7.50	7.50	7.65	7.80		
6.00	8.10	8.10	8.10	8.20	7.90	7.90	7.90	9.00	9.10		
8.20	8.25	8.25	8.30	10.00	9.25	9.50	9.60	9.75	9.75		



As you can see, the performance of Russian is more left-shifted than that of Chinese. If we randomly send 2 killers to conduct a mission involving climbing up a wall, the probability of the killers getting over the wall within 9 min will be like this,’ analyzed Perry.

According to the table,

P(Russian killers can finish within 9 min / 2 killers climbing up a wall)

$$\begin{aligned} &= \frac{C_2^{29}}{C_2^{30}} \\ &= \frac{406}{435} \end{aligned}$$

$$= 0.933$$

P(Chinese killers can finish within 9 min / 2 killers climbing up a wall)

$$\begin{aligned} &= \frac{C_2^{24}}{C_2^{30}} \\ &= \frac{276}{435} \end{aligned}$$

$$= 0.634$$

‘The mission can be conducted in an efficient way. Besides, **time getting inside the wall of the President’s House should not be too various because our time to assassinate the President is**



extremely limited. Therefore, Russian mercenaries are much better. Though there is just only extreme value of Russian, training one killer is less time-consuming.’

Valerie smiled secretly again.

‘Secondly, what if the killer caught is worse in unlocking handcuff? There are 5 extremely slow cases in Chinese while there is only 1 in Russian. Let’s calculate the possibility of one killer who is weak at unlocking handcuff if 2 killers are caught.

Time for Unlocking Handcuff (s)	
Russian	Chinese
40 40 41 43 45	10 18 19 20 27
49 52 56 56 57	31 36 38 40 45
57 59 61 61 61	50 52 54 54 60
62 63 63 65 65	60 60 63 67 70
66 69 70 70 77	72 81 89 90 99
77 80 83 87 122	108 120 120 121 122

P(1 is weak at unlocking handcuff / 2 Russian killers caught)

$$\begin{aligned}
 &= \frac{C_1^1 C_1^{29}}{C_2^{30}} \\
 &= \frac{29}{435} \\
 &= 0.0667
 \end{aligned}$$

P(1 is weak at unlocking handcuff / 2 Chinese killers caught)

$$\begin{aligned}
 &= \frac{C_1^5 C_1^{26}}{C_2^{30}} \\
 &= \frac{130}{435} \\
 &= 0.29885
 \end{aligned}$$

‘You see? It will be far more dangerous if we send Chinese killers without training to execute mission. However, it seems be time-consuming for training 4 rather than training 1.’

DANGEROUS

‘My dear, you’re fantastic! You get the strong foresight!’ appreciated Tiffany.

‘But how about task 2? Why can’t Chinese ...?’



‘Well, actually, I have talked about it. Don’t you have listened to?’

Perry shouted angrily.

Time for escaping secret chamber (min)										
Russian					Chinese					
20.00	29.50	29.95	32.15	32.40	20.50	34.00	35.50	35.75	36.00	
33.00	33.00	34.00	34.00	34.25	36.50	37.00	37.50	37.90	38.00	
35.00	35.50	36.00	36.50	38.00	38.25	38.80	39.00	39.00	39.00	
38.00	38.00	39.50	39.50	40.00	39.90	40.00	40.75	41.00	42.30	
40.00	40.00	40.00	42.50	43.00	44.00	47.00	49.00	49.20	50.50	
45.00	45.50	47.00	48.50	80.00	51.00	52.50	53.00	54.50	61.00	

Time for escaping secret chamber (min)		
	Russian	Chinese
Mean	38.658	41.945
Standard deviation	9.5872	7.8438

‘However, according to the standard deviation, Chinese really performed better!’ Valerie said with doing the calculations.

Tiffany surprised, ‘Oh really? Perry, did you estimate wrongly?’

‘Well, No. According to your analysis, each datum of Chinese is close to each other. However, the mean of Chinese is much larger than that of Russian. What if we remove one extreme value on both sides? Continued by Perry.

Time for escaping secret chamber (min)										
Russian					Chinese					
20.00	29.50	29.95	32.15	32.40	20.50	34.00	35.50	35.75	36.00	
33.00	33.00	34.00	34.00	34.25	36.50	37.00	37.50	37.90	38.00	
35.00	35.50	36.00	36.50	38.00	38.25	38.80	39.00	39.00	39.00	
38.00	38.00	39.50	39.50	40.00	39.90	40.00	40.75	41.00	42.30	
40.00	40.00	40.00	42.50	43.00	44.00	47.00	49.00	49.20	50.50	
45.00	45.50	47.00	48.50	<u>80.00</u>	51.00	52.50	53.00	54.50	<u>61.00</u>	

	Russian (Without datum 80)	Chinese (Without datum 61)
Mean	37.23	41.28
Standard deviation	5.8409	7.1199

‘Wow!’ The standard deviation becomes smaller!’

‘Of course. Since there is a datum in each group much larger than the mean, standard deviation should be reduced. Overall performance is better as Russians’ mean is smaller. Also, since the standard deviation becomes smaller, the consistency is higher. We can get a mercenary having similar performance easier,’ Perry analysed.

‘We may refer to this graph. Actually, the data of Russian tends to be on the left hand side except an extreme value,’ Perry continued.



‘If escaping the secret chamber within 35 minutes is not standard, most of them cannot pass the mission. We may have to find more killers with higher ability of analytic skills. If we need to find 100 more people escaping the secret chamber within the standard (i.e. 35 minutes), we use normal approximation. Just like this.’

Let R be time needed for a Russian killer escaping the secret chamber,
then $R \sim N(38.658, 9.5872^2)$

$$P\left(R < \frac{35 - \bar{x}}{\sigma}\right) = \frac{100}{n}$$

Total number of killers



$$P\left(R < \frac{35 - 38.658}{9.5872}\right) = \frac{100}{n}$$

$$P(R < -0.38) = \frac{100}{n}$$

$$P(R \leq 0) - P(-0.38 \leq R \leq 0) = \frac{100}{n}$$

$$0.5 - 0.1480 = \frac{100}{n}$$

$$n = 284.09$$

Let C be time needed for a Chinese killer escaping the secret chamber
then $C \sim N(38.658, 9.5872^2)$

$$P\left(C < \frac{35 - \bar{x}}{\sigma}\right) = \frac{100}{n}$$

Total number of killers



$$P\left(C < \frac{35 - 41.945}{7.844}\right) = \frac{100}{n}$$

$$P(C < -0.88) = \frac{100}{n}$$

$$P(C \leq 0) - P(-0.88 \leq C \leq 0) = \frac{100}{n}$$

$$0.5 - 0.3106 = \frac{100}{n}$$

$$n = 527.98$$

‘As you see, we only need to find 284 Russians, and we can get 100 extraordinary killers. For Chinese, we need to find 528 Chinese, but only get 100 excellent killers. The cost of finding 284 Russians is much cheaper than that of 528 Chinese which will be quite time-consuming and less convenient. Thus, Russians should be our only option,’ said Perry.

‘My dear, you are the best!’ appreciated Tiffany.

Valerie remained silent.



Part 7

Bang! Suddenly, the main door of the house was broken. A barrage of police holding machine guns rushed into the room and encircled them.



‘Put everything down and hand up or we shoot!’ shouted the police.

‘Wait! Wait! Wait! What’s up?’ Perry asked.

‘What’s going on?’ Tiffany yelled.

Valerie still remained silence.

‘According to the intelligence, we believe that you are going to participate in an assassination.’

‘No, we don’t! Does anything go wrong? We’re just citizens abiding by laws!’ argued Perry.



‘...Our mission is to slip into the President’s House and assassinate the President ES.Drump.’ Valerie was transmitting the sound record from her mobile. **‘...time getting inside the wall of the President’s**

House should not be too various as our time to assassinate the President is too limited. Therefore, Russian mercenaries are much better.’

She stopped her phone and put it into her pocket.

‘Why do you betray me? I’m your best friend!’ cried Tiffany.

‘Nope. My life has already been consecrated to my country. Anyone who want to assassinate the President of my country, he or she is killing my country. No matter who he or she is, I must put him into the jail until the end of his or her life!’



‘The crazy regime will be never success! There will always be someone to oppose the political power,’ Perry yelled angrily.

‘Oh really? Let’s see in the future. By the way, in my deepest notion,



actually I want you to choose Chinese mercenaries for your assassination. Therefore, our America government can have an excuse to attack and invade

China with nuclear power. I have been using Pygmalion effect to lure you choosing Chinese but only Tiffany is lured.'

'Pygmalion effect? What's it?' asked Tiffany.

'It's a psychological theory. It means that when someone tells you her opinions with emotional words or expresses her ideas indirectly and implicitly, he has actually affected our minds and action. Just like you, Tiffany. When Valerie said Chinese should be chosen with statistics, which actually misleading us, you deeply believed.' Perry explained.

'Wow! Incredible! I've never expected you are that smart! But still, caught by me,' laughed Valerie.

'I am a human, loyal to statistics. You? A cold and ruthless dog!' Perry shouted, 'We still have the chance. There was lots of Presidents being assassinated before such as Abraham Lincoln died in 1865, James Abram Garfield died in 1881, John Fitz Gerald Kennedy died

in 1963 and so on. By Murphy's Law, anything that can go wrong will go wrong. If someone has succeeded to assassinate the President, there will always be another one who can kill this Trump finally...' said Perry.



'Wow, that's a funny joke.' Valerie said.

'Though that person may not be one of us, we deeply believe that there is always a Savior in our society,' said Tiffany.

'By self-fulfilling theory, if we expect something would happen, it will finally come true! This is our prophecy! Let's see in our future' said Perry.



'That's enough! Put these two criminals into the jail immediately. I don't want to hear their voices anymore!' shouted Valerie discontentedly and angrily.

Perry and Tiffany were then pulled away by the police force. Soon, the room was empty except Valerie. Everything becomes silent.

(No. of words: 2499)

Reference

1. 2018 rank of president
<http://m.ltn.com.tw/news/world/breakingnews/2344851>
<http://news.tvb.com/world/5a8bf5a5e60383fc613f8b7e>
2. Standard deviation
https://en.m.wikipedia.org/wiki/Standard_deviation
3. Assassinated President
<http://www.timetw.com/4938.html>
<https://zh.m.wikipedia.org/wiki/詹姆斯·艾布拉姆·加菲爾德>
<https://zh.m.wikipedia.org/wiki/亞伯拉罕·林肯>
4. Murphy's Law
<https://zh.m.wikipedia.org/zh-hk/墨菲定理>
5. Pygmalion effect
<https://td026544.pixnet.net/blog/post/32006461-皮格馬利翁效應-暗示的力量>
https://en.m.wikipedia.org/wiki/Pygmalion_effect

優異作品：

常見的罕有

學校名稱：順德聯誼總會李兆基中學

學生姓名：陳心妍、劉欣桐

指導教師：許俊江老師

概率與我們日常生活息息相關，天氣、風險、經濟、遊戲等等都與概率有關。有些時候，我們會以直覺方式解讀概率，例如發生極低概率的事情是否很不可思議？此文章希望引起大眾探討對概率的誤解。

去年蘋果公司推出新產品 iPhone X，因主打功能 Face ID 而聞名，即可用手機主人的臉部特徵解鎖，吸引大量消費者。近日流傳一宗有關 iPhone X 的新聞¹，兩位青少年，其中一位打算代另一位發訊息，他望著螢幕就成功解鎖了，萬試萬靈，引起普羅大眾紛紛熱論。

蘋果公司的高層 Phil Schiller 在發佈會表示，隨機挑選另一個人能成功解鎖你的 iPhone X 的概率只有約百萬分之一，大多數

¹（詳見附件）【似樣咩？】自己手機被好友用Face ID 解鎖
蘋果曾稱機會率僅百萬分之一

<https://hk.news.appledaily.com/international/realtime/article/20180211/57806685>

人直觀地認為百萬分之一的概率很低，幾乎沒有可能發生，但當事件流傳，令人質疑蘋果公司所說的概率是否準確。事實上，發生被人無意解鎖這件事的概率有多大呢？我們嘗試作一些簡單的計算。

假設 E 為被人無意解鎖的這件事，為了計算 E 的概率，即 $P(E)$ ，我們有以下的假設：

- 我們假設 iPhone X 的用戶數量為 n 。
- 每一個用戶都曾經嘗試被一個朋友解鎖，能成功解鎖的概率為 $\frac{1}{1000000}$
- 每次解鎖是獨立事件。

$$\begin{aligned} \text{首先，} P(\text{第一部失敗}) &= 1 - \frac{1}{1000000} \\ &= \frac{999999}{1000000} \end{aligned}$$

$$P(\text{連續兩部都失敗}) = P(\text{第一部失敗}) \times P(\text{第二部失敗})$$

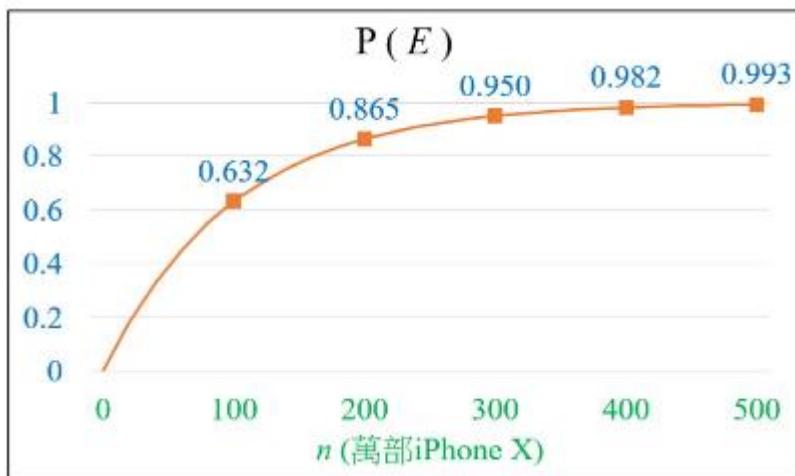
$$= \left(\frac{999999}{1000000} \right)^2$$

$$\text{如此類推，} P(\text{連續 } n \text{ 部都失敗}) = \left(\frac{999999}{1000000} \right)^n$$

$$\begin{aligned} P(E) &= P(n \text{ 部最少有 } 1 \text{ 部成功}) \\ &= 1 - P(\text{連續 } n \text{ 部都失敗}) \end{aligned}$$

$$= 1 - \left(\frac{999999}{1000000} \right)^n \dots\dots(*) ,$$

下圖顯示 n 與 $P(E)$ 的關係。



能成功解鎖的概率雖然很低，但當 iPhone X 的數量增加到 100 萬部的時候，最少一部成功的概率已經達到 0.632，當 iPhone X 的數量增加到 500 萬部的時候，概率已是 0.993。根據研究機構 Canals 的資料，2017 年第 4 季蘋果 iPhone X 出貨量達到 2900 萬支²，套用上述(*)的算式計算出成功被他人解鎖的概率竟高達 0.999999999999746！原來我們幾乎可以肯定這件事是會發生的，這則新聞是「必然」會出現。看似低概率的事情，但當經過很多人屢次嘗試後，會發生的機會變得愈來愈有可能，

²Media alert: Apple ships 29 million iPhone Xs in Q4 2017

<https://www.canals.com/newsroom/media-alert-apple-ships-29-million-iphone-xs-q4-2017>

情形就好像六合彩一樣，中頭獎的機會微乎其微，但當買彩票的人不勝枚舉的時候，便相應提高有人中頭獎的機會，你不會覺得中頭獎這事很稀奇，放諸解鎖事件道理一樣。

由此可見，大眾經常被「百萬分之一」等字眼所誤導，誤以為概率渺茫，沒作深入了解，事實上，世上「不可思議」的事情多的是呢！

(字數：782)

附件：

【似樣咩？】自己手機被好友用 Face ID 解鎖 蘋果曾稱概率僅百萬分之一

2018 年 02 月 11 日



蘋果 iPhone X 其中一個主打新功能 Face ID，原理是透過辨識用戶臉部的構成特徵，提供安全的認證方式。英國威爾特郡一位男子卻驚訝發現即使好友長相與他不甚相似，也能成功用 Face ID 解鎖他的 iPhone X，明顯存在安全隱憂。23 歲的克萊頓（Joe Clayton）與 22 歲的布徹（Brad Butcher）是青梅竹馬，二人早前一同到健身室健身，克萊頓意外受傷，於是將電話交給布徹叫他代發訊息通知朋友。克萊頓說：「他望著螢幕就成功解鎖了，萬試萬靈。其他人的臉不行，只有布徹可以，但我們並不長得那麼相似。」克萊頓及布徹都有深色頭髮及啡色眼睛，但二人的臉形、鼻子及嘴巴明顯不同。克萊頓說：「我們都有粗眉，但他的額頭較長。我認為有人說我們相似，是因為我們的下巴輪廓及深色頭髮，但 iPhone X 本來就是要辨認臉容深度，是完全不同的一回事。」成功解鎖意味著布徹可用克萊頓的電話進行身份認證及購物付款，本來有意購入 iPhone X 的布徹見 Face ID 功能如此簡單被破解後，也就打消念頭。蘋果網站資料顯示，「若在人口中隨機挑選一個人，看著您的 iPhone X 並使用 Face ID 進行解鎖」，成功率約為一百萬分之一，以力證是安全，不過公司卻拒絕評論克萊頓的遭遇。

英國《每日郵報》

優異作品：

報章「掩眼法」

學校名稱：順德聯誼總會李兆基中學

學生姓名：林果仙、蔡雅媛

指導教師：許俊江老師

摘要：

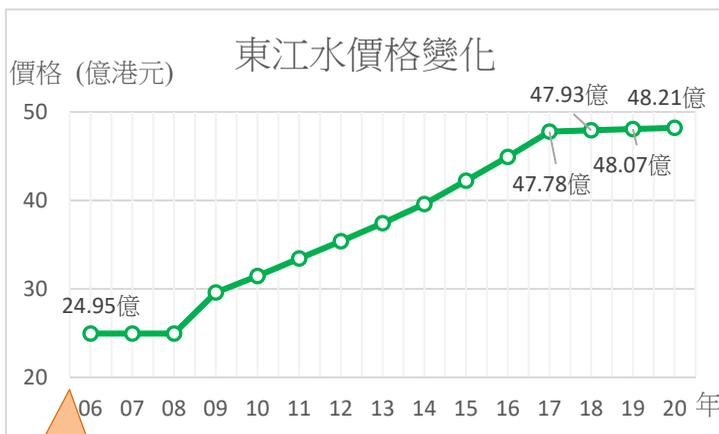
今時今日，社會各界講求數據及科學化以加強自己的觀點。傳媒為增加銷量而使用誤導字眼，吸引讀者閱讀文章。若傳媒誤用數據，文章的公信力及可信性則大打折扣。

引言

眾所周知，本港食水供應主要來自東江水，附件的報章提到「去年尾發展局與廣東省簽訂新供水協議，雙方同意未來3年(2018至2020年)水價每年加0.3%，總開支達144.2億元，較現行3年協議的134.9億元增加近7%。」。可惜，報章未能正確分析及表達數據，有誤導讀者之嫌。

一：縱軸不由 0 開始

圖一：報章上的圖表



縱軸由
20 億作起點

為了顯示數據的真實情況，統計圖中的數值尺度須由 0 開始。但圖一的數值較大而差異小，因此報章採用較大的尺度，將縱軸由 20 億作起點繪製統計圖，令讀者更容易觀察及比較之間的差異。不過這會誇大數據間真正差距而引致錯覺。為表達數據的真正關係，同時又能突顯它們之間的差異，其中一個常用的方法是使用折斷尺度的統計圖。這可方便讀者觀察數字間的差異，同時也能提醒讀者留意統計圖的表達方法或許會引起的錯覺。以下的圖二乃正確做法：

圖二：加入折斷尺度的折線圖



二：標題誇大數字

根據報章內容，標題為「東江水加價 7% 不按量收費」，驟眼一看，以為一年加價 7%；細看之下，內容實指 2018 至 2020 年的水價總額比 2015 至 2017 年的多了近 7%。然而標題並沒有標示數據的統計期，容易誤導讀者。另外，按年變動百分率會讓人較容易理解，見表一：

表一：水價的按年增幅

年份	固定水價總額(億港元)	按年增幅(%)
2006	24.95	---
2007	24.95	0.00%
2008	24.95	0.00%
2009	29.59	18.60%
2010	31.46	6.32%
2011	33.44	6.29%
2012	35.39	5.83%
2013	37.43	5.76%
2014	39.59	5.77%
2015	42.23	6.67%
2016	44.92	6.37%
2017	47.78	6.37%
2018	47.93	0.31%
2019	48.07	0.29%
2020	48.21	0.29%

事實上，水價是每三年商議，另一做法是計算歷來三年水價總額的增幅，再將增幅年度化，見表二：

表二：3 年水價總額的變動百分率及年度化的變動百分率

年份	3 年水價總額(億港元)	3 年水價總額的增幅	年度化的增幅**
2006-2008	74.85	---	---
2009-2011	94.49	26.2%	8.1%
2012-2014	112.41	19.0%	6.0%
2015-2017	134.93	20.0%	6.3%
2018-2020	144.21	6.9%	2.2%

**設 $r\%$ 為3年水價總額的增幅，年度化的增幅 $\sqrt[3]{(1+r\%)}-1$ 。

若以年度化的增幅來看，東江水價較前一協議每年增加約2.2%。

三：考慮通脹及匯率變動

若要判斷加幅是否合理，還要考慮通脹及匯率等因素。見表三：

表三：水價的實際按年增幅（扣除人民幣匯率及香港通脹兩項因素）

	固定水價總額	人民幣兌 港幣匯率	固定水價總額	綜合消費 物價指數	固定水價總額 (按 2006 年物價計 算)	實際 按年增幅
算式 年份	(a) (億港元)	(b) (港元)	(c) = (a) x (b) (億人民幣)	(d)	(e) = (c) / (d) * 74.7 (億人民幣)	以(e)計算
2006	24.95	0.9742	24.3	74.7	24.3	---
2007	24.95	1.0300	25.7	76.2	25.2	3.6%
2008	24.95	1.1236	28.0	79.5	26.3	4.6%
2009	29.59	1.1343	33.6	80.0	31.3	19.0%
2010	31.46	1.1503	36.2	81.8	33.0	5.4%
2011	33.44	1.2069	40.4	86.1	35.0	6.0%
2012	35.39	1.2304	43.5	89.6	36.3	3.7%
2013	37.43	1.2635	47.3	93.5	37.8	4.1%
2014	39.59	1.2590	49.8	97.7	38.1	0.9%
2015	42.23	1.2299	51.9	100.6	38.6	1.2%
2016	44.92	1.1664	52.4	103.0	38.0	-1.5%
2017	47.78	1.1552	55.2	104.5	39.5	3.8%
2018	47.93	---	---	---	---	---
2019	48.07	---	---	---	---	---
2020	48.21	---	---	---	---	---

由於東江水由內地供港，價格以人民幣計算較為合理。另外，若要考慮水價對香港的承擔能力的影響，可計算實際按年增幅

(即要扣除因通帳而減低貨幣購買力的影響)，若以實際增幅來看，東江水價由 2016 到 2017 年增加 3.8%。

總結

最後還可加入實際用水量、香港人口增長、內地通脹、水質、供水的穩定性、工商業對水的需求、香港集水區的集水量等考慮因素。

本文並非就東江水價格的高低作任何立場，只是想透過此事例說明一幅統計圖表所表達的資訊有限，而且數據受多方面因素影響，作出判斷時，考慮宜更全面。

(790 字)

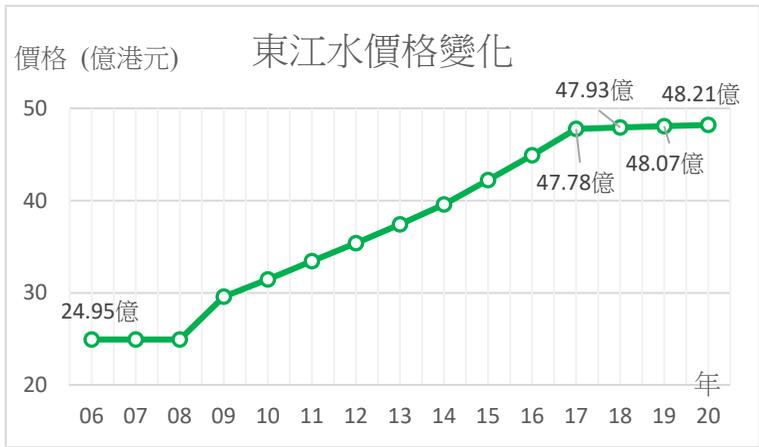
附件：

東江水加價 7% 不按量收費

明報 2017 年 11 月 22 日

網址連結：

https://news.mingpao.com/pns/dailynews/web_tc/article/20171122/s00002/1511288038655 (重繪下圖)



現行東江水供水協議今年底屆滿，發展局表示經磋商後，將與廣東省簽訂新供水協議，雙方同意未來3年(2018至2020年)水價每年加0.3%，總開支達144.2億元，較現行3年協議的134.9億元增加近7%。當局亦稱，新協議沿用「統包總額」而非「按量付費」，但指雙方已同意檢討收費模式。

港3年共多付9.3億

現時本港食水供應有70%至80%購自東江水。發展局向立法會提交文件，指在新供水協議下，2018至2020年水價分別訂為47.93億元、48.07億元及48.21億元，每年供水量上限維持8.2億萬立方米。

供水上限不變 港府：加價合理

發展局解釋，東江水價調整參照營運成本、匯率，以及粵港物價指數變化，當中 2014 至 2016 年人民幣兌港元匯率平均按年變動約-2.29%，同期粵港物價指數平均按年變動+2.66%，認為水價調整合理。

現行協議採用「統包總額」購水，香港每年支付固定費用，即使未用盡供水量，粵方亦不會退款，而過去數年本港均未用盡上限，有意見要求改用「按量付費」。發展局表示已獲粵方同意檢討收費模式，但涉及問題複雜，需研究對本港供水穩定影響、釐定單位水價等，因此新協議維持「統包總額」，雙方同意完成檢討後，就 2020 年後供水協議付費方式再作討論。

粵：不訂明供水量 難保乾旱供應

發展局引述粵方認為，若不在供水協議訂明每年供水量，將難確保本港在乾旱時可獲充足供水；並指若採用「按量付費」，收費最終或較現行「統包總額」為高。

立法會議員劉國勳認為，目前水費計算方式欠透明，對新協議沿用「統包總額」感失望。另一立法會議員毛孟靜促請政府採用「按量付費」模式購水。

優異作品：

「平均」中位數

學校名稱：香港真光書院

學生姓名：鄭曉欣、朱穎瑤

指導教師：關穎楠老師

在一次聚餐中，陳太問兒子今年的每月工資，兒子以討論甚麼是平均數和中位數回覆媽媽的提問的情況。

「平均」中位數

在年尾的一次聚餐，陳太忽然問：「兒子，終於都放年假啦。你今年每月工資有多少呀？」

兒子想了想，有點遲疑，眼神閃縮地說：「今年每個秘書平均月薪有三萬元。」

陳太說：「今年哪一個秘書有幸中了年終大獎？」

兒子答：「老闆說今年年終大獎是所有秘書月薪的中位數。」

陳太追問：「那月薪中位數是多少呀？」

兒子回應：「正是三萬元。」

陳太說：「那豈不是「太公分豬肉」，人人都有年終獎？」

兒子說：「媽，這您就錯了。讓我介紹一下平均數和中位數的分別吧！平均數是最常用的平均值，一組數據的平均數是所有數據的總和除以數據的數目所得的結果。平均數常用於表示統計對象的平均一般水平。我們既可以用它來反映一組數據的整體情況，也可以用它進行不同組數據的比較，可以更容易看出組與組之間的差別。」

陳太疑惑地問：「平均數和中位數有甚麼分別呢？」

兒子笑著回答：「你先別著急，中位數是另一種常見的平均值，中位數是一組按大小順序排列的數據中，位於中間位置的數值。如果數據的總數是奇數，中位數就是位於中間的數據，若數據的總數是偶數，中位數就是位於中間的兩個數據的平均數。」

兒子看一看媽媽，並對她說：「媽媽，你試舉一個例子，看看你究竟明不明白平均數和中位數的分別吧。」

媽媽思考了一會兒說：「我用今天到超級市場買東西作例子吧。今天我買了一袋九元的青瓜、一袋十二元的麵包、一袋十八元的餃子、一盒二十元的免治豬肉和一袋一百二十六元的米，它們價錢的平均數是價錢的總和除以項數，……等於三十七元，

而中位數是價錢排列順序中間的餃子，即十八元。」

兒子說：「沒錯，真聰明，難怪能生出我這種高智商的孩子。」

陳太笑著說：「當然！」

兒子看見媽媽洋洋得意的樣子，又繼續向媽媽解釋道：「在日常生活中，我們不難發現平均數或中位數誤用的地方。例如：在標題為"回歸以來青年置業最困難"的報道中，提到 2016 年 25 歲至 34 歲青年主要職業每月收入中位數為 1.5 萬，兩人家庭即三萬元。當中，兩人家庭每月入息中位數的計算方法有錯，因在 25 歲至 34 歲青年的收入中，有一半人的收入比 1.5 萬元高，亦有一半人比 1.5 萬元低，但結婚的不一定是每月收入 1.5 萬的男和 1.5 萬的女，所以我認為不能以 1.5 萬元相加作為 25 歲至 34 歲青年兩人家庭主要職業每月入息的中位數。中位數只是一組按大小順序排列的數據中處於中間位置的數值，在此情況下兩人家庭每月入息中位數應考慮 25 歲至 34 歲青年中兩人家庭的入息排列而得。另外在"香港 01"報道的"「新香港人」夫婦割房搬上海景樓 內地人妻：羨慕港男生有條件玩"一文中，提及香港月入一萬五千元或以下達五成五人口，而北京人均工資七千多，但花費比香港少三倍。這裡描述的香港月入一萬五千元以下人口和北京人均工資數目不能直接比較，是將平均數和中位數混為一談。如需比較，比較的數目需要相同單位。例

如可以直接比較香港人的平均工資和北京人的平均工資。」

媽媽聽完後似懂非懂的樣子，又問道：「甚麼平均數、中位數，其實你每月工資是多少？」兒子見狀，馬上三十六計走為上計。

(1175 字)

參考資料：

1. <https://www.hk01.com/%E5%A5%B3%E7%94%9F/149065/-%E6%96%B0%E9%A6%99%E6%B8%AF%E4%BA%BA-%E5%A4%AB%E5%A9%A6%E5%8A%8F%E6%88%BF%E6%90%AC%E4%B8%8A%E6%B5%B7%E6%99%AF%E6%A8%93-%E5%85%A7%E5%9C%B0%E4%BA%BA%E5%A6%BB-%E7%BE%A8%E6%85%95%E6%B8%AF%E7%94%B7%E7%94%9F%E6%9C%89%E6%A2%9D%E4%BB%B6%E7%8E%A9>

邀請作品：

港式女士品茶

想必大家都品嚐過茶餐廳的港式奶茶，港式奶茶又稱「絲襪奶茶」，由殖民地時代的英式奶茶，經過百年的逐漸演變，成為今天大家熟識的飲品。港式奶茶製作技藝更反映了中西飲食文化在香港的交融與發展，是香港的非物質文化遺產。港式奶茶製作技藝包括調配茶葉、煲茶、撞茶、座茶、撞奶等步驟，這些步驟對於奶茶的品質和味道都非常重要。

巧合的是統計學發展上也有一段有關奶茶製作的經典小故事，讓我們先在此重溫這段趣事，並簡單敘述它和現今常見的假設檢定的不同之處。

經典女士品茶

故事發生在一個風和日麗的下午，幾位女士在英國街頭一如平日享受她們的下午茶，並談到奶茶的製作方法。其中一位女士妙麗葉兒·布里斯托爾(Muriel Bristol)提到「撞奶」此一步驟，即是將泡好的紅茶撞入常溫牛奶，如果調轉把牛奶加入紅茶，味道會大不相同。但其餘幾位女士就對此說法不以為然，兩種方法都是把紅茶和牛奶溝在一起，怎麼會有分別呢？用化學角度

去看這個問題，同樣是兩種液體，把液體甲加進液體乙和把液體乙加進液體甲，又怎麼會有不同的結果呢？在用化學角度去思考這個問題前，先去想的是我們如何得知這兩種製作方法是否真的對奶茶的味道有所影響。

劇情發展說到當時一位統計學家羅納德·愛爾默·費雪 (Ronald Aylmer Fisher) 經過並聽到這一問題，並提出設計一個實驗去證明布里斯托爾女



士的說法。他提出準備 8 杯奶茶，其中 4 杯是用「撞奶」的方法製作，另外 4 杯是把牛奶加入紅茶製作而成，再以隨機的次序讓布里斯托爾用舌頭去分辨那 4 杯是「撞奶」奶茶。這些實驗設計的資訊亦都對布里斯托爾清楚說明，當然她唯一不知道的就是「撞奶」奶茶的出現次序。

這個實驗中奶茶的組合方式有 $C_4^8 = 70$ 種，而正確的組合只有一種，估錯一杯的可能組合是 $C_3^4 C_1^4 = 16$ 種，如下表所示。

正確數目	可能組合的種類
4 杯	$C_4^4 C_0^4 = 4$

正確數目	可能組合的種類
3 杯	$C_3^4 C_1^4 = 16$
2 杯	$C_2^4 C_2^4 = 36$
1 杯	$C_1^4 C_3^4 = 16$
0 杯	$C_0^4 C_4^4 = 1$

費雪的思考是在這個假設檢定中的虛無假設是布里斯托爾沒有分辨「撞奶」奶茶的能力，若果實驗結果是她能完全正確分辨出 4 杯「撞奶」奶茶的話，單靠運氣的機率是 $1/70$ 或 1.43%，這就是這個假設檢定中的 p 值，即虛無假設下 4 杯全中的機率。若果結果是她只能分辨出 3 杯「撞奶」奶茶， p 值則是 $17/70$ 或 24.29%，即虛無假設下正確分辨 3 杯或 4 杯的機率。費雪認為觀察到越小的 p 值可令我們越有力去拒絕虛無假設。

傳聞故事的結局是布里斯托爾能夠準確分辨出該 4 杯「撞奶」奶茶，費雪因此拒絕了這個假設檢定中的虛無假設，可見我們也有理由相信兩種製作方法是真的對奶茶的味道有所影響，難怪港式奶茶仍舊保存此「撞奶」步驟，也算是不無原因吧！

現今常見的假設檢定

說到這裡，對假設檢定有所認識的人可能會問，「為甚麼沒有對立假設？」和「說好的統計顯著性水平呢？」。

的確，在費雪當時的假設檢定模型中並沒有對立假設和定立確切的統計顯著性水平，如上文所說，他認為只需通過觀察所得的 p 值去做拒絕虛無假設的判斷。我們都能夠在費雪的故事中看到現今常見的假設檢定的影子，其後再有另外兩位統計學家耶日·內曼(Jerzy Neyman)和埃貢·皮爾遜(Egon Pearson)提出包含對立假設和統計檢定力等元素的假設檢定方法，有如奶茶一樣經過多年的演變，變成現今常見的假設檢定。讓我們由百年前的英國回到今天的香港，說一個有關港式奶茶的小故事。

品味港式奶茶的假設檢定

小河已經在茶餐廳打滾了四十多年，是港式奶茶製作技藝的老師傅。他對併茶、煲茶、撞茶、座茶、撞奶等步驟一絲不苟，因為他相信客人是可以分辨一杯奶茶是好是壞。另外，小河師傅更自豪就是他的品茶技巧，他常常在兒子面前吹噓，一杯奶茶的質素，用的茶葉品質如何、併茶份量、撞茶力度等，全部都喝得出來。

要把故事說下去，我們都知道需要一個懂得假設檢定的人去驗

證小河師傅的說法，正如大家預期一樣，他的兒子正在大學修讀統計學，並決定效法費雪重演一次女士品茶的實驗。同樣，測試都是「撞奶」這一個最後步驟，先加淡奶後加茶還是先加茶後加淡奶。

小河師傅一聽到這個考牌試，已經充滿信心地說道：「只有外行人才會犯上先加茶後加奶這種低級錯誤，我又怎會喝不出來呢？我有 9 成的把



握。」說到這裡，我們可以理解為小河師傅認為他分辨「撞奶」奶茶的能力是依從成功的概率(p)為 0.9 的伯努利分佈，因此兒子定立這個假設檢定中的虛無假設為($p = p_0 = 0.9$)，而對立假設則是($p < p_0 = 0.9$)，即小河師傅並無如他聲稱高達 0.9 成功率的分辨「撞奶」奶茶能力。值得一提的是，這是一項單尾檢定。

從費雪女士品茶的故事當中，我們知道實驗的設計相當重要。而兒子的實驗設計也是準備 8 杯奶茶，不同的是該 8 杯奶茶的製作方法都是獨立的，換句話說，他並沒有如費雪一般去限制其中 4 杯是用「撞奶」的方法製作，另外 4 杯則把淡奶加入紅

茶，每杯奶茶的製作方法都是隨機的。

檢驗統計量(X)就是成功分辨的次數，在統計顯著性水平(α)初步定立為 5%的情況下，拒絕域則為 $\{X \leq 5\}$ 。事實上，當拒絕域

為 $\{X \leq 5\}$ 的時候，而 X 依從二項分佈 $\text{Bi}(x; 8, 0.9)$ ， α 實則上為，

$$\sum_{x=0}^5 C_x^8 0.9^x (1-0.9)^{8-x} = 3.8\%$$
 這亦是犯了第一型錯誤（錯誤地拒絕虛無假設）的概率。

假若小河師傅只是自吹自擂，實際上他只能靠瞎猜去分辨「撞奶」奶茶，即對立假設下 $p = 0.5$ 的情況，則 X 依從二項分佈 $\text{Bi}(x; 8, 0.5)$ ，那犯下第二型錯誤（錯誤地未能拒絕虛無假設）的

概率 (β)則為
$$\sum_{x=6}^8 C_x^8 0.5^x (1-0.5)^{8-x} = 14.5\%$$
。因此，這個假設檢定的統計檢定力則為 $1 - \beta = 85.5\%$ 。

由此可見，內曼和皮爾遜的貢獻使我們對運用假設檢定模型有了更多的掌握，尤其是我們可以利用對立假設去尋找假設檢定的統計檢定力。在這個實驗中，兒子用上了 3.8%較為嚴格的統計顯著性水平，犯下第二型錯誤的概率卻有 14.5%。實際上，當其他情況不變下， α 下降會導致 β 上升，反之亦然。

品茶的故事來到最後，老師傅當然不負眾望，成功分辨出 8 杯奶茶的製作方法， p 值為 1 比 α 為 3.8% 大，兒子當然無法在此統計顯著性水平拒絕這個假



設檢定中的虛無假設，未能推翻老父的說法。小河師傅最後再補上一句：「先落茶後落奶，淡奶都凝固成薄片狀浮現奶茶表面，看都看得出來，怎會分辨不了呢！」

參考資料:

1. 港式奶茶製作技藝
香港非物質文化遺產代表作名錄
https://www.lcsd.gov.hk/CE/Museum/ICHO/zh_TW/web/icho/representative_list_milk_tea.html
2. 沖調技藝三部曲 港式奶茶大學問
<https://www.afoodieworld.com/foodie-ch/7500>
3. David Salsburg. The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century. 2016 年中譯本 江西人民出版社
4. Perezgonzalez JD (2015). Fisher, Neyman-Pearson or NHST? A tutorial for teaching data testing.
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4347431/>
5. E.L. Lehmann (1992). The Fisher, Neyman-Pearson Theories of Testing Hypotheses: One Theory or Two?
<http://digitalassets.lib.berkeley.edu/sdtr/ucb/text/333.pdf>

邀請作品：

背叛，還是合作

故事發生在 1950 年代的美國，彼得和艾倫涉嫌打劫銀行而被警方拘捕。由於證據不足及他們選擇保持沉默，警方顯得束手無策。這時聰明的警長分開他們囚禁，並分別與二人見面，他首先向彼得說以下的話：

「如果你們雙方都保持沉默，我們也有權關你們在這裡半年，配合我們的調查。」

「如果你認罪並檢舉艾倫，而他保持沉默的話，你將會被無條件釋放而艾倫將被監禁十年；反之，如果艾倫認罪並檢舉你，而你保持沉默的話，你將被監禁十年而他會被無條件釋放。」

「如果你們都認罪並互相檢舉，我們會向法官求情，你們將會被判監五年。」

「當然你不會知道艾倫的選擇，我給你一晚時間考慮吧！」

警長然後對艾倫說相應的話。這一晚，彼得和艾倫因為警長開

出的條件而茶飯不思，費煞思量。同時，彼得心中的天使與魔鬼不斷在他們的腦海中遊說。

「選擇沉默吧！艾倫是與你出生入死的好兄弟，他一定不會背叛你的。不要忘記，只要你們都保持沉默，你們只需被關半年！」

「你是傻的嗎？在最關鍵的時刻還說甚麼義氣？你不背叛他，你能擔保他不會背叛你嗎？你甘心被關十年但他在外面風流快活嗎？如果你仔細分析的話其實不難發現，若果他沉默，你背叛他會讓你獲釋。若果他檢舉你，你沉默就被關十年，他就逍遙法外；檢舉他就一起被關五年。無論艾倫選擇甚麼，檢舉他還是比較划算的啊！」

翌日，警長分別詢問彼得和艾倫的選擇。最終他們都要被關在同一間囚室內五年。在囚室內，彼得和艾倫以仇視的眼光望向對方，不發一言，終於艾倫忍不住道：「枉我和你出生入死這麼多年，想不到你竟然背叛我……如果我們合作死不認罪，根本不用關那麼久……」彼得聽後怒火中燒，說：「你也不是背叛了我嗎？我就知道你會背叛我，我才要自保呢！」艾倫無言以對。

以上是非零和博弈論中最經典的「囚徒困境」。理性的彼得和艾倫在作出選擇時，通常以個人的利益為依歸。即使他們知道雙

方合作可以令他們的利益最大化，但是由於彼此猜疑的關係，最終他們雙雙背叛對方，達到此博弈論的納什均衡 (Nash Equilibrium)。

	艾倫		
		沉默	認罪並檢舉對方
彼得			
	沉默	半年	獲釋
	認罪並檢舉對方	十年	五年
		獲釋	五年

當然「囚徒困境」的故事純屬虛構，但是「囚徒困境」背後的理論能充分反映在現實的例子中。例如簡單的「囚徒困境」理論可以解釋為甚麼女士們都喜愛化妝。化妝是傷害皮膚的，所以化妝是會帶來 -5 的損失。但是如果甲化了妝而乙沒有化妝的話，甲會因為比乙漂亮而得到+10 的優越感，乙會因為比甲醜而得到 -10 的自卑感，反之亦然。雙方都不化妝明顯地是總體利益最大化的情況，但是由於無論對方化不化妝，化妝都會為自己帶來更佳的结果，所以最終雙方都會化妝。

女士乙 女士甲	化妝	不化妝
化妝	傷皮膚(-5)	醜而自卑(-10) 傷皮膚但有優越感(-5+10)
不化妝	傷皮膚但有優越感(-5+10) 醜而自卑(-10)	0 0

除了解釋個人行為之外，相似的理論也可以應用在大量經濟學及政治學的例子，例如價格競爭、軍備競賽、貿易戰等等。

參考資料：

<https://zh.wikipedia.org/wiki/%E5%9B%9A%E5%BE%92%E5%9B%B0%E5%A2%83>

邀請作品：

區別相關性與因果關係

統計是甚麼？若有人向閣下提出這個問題，你會怎樣回答呢？簡言之，統計是從數據中找出訊息並帶出想法。統計可幫助我們透過客觀的資料分析來理解變異來源的一門科學。隨著科技的進步，我們從日常生活中很容易獲得大量的數據，運用這些數據作研究分析較以往來得方便、省時。

我們每天閱讀報章刊物，很多時均會看到一些研究報告，有些結果與日常的認知有很大的出入，大家會怎樣對待呢？一笑置之，或認真地看其研究的細節，判斷結果是否值得參考。由於我們每天均接觸大量的資訊，我猜想大部分人均只會看完報導便算，無暇深究，報告對我們的影響其實不大。可是，若我們錯誤解讀一些研究報告的結果，便會帶來麻煩。讓我舉一個例子，若閣下接觸某個傳媒的報導，根據最近的數據顯示，聰明的人具備以下這些特點：戴眼鏡、使用 iPhone、左撇子、睡眠時間不超過 6 小時。閣下十分渴望孩子較其他同學更聰明，你會怎樣做？根據報告的結果，立刻為孩子配備一副眼鏡、買一部智能手機給他、訓練他使用左手、並限制他的睡眠時間。你當然希望孩子因而變得聰明，但實情又是否真的如此？

很多時，我們只是讀了幾篇看來是從統計分析中得出錯誤結論的研究，誤把「相關性」和「因果關係」混為一談，得出有誤導的結論。媒體時常以錯誤的方式報導科學研究的發現，說成一種因果關係，因而引發爭議。

要成為聰明的數據使用者，當面對相關性與因果關係時，我們應怎做？最近，筆者翻閱由約翰·強森和麥可·葛拉克所編寫的書《一次看懂小數據-MIT 經濟學家帶你搞懂生活中的統計數字，聰明做決定》，內裡提出一些建議，可供我們參考。

不要單看新聞報導或研究中的標題便下結論，例如：「使用智能手機便有智慧」，應詳細審視內容究竟說了些甚麼？很多時報導只是暗指因果關係，但若我們深入挖掘，就會發現研究只是提出某種相關性。

對於新聞報導的研究結果，我們應看看是否符合常理(common sense)。我們應用常理去理解任何統計分析。例如，「聰明人都用 iPhone」、「毒販越多，房價越高」、「雪糕銷售量與謀殺率之間有強烈關係」，這有道理嗎？符合我們的認知嗎？

若我們要找出兩件事件之間的關係，應小心考慮有沒有其他因

索引發所得的現象？遺漏變數是否存在，它是理解這項關係的重要因素。例如，從一個有關大學生對男女教授的評鑑研究結果發現，被評為「無趣」的教授男性居多，性別真的有差異。但我們應想想有沒有其他因素影響學生的評分呢？例如：課程與教材是否太深奧？學生能否獲得好成績？學生需要完成多少功課？是否早上 8 時上課？教授樂於指導學生？該課是必修課還是選修課？所有這些因素對學生的評分皆有影響。相信大家數年前亦聽過「適度喝酒有益健康」的研究報告，有些科學家認為其結果反映的不是與喝酒有關，反而是與喝酒人士的社會地位有關。研究發現能夠負擔喝酒這習慣而同時對喝酒有節制，不會大喝的人，一般經濟環境稍為富裕而且注意健康。所以適度喝酒這習慣可算是一個較高社會地位的象徵，而並非喝酒會為健康帶來好處。我們應時刻保持頭腦清醒，尋根究底，知道如何與何時要質疑數據。

另外，即使我們找到統計上的相關性，亦要小心**倒因為果**。例如，有一隻小狗每日清晨六時均會向著門外吠一會，幾分鐘後，垃圾車便到來收垃圾。究竟是狗吠聲引來垃圾車，還是小狗先聽到垃圾車的聲音，作出反應。後者的機會似乎大得多。

對於科學家而言，要證明因果關係實際上門檻很高，需要付出大量的時間和精力投入觀察與實驗，分析數據以證明相關性，

同時注意是否有遺漏因素。打破砂鍋問到底需要訓練和努力，要找出真相不是這麼容易。

參考資料：

約翰·強森 John H. Johnson 和 麥可·葛拉克 Mike Gluck 著，吳青榆(譯)(2017)。《一次看懂小數據-MIT 經濟學家帶你搞懂生活中的統計數字，聰明做決定》。台北：時報文化出版

邀請作品:

The Use of Neighbours in Improving Prediction

Simon K.C. Cheung

Department of Statistics and Actuarial Science,

The University of Hong Kong

The use of neighbours in providing a ball-park approximation is very common. We can find applications in valuation of a real estate, supplying the quotation of a new product and even pricing of a Lego product that was ceased in manufacturing. The underlying reasons why it works are multi-folded. Firstly, we like to compare. It happens in many of our daily life activities. First, when two people performed similarly, their performance scores should be identical. Otherwise, the one who receives a lower score would complain. Second, when we are offered a certain price for a service, it would be hard for us to accept a higher offer elsewhere if the quality and reliability of both companies are identical. Third, when our friend purchased a product, for example, an Apple watch, we always want to own an identical or similar product. The concept of comparison is deep down in all of us. It also makes the idea of neighbour intuitively appealing. Secondly, fundamental to our value system, matters are composed of features. Value of a target variable on an object is determined by the application of a set of weights on the features or functions of features. Thus, objects with similar features have similar values on the target variable. The same reason also serves to explain why multiple regression does not work sometimes. In a multiple

regression analysis, we specify a dependent variable (target variable) and a set of explanatory variables. Each explanatory variable can be considered a function of the features. Since the explanatory variables aggregately cannot cover all the aspects of the features, it follows that the residuals in the regression analysis and the explanatory variables are not independent. We often describe such situations as missing important explanatory variables. Lastly, the data generating mechanism can actually originate from a dependent process. Instead of generating the observations independently, observations are generated as a single realization. The residuals from such a regression can never be independent. Perhaps the best example of this kind of data is spatial data. Air pollution on different locations are dependent. The closer the locations, the higher is the correlation. Other examples can be found in relational databases. Companies and shareholders are linked together in a database of listed companies and movie ratings connect users and movies in a movie database.

Although we are interested in estimating the target variable of an arbitrary instance in a population, we only observe a random sample from the population. For example, we usually would like to find a fair market price of an apartment we would like to purchase or sell. But what is available is the transaction prices of a list of apartments. Another example is that we usually need to estimate a fair value of a company before it is going to be listed and such estimation comes from the market values of a set of listed companies each of which operates businesses in the same or similar industries. Thus, we arrive at the following observations. First, due to limited resources, we can only select neighbours from the observed sample.

This is in fact a source of inaccuracy in using neighbours for prediction. Second, when a network exists in the population either tangibly or intangibly, the local Markov property assumes that the conditional distribution of any observation given its neighbours is identical to that given the rest of other observations. In other words, conditional on its neighbours, an observation is independent on the rest of other observations. This conditional independence property provides a theoretical ground for us to use neighbours to form prediction of an observation outside the sample. Third, when a network is not clearly defined, the common practice is to define a distance function between any pair of observations. The belief behind the use of a distance function is that the smaller the distance between two observations, the higher is their correlation. The most commonly use distance function is the Euclidean distance function. Note that the use of a distance function implicitly assumes that any two observations are neighbour of each other. The stationarity assumption asserts that the joint distribution between any pair of observations depends only on their distance. A weaker form of this stationarity assumption asserts that the covariance between any pair of observations is a function of their distance. This assumption portrays the dependence structure among observations in the population. It therefore forms the basis of prediction.

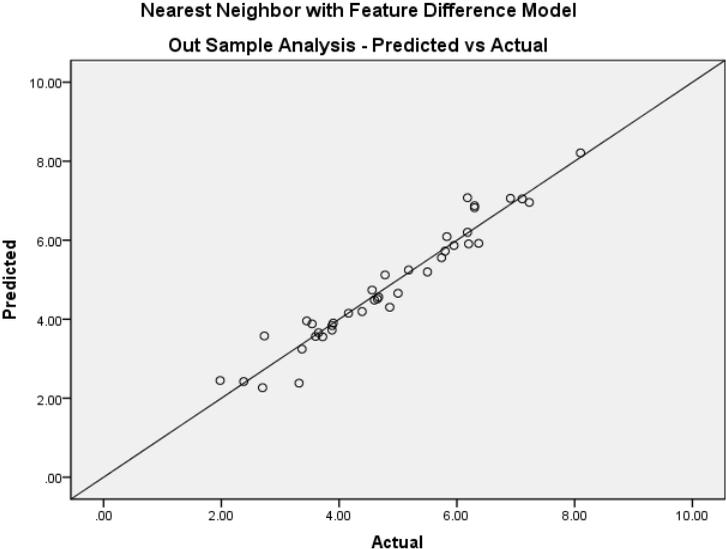
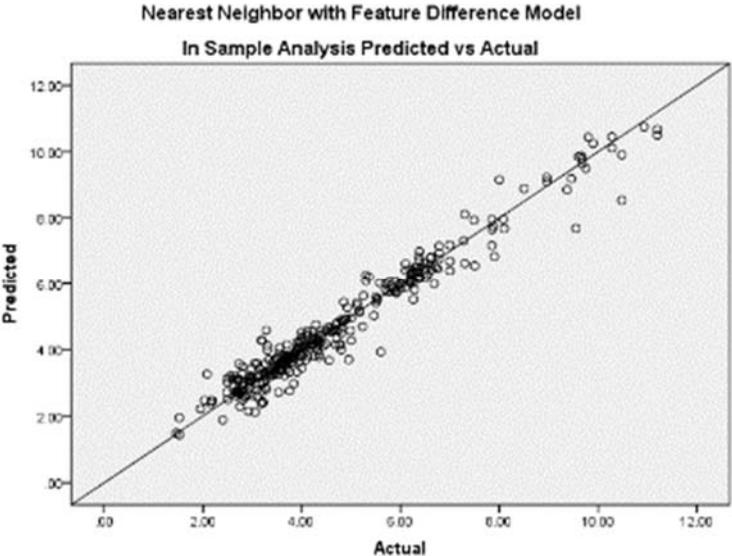
We would like to present three cases in which neighbours are used to assist in predictions. The first exercise is to predict the credit risk (credit risk = 1 indicates high risk and credit risk = 0 otherwise) of individuals based on a set of available information. The data set was downloaded from Kaggle

(<https://www.kaggle.com/uciml/german-credit/home>). Despite the credit risk indicator variable, the other variables are age, credit amount, credit history, duration, employment duration, housing status, instalment rate, job skill, number of credits, gender, savings and purpose. Since most variables are categorical, we assigned scores to each of the categorical variables. For example, we assigned 1000 marks for skilled employee, 100 marks for unskilled, 1000 marks for management/self-employed/highly qualified employee and 0 mark for unemployed/unskilled job variable. The assignment of scores are actually somewhat arbitrary. With the L2 distance function, we determined the distances between every pair of individuals in the database. By ranking these distances, we can find the top n nearest neighbours for each individual. The proportion of the top n nearest neighbours of an individual whose credit risk equals 1 provided an estimate of the chance of the event that the individual also has credit risk of 1. We tried different values of n in order to obtain the best prediction accuracy possible. It turned out that the best prediction accuracy is 83% for $n=1$ or $n=2$. This accuracy drops to 72% as we increase n to $n=10$. The lack of a set of standard rules in defining the distance between any two observations causes uncertainty and the arbitrary assignment of scores to categorical variables may also create uneven in scales among these variables. Since we need to find distances for every pairs of observations, the implementation is time consuming. The computation complexity is N^2 , where the number of observations in the database is N .

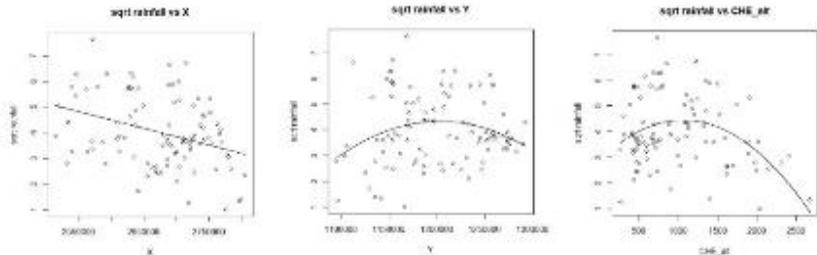
The second exercise is to predict the transaction prices per square foot of real estates in Ma On Shan. We collected 340

transaction data in Ma On Shan between September and November of 2016. Besides transaction prices, we also collected features on each of the properties in the data. The features include size, ceiling height, number of bath rooms, number of bedrooms, size of dining room, proximity to market or shopping mall, proximity to MTR, proximity to swimming facility, proximity to sporting facilities, proximity to park, Sea view, facing direction and so on. Since there is a natural spatial distance between any pair of real estates, we first estimate the log transaction price per square foot of a property by the average log transaction prices per square foot of its neighbours in the data. Then we adjust this estimate by a weighted average of the difference in the features between the real estate and its neighbours. This is in consistent with the usual practice in valuing a property in the valuation department of most banks. By assuming a normally distributed measurement errors, we estimate the model parameters by the method of maximum likelihood. The data was randomly divided into a training set and a testing set. The training set was used to obtain the parameter estimates and the predicted log transaction prices per square foot in the testing set are compared with their actual

log transaction prices per square foot. The following is a diagram of such a comparison.

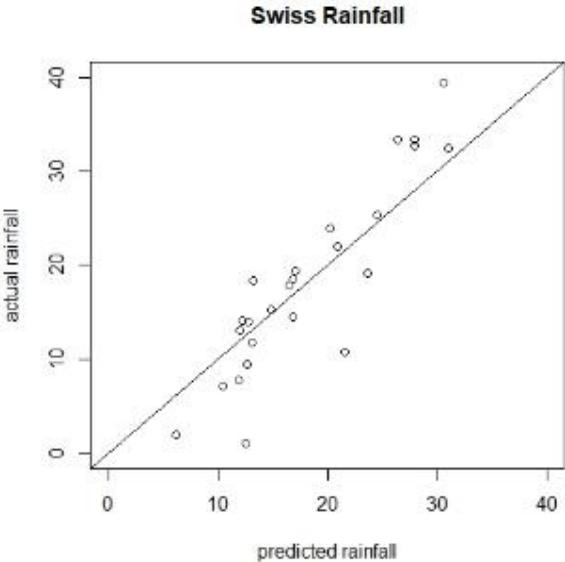


The last exercise relates to the spatial data set **swissRainfall** provided by R package **geostatsp**. The data consists of 100 daily rainfall measurements on the 8th of May 1986 and their corresponding X, Y and Z locations, where a rainfall measurement was taken in the position with X = longitude, Y = latitude and Z = altitude. With X, Y and Z as the explanatory variables and the square root of the rainfall measurements as the dependent variable, we illustrate the relationships between the dependent variable and each of the X, Y and Z with their corresponding scatter plots as given in the following diagrams respectively.



Since the residuals from a regression of the dependent variable on quadratic functions of the explanatory variables are not independent, we adopted a parametric approach to estimate a covariance function of the residuals. The basic idea is that the covariance between any two points in space is inversely related to their distance. There are two main family of parametric covariance functions, namely the Matérn family and the radial basis family. A weighted regression with the Matérn covariance function was employed to model the dependent variable. Based on a normal distribution assumption on the residuals, estimation of the model parameters was done by the method of maximum likelihood. Because of the dependence structure on the residuals, the prediction

of the dependent variable on a location outside where the data was observed should be based on the conditional distribution of the residual given the data. Hence the proper method to do prediction of the rainfall measurement on a new location should be in three steps. The first step draws random numbers from the above mentioned conditional distribution to form random numbers of the dependent variable on the new location. The second step then back transforms the random numbers to obtain the random numbers of the rainfall measurement on the new location. The final step is to use the average of the random numbers to form a prediction on the rainfall measurement on the new location. We divided randomly the data set into a training set and a testing set. Only the training set was involved in the parameter estimation. Then we predicted the rainfall measurements on the testing set and compare them with the actual measurements. The following diagram presents such a comparison.



I hope that this article can give the readers a general introduction on the use of neighbours in improving prediction. After reading this article, I also hope that you will be interested in exploring more on the topics of spatial data and network data analysis and better yet to pursue a higher degree in statistics.