ISBN 978-988-8370-33-7

政府物流服務署印

數學百子櫃系列(二十三) 2016/17 中學生統計創意寫作比賽 作品集

教育局數學教育組

數學百子櫃系列 (二十三)

2016/17 中學生統計創意寫作比賽

Statistics Creative-Writing Competition for Secondary School Students

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（二十三）

2016/17 中學生統計創意寫作比賽

Statistics Creative-Writing Competition for Secondary School Students

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

©2017 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 978-988-8370-33-7

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《2016/17 中學生統計創意寫作比賽作品集》，是這個系列的第二十三冊。本書輯錄的文章，大部分是「2016/17 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概率和統計的知識。

本書共有 12 篇文章，第 1 至 7 篇為「2016/17 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 5 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，內容有趣，期望讀者閱讀後能增加統計知識，並能善用統計決策、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，增強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「 離散概率分佈 」。我們十分感謝香港大學統計及精算學系的鍾玉嘉博士在簡介會中介紹相關的概念，並鼓勵同學在這課題上發揮創意。繼承以往的優良成績，本屆比賽收到約 40 份參賽作品，當中不乏精彩之作。文章取材創新，趣味盎然；同學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

2017 年 10 月 15 日

目錄

編者的話　3

前言　5

目錄　7

冠軍作品: 誰是誰非－透過概率分析民調的可信性.....8

亞軍作品: **Mission DESTROY - The Boss vs. The Targets** 21

季軍作品: 消防演習背後，香港的消防局 56

優異作品: 特首寶座，花落誰家？ 70

優異作品: 「盲」衝直「撞」 80

優異作品: **Usage of Priority Seats in Hong Kong** 94

優異作品: 離散概率有訣竅，遊戲抽獎抽到笑 108

邀請作品：人口普查與人口登記冊 120

邀請作品：如何提高隊伍在「中學生統計創意寫作比賽」當中獲獎的機會？對過往三屆參賽作品的統計分析 128

邀請作品：統計的學與教 137

邀請作品：**Probability and Prime Numbers**.....143

邀請作品：**Bayesian Criminal Tracking** 148

冠軍作品:

誰是誰非－透過概率分析民調的可信性

學校名稱：瑪利諾神父教會學校

學生姓名：張芷晴

指導教師：郭宇輝老師

文章摘要

二零一五年，政改民調出現了支持與反對比例相同的情況。當時立法會議員湯家驊質疑民調的可信性，認為雙方人數一樣的機會為幾百萬分之一，但一名數學老師指出，每一個結果只有約一百五十分之一機會出現。在這篇文章中，我們會透過計算概率，嘗試分析誰對誰錯。

誰是誰非—透過概率分析民調的可信性

引言:

二零一五年，三間大學政改民調首次出現支持與反對比例相同的情況。當時立法會議員湯家驊質疑民調的可信性，認為雙方人數一樣的機會為幾百萬分之一，其後，一名數學老師批評，認為出現的結果只有約一百五十個，而每一個結果只有約一百五十分之一機會出現，其有關爭議無休無止。事實上，統計學能為此找出答案。在一個中午，允行和一心恰巧正在翻看其有關文章.....

2015 年 6 月 10 日　蘋果日報

民調正反比率平手　湯家驊下午質疑晚上道歉

對於最新一輪民調顯示反對及支持政改方案的比率同為 42.8%，公民黨立法會議員湯家驊下午回應時對民調結果表示驚訝及難以置信,指 正反比率同為 42.8%會否太過巧合，「係咪正反人數真係完全一樣呢？」其言論備受批評，湯家驊晚上於 Facebook 就此致歉，「對不起！我理解錯了！」

湯家驊今午指,三大滾動民調最新數據出現支持與反對政改的比率平手,質疑只有數百萬分之一,是否太巧合。有數學老師即向他致公開信，推算機會只是大約一百五十分之一，絕非罕有現象，批評其言論違反科學、甚至基本常識。

湯家驊晚上在 Facebook 回應，稱數學從來不是他的強項。「如你所說的話，我的回應確有誇大之詞！對不起！我理解錯了！」他強調， 今天的回應重點並非在此，而是看不到民調對解決政改有甚麼幫助，並希望三大民調機構接受其道歉！

湯家驊下午回應民調顯示「平手」時，反問「何喜之有」，指根本無助消解當前的政改問題，只會令雙方立場更加「強硬」。他覺得現時需要的可能是不進行民調，讓雙方靜思一下「面對呢個困局應該點去解決，咁仲可能有意義少少囉」。

一心: 允行，依你看，誰的看法較準確？

允行: 要判斷對錯，首先要找出理據支持。民調，在數學上是抽樣調查的一種，是通過從總體中隨機抽取一個樣本，通過樣本的特質來推斷或估計總體的特質。

一心: 在這次調查上，總體即為全港選民，而樣本是受訪的人數約 1000 人。

允行: 民調可信性在於其隨機抽樣，換言之，市民被抽中作為樣本的機會均等，你看以下是由港大、中大和理大合辦的 2017 年行政長官選舉辦法滾動民調數表結果其中一小部分：

調查日期	樣本人數	支持	反對	沒意見
31/5-4/6/2015	1,119	45.8%	39.6%	14.7%
1-5/6/2015	1,121	43.7%	41.6%	14.7%
2-6/6/2015	1,118	42.8%	42.8%	14.5%
3-7/6/2016	1,115	41.7%	43.0%	15.3%

一心: 上表顯示在 6 月 2 日至 6 月 6 日期間，正反雙方的百分比完全一致。這種情況是不是非常巧合呢?會不會令人覺得有造假的嫌疑呢?

允行:到底湯家驊和該名數學老師的意見是否正確?我們可以先考慮一個簡單例子。數學老師認為:「每個結果只有約一百五十分之一機會出現。」針對此點，現舉出一例子，假設四名受訪者回答民調問題，四人皆只能回應贊成和反對，所出現的結果可以下表表示:(✓代表贊成，而×代表反對)

	受訪者 A	受訪者 B	受訪者 C	受訪者 D
1	✓	✓	✓	✓
2	×	✓	✓	✓
3	✓	×	✓	✓
4	✓	✓	×	✓
5	✓	✓	✓	×
6	×	×	✓	✓
7	×	✓	×	✓
8	×	✓	✓	×
9	✓	×	✓	×
10	✓	×	×	✓
11	✓	✓	×	×
12	✓	×	×	×
13	×	✓	×	×
14	×	×	✓	×

	受訪者 A	受訪者 B	受訪者 C	受訪者 D
15	×	×	×	✓
16	×	×	×	×

允行: 假設每人贊成和反對的機會均等。

每組組合出現概率 $=\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}=\frac{1}{16}$

所以，我們可以得出

P(4 人贊成) $=\frac{1}{16}$

P(3 人贊成) $=\frac{4}{16}=\frac{1}{4}$

P(2 人贊成) $=\frac{6}{16}=\frac{3}{8}$

P(1 人贊成) $=\frac{4}{16}=\frac{1}{4}$

P(0 人贊成) $=\frac{1}{16}$

在上例的 5 個結果中，每個結果出現概率不一定等於 $\frac{1}{5}$，而且正反雙方人數一樣的概率為 $\frac{3}{8}$。類似地，若有 N 人受訪，而且只有贊成及反對，在 $(N+1)$ 個結果中，每個結果出現概率不一定等於 $\frac{1}{N+1}$，而且正反雙方人數一樣的概率亦非 $\frac{1}{N+1}$。

因此可知，上文中那位數學老師的意見認為「每個結果只有約一百五十分之一機會出現。」並不一定正確。

一心: 如果要計算1100人中每個結果出現的機會，豈不是要畫1000多個圖表？

允行: 那可說不定，有更簡單的方法可以推算出結果，你聽過二項分佈 (Binomial distribution)嗎？上述文章中所提及的政改民調大約有9百多人有回覆贊成或對。現在我們假設只考慮贊成和反對的人數，總人數為900，而贊成及反對的概率均等，均為 $p = 0.5$，則根據二項分佈的公式可得

P (正反雙方人數一樣)

= P (450人贊成和450人反對)

$= C_{450}^{900}\, p^{450}\left(1-p\right)^{900-450}$

$= C_{450}^{900}\, p^{450}\left(1-p\right)^{900-450}$

= 0.0265887768

$\approx \dfrac{1}{38}$

一心: $\dfrac{1}{38}$ 的機會遠大於湯家驊議員所說的幾百萬分之一，以及數學老師所指出的 $\dfrac{1}{150}$。那麼，當 p 的值有所改變時，其

概率又會出現什麼變化呢?現在我會舉出其他例子。根據二項分佈的公式，

P(正反雙方人數一樣)

= P(450人贊成和450人反對)

$$= C_{450}^{900} p^{450} (1-p)^{450}$$

現在我可以透過電腦試算表程式幫忙下，代入不同的 p 值去計算正反雙方人數一樣的概率。以下就是我得到的結果：

p 的值	概率Probability(M)	$N=\frac{1}{n}$
$p=0.5$	0.026589	$\frac{1}{38}$
$p=0.527702$	0.0066666	$\frac{1}{150}$
$p=0.574813$	9.99×10^{-7}	$\frac{1}{1,008,816}$
$p=0.578695$	3.33×10^{-7}	$\frac{1}{3,002,281}$

一心: 所以當$p=0.527702$, 正反雙方人數一樣的概率大約為一百五十分之一(數學老師所指出的數目)，而當$p=0.578695$時，正反雙方人數一樣的概率大約為三百萬分之一 (湯家驊

議員所提出的「幾百萬分之一」)

允行: 原來湯家驊議員和那位數學老師所提出的數值並不是完全錯誤，只是在特定的情況下才是正確的。

一心: 等一等!我們以上的計算是假設受訪者只可表示贊成或反對，但實際的情況是受訪者可選「無意見」。

允行: 那麼讓我們試計算在不排除沒有意見的情況下的結果。首先，我們假設贊成的概率為 p，反對的概率為 q，無意見的概率為 r (其中 $r = 1 - p - q,$ 且 $r \neq 0$)。假設受訪的總人數為 N，贊成和反對的數目均為 k，餘下$N-2k$ 人選擇「無意見」。那麼我們可以再進一步，使用多項分佈(Multinomial distribution)的公式：

P (贊成＝反對＝k 及無意見$= N - 2k$)

$$= P_k = \frac{N!}{k!k!(N-2k)!} p^k q^k r^{N-2k}$$

要計算正反雙方人數一樣的概率，就要將以上一系列的概率加起來。換句話說，

P(正反雙方人數一樣)

$$= \sum_{k=0}^{\left[\frac{N}{2}\right]} P_k = \sum_{k=0}^{\left[\frac{N}{2}\right]} \frac{N!}{k!k!(N-2k)!} p^k q^k r^{N-2k}$$

一心: 這個公式很複雜。我們試試再運用電腦試算表程式去計算正反雙方人數一樣的概率。下表列出了在總人數為1100時一些不同的 p、q 及 r 值下正反雙方人數一樣的概率：

同意比率(p)	反對比率(q)	無意見比率(r)	P(正反雙方人數一樣)(M)	$M=\frac{1}{n}$
0.49	0.49	0.02	0.01214795	$\frac{1}{82}$
0.4	0.4	0.2	0.013445667	$\frac{1}{74}$
0.333	0.333	0.334	0.014736782	$\frac{1}{68}$
0.25	0.25	0.5	0.017009023	$\frac{1}{59}$
0.1	0.1	0.8	0.026902813	$\frac{1}{37}$
0.8	0.1	0.1	5.1455×10^{-197}	$\frac{1}{1.94\times10^{196}}$
0.7	0.1	0.2	1.772×10^{-153}	$\frac{1}{5.64\times10^{152}}$
0.6	0.1	0.3	3.2506×10^{-115}	$\frac{1}{3.08\times10^{114}}$
0.5	0.1	0.4	1.0259×10^{-81}	$\frac{1}{9.75\times10^{80}}$
0.4	0.1	0.5	8.37584×10^{-53}	$\frac{1}{1.19\times10^{52}}$

同意比率(p)	反對比率(q)	無意見比率(r)	P(正反雙方人數一樣)(M)	$M=\frac{1}{n}$
0.4473	0.415	0.1377	0.006666661	$\frac{1}{150}$
0.3713	0.5	0.1287	3.3333×10^{-7}	$\frac{1}{3,000,027}$

一心: 從上表我們可看到若正反雙方的比率相近，樣本中正反雙方人數一樣的概率大約為幾十分之一。但是正反雙方的比率有一些距離時，樣本中正反雙方人數一樣的概率可以低於至10的幾百次方分之一，即是幾乎是0。

另一方面，當 $p=0.4473$ 而 $q=0.415$ 時, 正反雙方人數一樣的概率大約接近一百五十分之一（數學老師所指出的數目），當$p=0.3713$ 而 $q=0.5$ 時，正反雙方人數一樣的概率大約接近三百萬分之一 (湯家驊議員所提出的「幾百萬分之一」)，因此湯家驊議員和那位數學老師所提出的數值在某些特殊情況下是成立的。

一心: 本來想參考你的意見判定是非，現在發現兩人的看法皆非盡善盡美，亦非完全錯誤，這次的結果真有趣，哈哈！

允行: 在未搞清事情的根源前，自作聰明，妄下定論，是一種自打嘴巴的行為。在生活上，須講究真憑實據。我們不能在

沒有依據下作結論，否則可能會冤枉他人；亦不能盡信權威在媒體的意見，否則會失去獨立思考的空間。

(字數：2135字)

參考資料

1. “民調正反比率平手湯家驊下午質疑晚上道歉”，蘋果日報，2015年6月10日

http://hk.apple.nextmedia.com/realtime/news/20150610/53837274

亞軍作品:

Mission DESTROY - The Boss vs. The Targets

School Name: St. Francis Xavier's College

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

CHAN Pui -wing

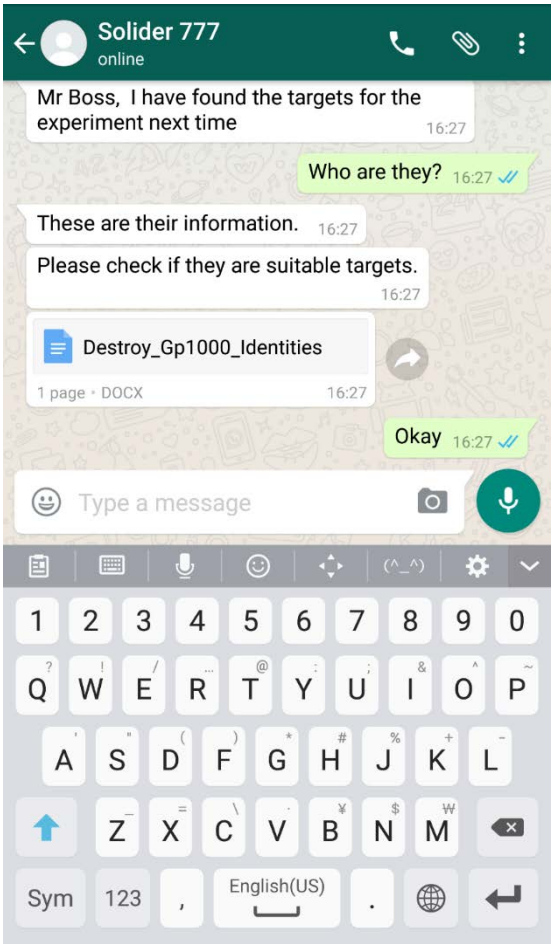
Supervising Teacher: Mr. Chan Pak Kei

Abstract:

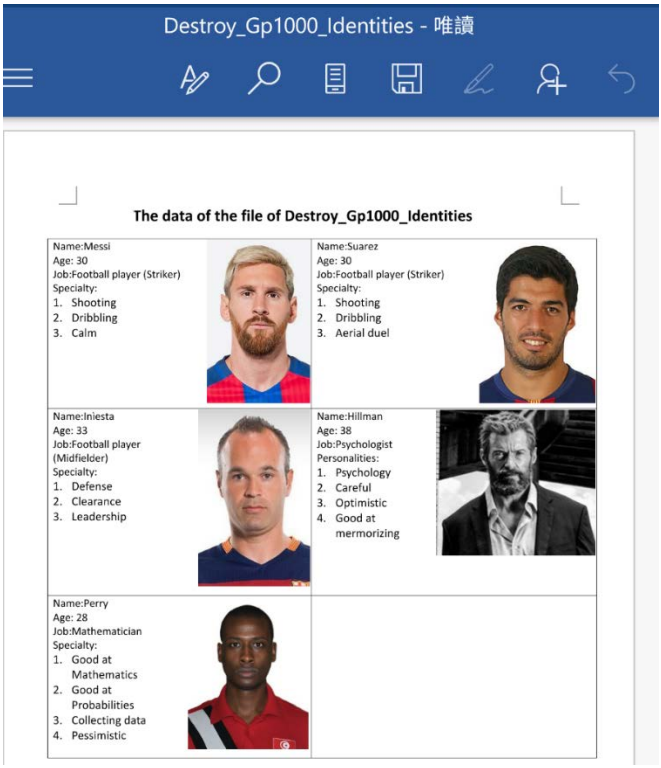
Messi, Iniesta, Suarez and other two people were caught to join a deadly real-life experiment. If they lost a task, the world would suffer damage by bomb attacks. If finally lost all the tasks, the human beings would definitely die. Therefore, they stepped forward to rescue the world. Although failing in the first task, they passed the last task that could eventually save the world. Moreover, they found out the one who was manipulating the whole 'game' from behind the scenes.

Part One

One day, a message was sent to a man.



Then, the man clicked into the file and saw the five targets.


Destroy_Gp1000_Identities - 唯讀
The data of the file of Destroy_Gp1000_Identities
Name:Messi
Age: 30
Job:Football player (Striker)
Specialty:
1. Shooting
2. Dribbling
3. Calm
Name:Suarez
Age: 30
Job:Football player (Striker)
Specialty:
1. Shooting
2. Dribbling
3. Aerial duel
Name:Iniesta
Age: 33
Job:Football player (Midfielder)
Specialty:
1. Defense
2. Clearance
3. Leadership
Name:Hillman
Age: 38
Job:Psychologist
Personalities:
1. Psychology
2. Careful
3. Optimistic
4. Good at mermorizing
Name:Perry
Age: 28
Job:Mathematician
Specialty:
1. Good at Mathematics
2. Good at Probabilities
3. Collecting data
4. Pessimistic

Solider 777
1 page · DOCX 16:27
Okay 16:27
Yes. They are good enough 16:30
But is the footballer's bodyguard easy to deal with? 16:30
plz send me the statistic report about their bodyguards 16:31
Expected time to faint an op...
1 page · DOCX 16:31
Type a message

After the second file was sent to him, he opened it to check the statistics of the guard of the targets.

Expected value to faint an opponent - 唯讀

The statistics of stunning certain number of opponents in 1 minute

Messi's security

Number of opponents	50	60	70	80
Probability of stunning certain number of opponent in 1 minute	$\frac{11}{16}$	$\frac{1}{8}$	$\frac{5}{32}$	$\frac{1}{32}$

Expected number of opponent can be stunned in 1 minute

$$= 50 * \frac{11}{16} + 60 * \frac{1}{8} + 70 * \frac{5}{32} + 80 * \frac{1}{32}$$

$$= 55.3125 \text{ opponents}$$

$$\approx 55 \text{ opponents}$$

Expected value is based on the probability calculated to predict the result in certain number of trials.

Iniesta's Security

Number of opponents	50	60	70	80
Probability of stunning certain number of opponent in 1 minute	$\frac{13}{16}$	$\frac{1}{16}$	$\frac{1}{12}$	$\frac{1}{24}$

Expected number of opponent can be stunned in 1 minute

$$= 50 * \frac{13}{16} + 60 * \frac{1}{16} + 70 * \frac{1}{12} + 80 * \frac{1}{24}$$

$= 54.8884$ opponents

≈ 54 opponents

Suarez's Security

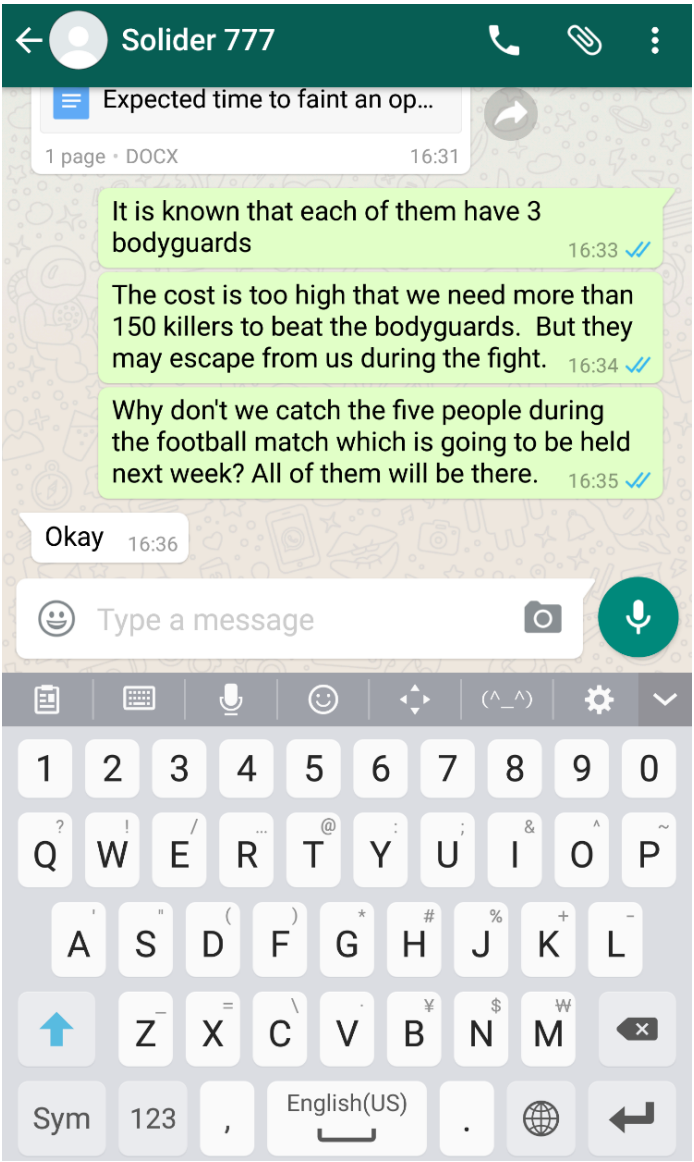
Number of opponents	50	60	70	80
Probability of stunning certain number of opponent in 1 minute	$\frac{23}{32}$	$\frac{1}{7}$	$\frac{31}{448}$	$\frac{31}{448}$

Expected number of opponent can be stunned in 1 minute

$$= 50 * \frac{23}{32} + 60 * \frac{1}{7} + 70 * \frac{31}{448} + 80 * \frac{31}{448}$$

$= 54.8839$ opponents

≈ 55 opponents



Soon after the message was sent, the targets were captured.

Part Two

The targets were kept in a dark room with no exit. Suddenly, light was turned on and the targets were wakened up.

‘Where am I?’ Messi said.

‘Who are they? Hey, are you guys okay?’ Suarez asked.

‘I’m fine,’ answered Hillman.

Suddenly, they heard a voice from the speaker on the top of the room.

‘Welcome, everyone. This is a deadly real-life experiment. You are now given three tasks to deal with. You are the last group to participate in this experiment. Let’s look at the screen.

Some data was shown on the screen held on the wall.

Bernoulli distribution is to the probability to get a success

Total number of groups tested	No. of groups that passed all the tasks
999	5

By Bernoulli distribution

$$P(\text{A group win}) = \frac{5}{999}$$

$$= 0.005005$$

$$= 0.5005\%$$

$$P(E) = \frac{\text{No of favourable outcome}}{\text{total possible outcome}}$$

"That means you have only 0.5005% to pass all the tasks."

Everyone was shocked!

Iniesta broke the silence and asked, 'Err…What is Bernoulli distribution?'

Perry replied, 'Bernoulli distribution is a method to find the probability of an event to get a success such as the probability of tossing a coin.'

'How do you know?'

'I am good at Mathematics. Actually, 0.5005% is extremely low in probability. So…'

‘Wait! As I remember, we were having a derby with Real Madrid,’ disturbed Suarez.

‘I remember I was watching that competition as well. It is impossible for me to be here,’ screamed Hillman.

‘Oh yes! I remembered the result is 1 to 1 in the 55th minute,’ added Perry.’

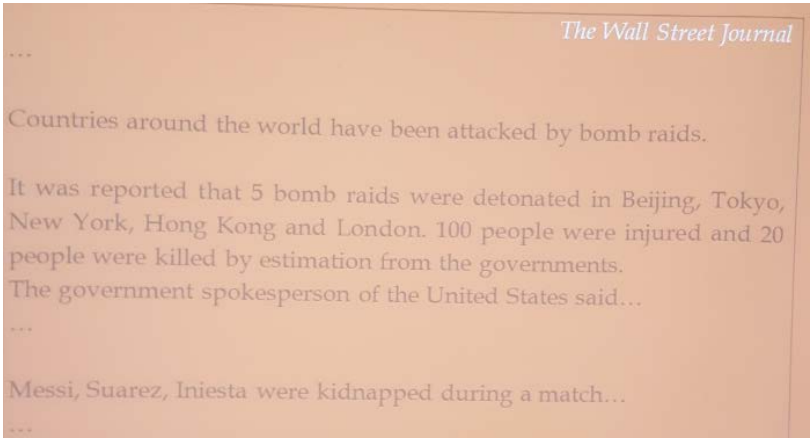
‘As you doubted the “game”, I’ve just exploded 5 bombs worldwide. The 5 bombs are planned to be exploded originally if you failed the first task, but it seems that I should show you the reality first.’

‘Now, you will have 30 minutes to consider whether you join or not. Just to remind you, think carefully before you make your decision.’

‘One more thing is, there will be 24 bombs exploding every hour so take your time and finish the “game”.’

‘Impossible! That’s crazy! What the heck!’ said Suarez.

Suddenly, a message was shown on the screen.



The Wall Street Journal

…

Countries around the world have been attacked by bomb raids.

It was reported that 5 bomb raids were detonated in Beijing, Tokyo, New York, Hong Kong and London. 100 people were injured and 20 people were killed by estimation from the governments.
The government spokesperson of the United States said…
…

Messi, Suarez, Iniesta were kidnapped during a match…
…

The room was immediately in a mess.

‘Oh my god, this looks real. What should we do?’ asked Perry anxiously.

‘We have no choice,’ replied Hillman, ‘Let’s do it, bro. OH! I forgot to introduce myself. I’m Hillman, a psychologist working in the urban hospital of Barcelona.’

‘I’m Perry. I’m sorry that I have autism, so I am not good at talking with others.’

Then, they introduced themselves one by one.

‘Emm…According to a survey, when 100 people in an unknown environment, 97 of them will be in a mess just like us.’ explained Perry

$$P(\text{people to be afraid})$$

$$= \frac{97}{100} * 100\%$$

$$= 97\%$$

Then, we try to find a person that is not afraid in an unknown environment within 5 people.

p is probability of getting one success

x is total number of trials

$(1-p)^{x-1}$ is the probability that getting fail.

In x trials, find the probability that getting a success in the xth trial.

By Geometric distribution,

$$X \sim Geo(p)$$

$$P(X = x) = (1-p)^{x-1}p$$

X is a random variable

x is the value of X

$$X \sim Geo(0.03)$$

$$P(\text{the 5th person that are not afraid in 5 people})$$

$$= \left(\frac{97}{100}\right)^4 * \left(\frac{3}{100}\right)^1$$

$$= 0.02576$$

That means nearly everyone is in a mess. We can't blame ourselves that not to believe him.'

Everyone was in a silence.

‘Moreover, there will be 24 bombs exploding every hour, which means 2 bombs will explode every 5 minutes in average. In order to find the probability of 2 bombs exploding in every 5 minutes, we can use Poisson distribution to find it.

Base on the mean, find out the probability of something happen in a specific period.

By Poisson distribution,

$$X \sim Po(\lambda)$$

$$P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

X is a random variable
x is the value of X

λ is mean of an event happened
x is the number of occurrences in an event
x! means $1 * 2 * 3 * \ldots * (x-1) * x$

$$P(X=0) = \frac{e^{-(2)}2^0}{0\ !} = 0.135$$

$$P(X=1) = \frac{e^{-(2)}2^1}{1\ !} = 0.271$$

$$P(X=2) = \frac{e^{-(2)}2^2}{2!} = 0.271$$

‘It is expected that 27.1% having 1 bomb explode and 27.1% having 2 bombs explode,’ said Perry.

Everyone was in silence.

‘According to the news, 100 were injured and 20 were killed in 5 bomb explosions which means 20 were injured and 4 were killed in a bomb explosion in average.’

Expected number of people died every 5 minutes

$$= 0 \times P(X=0) + \frac{20}{5} \times P(X=1) + \frac{20}{5} \times 2 \times P(X=2) + \frac{20}{5} \times 3 \times P(X=3) + \ldots \to \infty$$

≈ 8

Expected number of people injured every 5 minutes

$$= 0 \times P(X=0) + \frac{100}{5} \times P(X=1) + \frac{100}{5} \times 2 \times P(X=2) + \frac{100}{5} \times 3 \times P(X=3) + \ldots \to \infty$$

≈ 40

‘WHAT!!! About 480 people will be injured and 96 people die every hour!’ shouted Iniesta frustratingly.

‘That’s why we can’t fail in the following tasks,’ comforted Messi.

‘If you have decided to join, just walk into that door.’ announced by the speaker.

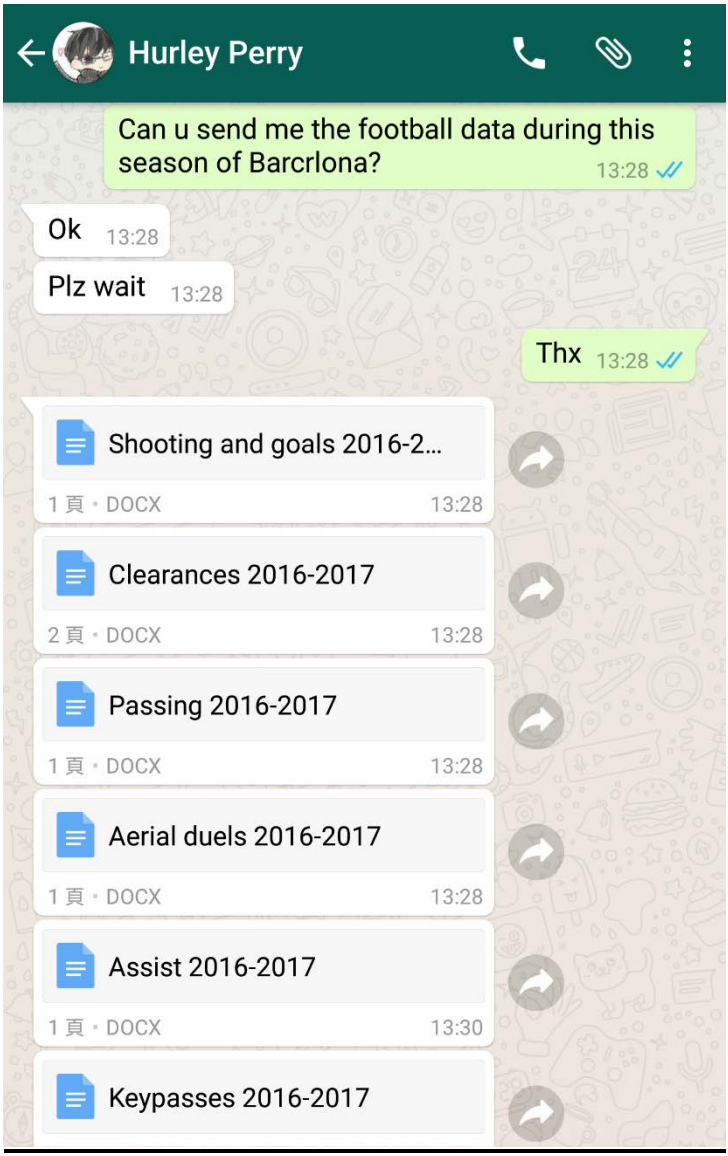
Part Three

‘The first task is participating in a 3 vs 3 football competition. The only way to win is to score. If you fail, 300 bombs will be exploded. If you fail the next tasks as well, a fatal virus will then be spread around the world. You have 30 minutes to discuss for the tactics,’ the speaker broadcasted.

‘Well, that should be us, right?’ asked Iniesta.

‘Yes. Let me see where you should play,’ Perry said while he was turning on his phone, ‘based on the data from my friend.’

Then, he showed the WhatsApp record to the others.



In order to facilitate the calculation of probability, he jotted down data on the ground.

> What is marking?
> A person follows another person in order not to give him a chance to receive a ball, pass or dribble.

He explained, ‘Normally, a guy marks an opponent, so aerial duel in the competition is usually one on one.’

	Percentage of an aerial duel win	Percentage of an aerial duel lose
Messi	10%	90%
Iniesta	20%	80%
Suarez	30%	70%

Perry assumed them to play the game ten times, and calculated their probabilities of winning exactly 5 aerial duels.

B means Binomial distribution
n means total number of trials
p means the probability to clear
a ball successfully

He said, 'By Binomial distribution,

$$X \sim B(n, p)$$

$$P(X = x) = C_x^n p^x (1-p)^{n-x}$$

X is a random
variable
x is the value of X

n is the total number of trials
x is number of time that succeed
p is the probability to get one success
(1-p) is the probability to get one fail
C is the total combination

$$X \sim B(10, 0.1)$$

P(Messi wins 5 aerial duels)

What is clearance?
It means kicking away
the ball from his own
half when defending

$$= C_5^{10} 0.1^5 (1-0.1)^5$$

$$= 1.488 * 10^{-3}$$

$$= 0.1488\%$$

$$X \sim B(10, 0.2)$$

P(Iniesta wins 5 aerial duels)

$$= C_5^{10} 0.2^5 (1-0.2)^5$$

$$= 0.02642$$

$$= 2.642\%$$

$$X \sim B(10, 0.3)$$

$$P(\text{Suarez wins 5 aerial duels})$$

$$= C_5^{10} 0.3^5 (1 - 0.3)^5$$

$$= 0.1029$$

$$= 10.29\%$$

$\because$ P(Suarez wins 5 aerial duels) $>$ P(Iniesta wins 5 aerial duels) $>$ P(Messi wins 5 aerial duels)

$\therefore$ Suarez has great aerial duel skill.

'Not only strikers need to have great aerial duel skill, a defender also requires it. Iniesta, who is the average one, should be our defender. Also, his clearance ability seems good,' said Perry.

Then, he wrote the data of clearance on the ground.

	Percentage of clearing a ball successfully	Percentage of fail to clear a ball
Messi	10%	90%
Iniesta	90.4%	9.6%
Suarez	48.6%	51.4%

Perry assumed them to play the game ten times again, and calculated their probabilities of exactly 8 success clearance.

B means Binomial distribution
n means total number of trials
p means the probability to clear a ball successfully

By Binomial distribution,

$$X \sim B(n, p)$$

$$P(X = x) = C_x^n p^x (1-p)^{n-x}$$

X is a random variable
x is the value of X

n is the total number of trials
x is number of time that succeed
p is the probability to get one success
(1-p) is the probability to get one fail
C is the total combination

P(Messi have 8 success clearance)

$$X \sim B(10, 0.1)$$

$$= C_8^{10}(0.1)^8(1-0.1)^2$$

$$= 3.645 * 10^{-7}$$

$$= 3.645 * 10^{-5}\%$$

What is clearance?
It means kicking away the ball from his own half when defending

P(Suarez have 8 success clearance)

$$X \sim B(10, 0.486)$$

$$= C_8^{10}(0.486)^8(1-0.486)^2$$

$$= 0.0370$$

$$= 3.70\%$$

$\because$ P(Iniesta have 8 success clearance) > P(Suarez have 8 success clearance) > P(Messi have 8 success clearance)

$\therefore$ Iniesta is better in clearance.

‘Having average skill in aerial duel and extraordinary clearance skill, he is the best choice to be our defense.’

‘Yes, this is just what our plan is! You idiot!’ grumped Suarez. Perry ignored Suarez and continued to say,

‘One more thing is, not only clearance and aerial duel, we should also check the probability of successful passing. This is the most important data of a football player.’

	Succeed to passing a ball	Fail to pass a ball
Messi	0.802	0.198
Suarez	0.731	0.269
Iniesta	0.873	0.127

$$\text{Mean of a success passing}$$

$$= \frac{\text{Messi} + \text{Suarez} + \text{Iniesta}}{3}$$

$$= \frac{0.802 + 0.731 + 0.873}{3}$$

$$= 0.802$$

‘Well, I’ve never thought I have such a high success passing rate!’ exclaimed Suarez. However,’ Perry continued.

B means Binomial distribution
n means total number of trials
p means the probability to clear a ball successfully

n is the total number of trials
x is number of time that succeed
p is the probability to get one success
(1-p) is the probability to get one fail
C is the total combination

'By Binomial distribution,

$$X \sim B(n, p)$$

$$P(X = x) = C_x^n p^x (1-p)^{n-x}$$

X is a random variable
x is the value of X

$$X \sim B(10, 0.802)$$

P(success passing without being tackles in 10 times)

$$= C_0^{10}(0.802)^{10}(1-0.802)^0$$

$$=0.1101$$

$$=11.01\%$$

$$X \sim B(20, 0.802)$$

P(success passing without being tackles in 20 times)

$$= C_0^{20}(0.802)^{20}(1-0.802)^0$$

$$= 0.01212$$

$$= 1.212\%$$

‘What does it indicate? Is that still high?’ asked Suarez.

‘It indicates that probability of being tackled is high when you pass. The success rate is less than 12%...’ replied Perry.

‘So, no matter how strong and careful we are, we will still lose in passing, ’Iniesta worried.

‘ARR…Enough!’ shouted Hillman, ‘You have brought down our morale. Boost our morale instead of reducing our own courage.’

‘Yes. We are strong. In Barcelona, we are seldom tackled. We can do it,’ Messi encouraged.

‘Okay! We can pass it but the first thing we should stay calm,’ said Suarez.

‘Remember, if you lose, a lot of people will die because of you,’ added Perry before the competition.

‘Time’s up! You only have one chance. The one who first scores wins. And your opponents are these people-like robots.’ Three robots appeared from the floor.

‘Kick off!’

Messi passed the ball to Iniesta. After passing the ball to Suarez, he stayed at the back of the field. Suarez did a one-touch pass to Messi. When Messi ran faster with the ball, he whispered, ‘Can we win in this competition?’

At that time, Suarez gave a signal to Messi to pass the ball. When he wanted to receive it, he was intercepted by a robot.
‘The robot has taken the ball!’ shouted Perry, ‘DEFEND! You have millions of lives in your hands!’

The robot ran faster than all of them. Even Iniesta could defend it, he is pushed away and got hurt.

At last, no matter how they chased, the robots were the one who first scores. They failed the task. All of them were frustrated, as they had never met this kind of opponent. Moreover, they knew they had killed millions of people due to their lost. ‘Are you okay,

Iniesta? You seemed to have a serious injury,’ asked Hillman.

Iniesta replied, ‘No problem. I can handle it.’

‘Take care. If you can't participate, just rest.’

‘Hey guys, I know you are worried, but don't be pessimistic. We still have one more chance. If we passed the task, we can save the rest of the lives!’ said Hillman, ‘Let's cheer up!’

‘That means if we lose, we make human this kind of species distinct,’ intercepted Perry.

‘Shut the heck up! Perry! You are discouraging them!’ screamed Hillman angrily.

Part Four

‘300 bombs have been exploded. Now, this is the last task. Let us see if you can save the humankind. Penalties. You are going to have three chances. One person, one shot. If you get one goal, then you win. However, the goalkeeper knows you guys well. Therefore, good luck. You have 10 minutes to prepare.’

‘Ha-ha idiots!!! That’s should be easy. Just gets a goal having 3 chances,’ said Suarez happily.

‘I don’t think so,’ replied Perry.

Perry immediately wrote down the data of their success scoring in the 2016-2017 season.

	Messi	Iniesta	Suarez
Game played	30	30	30
Average shots per game	5	0.9	3.3
Total number of goal in 2016-2017	23	1	23

$$\text{Number of shooting} = \text{Average shots per game} * \text{Game played}$$

$$\text{Number of shooting of Messi}$$
$$= 5 * 30$$
$$= 150$$

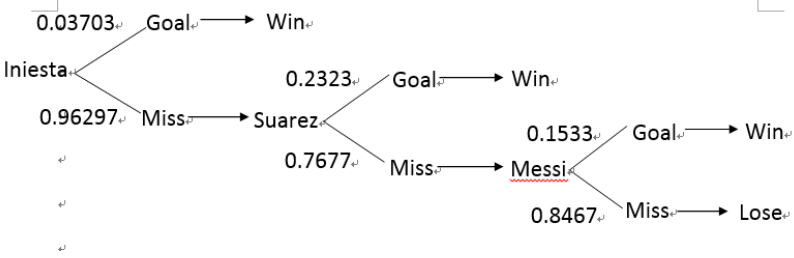
$$\text{Percentage of success goal}$$
$$= \frac{23}{150} * 100\%$$
$$= 15.33\%$$

$$\text{Number of shooting of Iniesta}$$
$$= 0.9 * 30$$
$$= 27$$

$$\text{Percentage of success goal}$$
$$= \frac{1}{27} * 100\%$$
$$= 3.703\%$$

$$\text{Number of shooting of Suarez}$$
$$= 3.3 * 30$$
$$= 99$$

$$\text{Percentage of success goal}$$
$$= \frac{23}{99} * 100\%$$
$$= 23.23\%$$



Tree diagram can let people to identify all the cases included.
< means there are two possible results.
The number represents the probability.

Perry said, "According to this tree diagram,

$$P(\textbf{Iniesta goal and wins the game}) = 0.03703$$

$$P(\textbf{Suarez goal and wins the game})$$

$$= P(\textbf{Iniesta not goal})P(\textbf{Suarez goal})$$

$$= 0.96297 \times 0.2323$$

$$= 0.2237$$

Why is P(Iniesta not goal) timed?
If Iniesta doesn't miss, Suarez doesn't have to shoot and get a goal. Thus, we need to calculate it.

$$P(\textbf{Messi goal and wins the game})$$
$$= P(\textbf{Iniesta not goal})P(\textbf{Suarez not goal})P(\textbf{Messi goal})$$
$$= 0.96297 \times 0.7677 \times 0.1533$$
$$= 0.1133$$

In terms of overall shooting, that may not be as high as imagined."

Assume Iniesta, Suarez and Messi performance are independent,

P(Win) = 0.03703 + 0.2237 + 0.1133

= 0.374

'Only 37.4 %,' said Perry, 'This is a very low probability to win in this task…. Remember, you have all the lives of the world. You can't lose!'

'Game starts!'

A goalkeeper appeared in front of a goal. A robot that was as strong as the best goalkeeper Neuer appeared in front of the goal. It seemed to be a very good goalkeeper.

'Send one of your teammates.' Then Iniesta immediately came out and prepared for the shot.

A whistle was sounded. Iniesta tries very hard to shoot the ball. However, as he is injured, he is not able to shoot the ball well. He

fell down but touched the ball.

‘ARRRRRRRRRR…My leg!!’

‘Bang!’ Neuer hold the ball tightly.

Perry scolded, ‘If you had known you are hurt, then you should not have participated in this shoot! You waste this chance!’

‘That's enough. As our teammate, he has tried his best. If you don't know how to play football, then shut up!’ shouted Suarez.

‘It's okay. We are a team. Internal strife should not appear in our team. We will lose our temper, patience and trust to overcome these tasks. Listen to me. No more arguing,’ told Messi.

Then Suarez walked towards the goal and ready for another shoot. He tried not to think about what Messi had told him. However, when he looked at Perry, he started to

think of the things he did to Iniesta. Due to his deconcentrating, he missed the target.

‘Suarez, never mind. We can still win if I get a goal!’ said Messi.

‘Oh my God. We are going to die. The fatal pathogen will be spread around the world,’ said Perry frustratingly while Messi was walking to the football field

‘Don't worry. I can do it. No matter how accurate and how low the probability is, I will always perform a miracle and save the world, ‘said Messi like a hero.

Messi wanted to save the world. He shot the ball without hesitation and planning. The ball directly went towards right. The robot jumped to that direction. When it nearly touched the ball, the ball turns left and flew directly to the net.

‘Goal! Goal! Goal…’ When Messi shot into the goal, everyone is thrilled as they saved the world. However, Perry was upset and cried.

‘Why are you so upset? We have saved the world!’ Iniesta asked Perry.

‘Nothing special,’ murmured Perry.

Although Perry did not clearly answer them the reason why he was upset, Hillman found the answer.

Part Five

‘We have just ruined his plan. Perry, you are the one who manipulates behind the scenes!’ shouted Hillman.

‘Are you crazy?’ cried Perry.

Hillman said calmly, ‘You applied the self-fulfilling prophecy which affects our mind. While we were participating in the “game”, you added lots of pressure on us. I remembered you said something like this. “You should know that if you lose, a lot of people will die for you,” “That means if we lose, we will make mankind distinct.” “Remember that you guys controls the lives of the all human beings.” These sentences pressurized us, which influenced our thinking secretly.

Also, the reason why you blamed Iniesta is not only failure of shooting, but to affect the emotion of Suarez. Then he was distracted while shooting. By affecting Suarez, you can ensure the one with the highest chance to score fails.’

‘No, I’m not!’

‘The major evidence is that you said the result of the competition is 1 to 1 in the 55th minute.’

‘Yes, but what is the problem?’

‘The problem is in the 55th minute there was no goal for both teams,’ said Suarez.

Perry finally admitted that he was the ‘game setter’

‘But why do you do that?’ asked Iniesta.

‘In this world, disparity between the rich and the poor becomes larger and larger. Reports from different organisations show that the gap between rich and poor is far greater than that was expected.

The richest 8 in the world

The richest eight in the world own the half of the wealth of the world, which is seemingly unfair to the poor around the world.

Moreover, there are a lot of refugees waiting to enter Europe to escape wars. However, most of them were sent back to their origins,’ Perry said. Then he drew a table.

	Refugees	Sent back	Percentage
2013	431000	168090	39%
2014	626715	192445	30.71%

	Afghanistan	Sent back	Percentage
2015	124000	62000	50%

‘As you can see 30% to 40% of refugees were sent back.
By Bernoulli distribution,

$$P\left(A \text{ refugee sent back in } 2013\right) = \left(0.39\right)^1 \left(0.61\right)^0 = 0.39$$

$$P\left(A \text{ refugee sent back in } 2014\right) = \left(0.3071\right)^1 \left(0.6929\right)^0 = 0.3071$$

Although it slightly decreased, number of Afghanistan sent back was about 50%

$$P\left(A \text{ Alghanistan sent back in } 2015\right) = \left(0.5\right)^1 \left(0.5\right)^0 = 0.5$$

This is unfair to them just like the disabled, or like me, being bullied by others. The unfairness has spread around the world. It has been turning into a worse situation. I have to purify the world!’

‘Well, you are not totally evil. If you are, you won’t force us to participate in this “game”. We can change our ending, and human beings, our world will be kind like then. Even if the calculations are correct, we can perform miracles to change and save the world again. We should do it together. Come with us, Perry,’ Messi said.

‘However, my plans were destroyed so I don't think I have any meaning to continue my life. Goodbye, my buddies,’ said Perry. He took out a pistol and killed himself.

Everyone was shocked. Finally, Messi, breaking the silence, said, ‘Let’s go and turn our world better.’

–End –

(No. of words : 2441)

Reference:

1. Football data

http://www.bdfutbol.com/en/p/j1753.html?temp=2014-15

https://www.whoscored.com/Teams/65/Show/Spain-Barcelona

2. Iniesta

http://www.bdfutbol.com/en/j/j159.html

3. Messi

http://www.bdfutbol.com/en/j/j1753.html

4. Suarez

http://www.bdfutbol.com/en/j/j95495.html

5. Self-fulfilling prophecy

https://en.wikipedia.org/wiki/Self-fulfilling_prophecy

6. Disparity

https://www.oxfam.org/en/pressroom/pressreleases/2017-01-16/just-8-men-own-same-wealth-half-world

7. Refugees

https://zh.wikipedia.org/wiki/%E6%AD%90%E6%B4%B2%E7%A7%BB%E6%B0%91%E5%8D%B1%E6%A9%9F

季軍作品:

消防演習背後，香港的消防局

學校名稱：順德聯誼總會翁祐中學

學生姓名：劉永豪、潘以諾

指導教師:阮啟崙

消防演習背後，香港的消防局

正當同學們對著數學測驗卷毫無頭緒時，突然間，所有人均起身逃跑，一陣風捲起了課室的試卷，試卷在空中翩翩起舞，猶如白色的蝴蝶自由地飛翔。課室外刺耳的聲音使人感到厭煩，原來學校正在進行火警演習。

數學才子劉同學問：火警演習會不會弄錯，錯誤地報警求救？

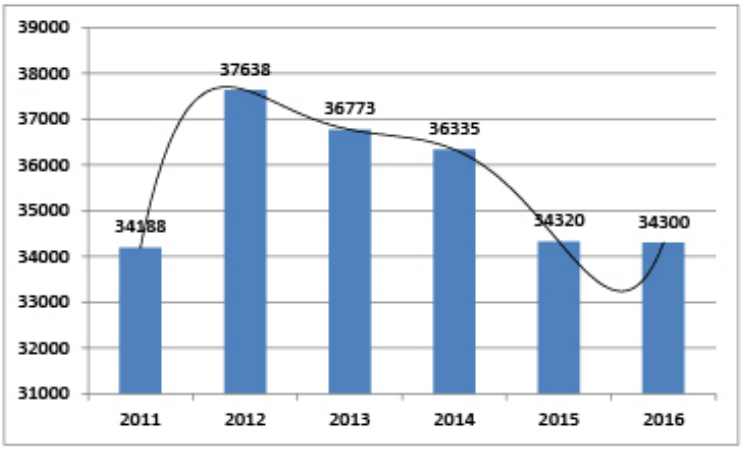
潘：如果消防局出動消防車來到我們學校，同時附近亦發生火警，這樣會不會導致他人因錯過黃金救援時間而不治？

阮老師：哈哈，真好笑，學校演習怎樣會有消防員出現呢！中一級的同學們小息時亂跑撞到警鐘就會。不過的確可能在同一時間發生多宗火警，消防局又會不會不夠人手呢？

潘：平時經常發生電梯困窘等意外都會需要消防員的幫手，消防局到底是怎樣運作的呢？

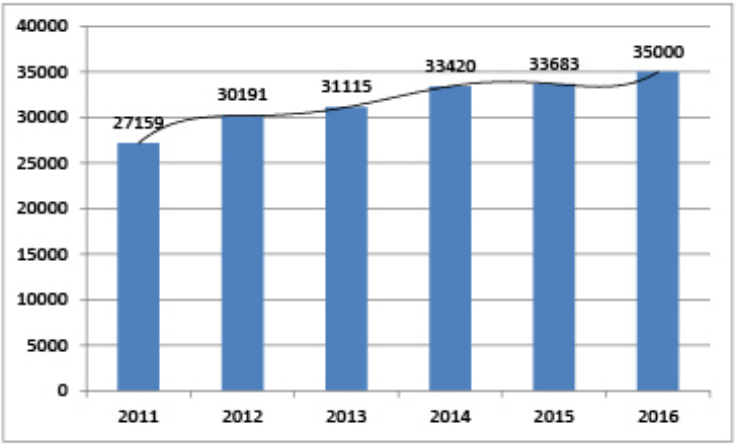
情況一

每年火警召喚總數



圖一

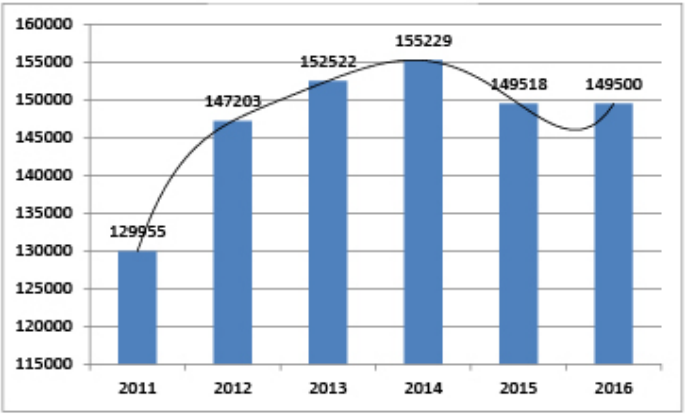
每年特別服務召喚



圖二

劉：根據香港政府網提供的資料，平均每年真正火警召喚總數為35592宗（如圖一），平均每年特別服務召喚總數為31761宗（如圖二）。而全港有87所消防局，所以平均每天每一間消防局有2宗消防召喚。

每年消防車輛出動處理緊急召喚的次數數目



圖三

潘：消防處共有402輛行動組車輛,平均每個消防局有5輛消防車輛，平均一年消防車輛出動處理緊急召喚的次數為147321次（如圖三），平均每天每一間消防局出動5架次。

消防局接到召喚的這個事件發生的次數是獨立的，同時也是

小概率事件，發生的概率也是穩定的。所以消防局每天接到召喚次數是遵從泊松分佈（Poisson Distribution），

$$P\left(X=k\right)=\frac{\lambda^{k}}{k!}e^{-\lambda},k=0,1,2,...$$

其中λ為單位時間內隨機事件的平均發生概率，k為發生的次數。假設一輛消防車出勤和出勤後大概需要4個小時才能再次出勤。所以平均每4個小時會發生$\frac{1}{3}$宗火警召喚，平均每4個小時一間消防局出動$\frac{5}{6}$架。λ即是每4個小時每一間消防局會發生$\frac{1}{3}$宗消防召喚，因爲一共有5輛消防車，而平均每4個小時出動$\frac{5}{6}$架，所以當$k\geq 6$，這個消防局就會進入癱瘓狀態，因爲所有的消防車輛都處於使用中的狀態。所以消防局不會進入癱瘓的概率是（見下表）：

每4個小時消防局接到召喚次數（k）	概率（P）	累積概率
X=0	0.7165	0.7165
X=1	0.2388	0.9554
X=2	0.0398	0.9952
X=3	0.0044	0.9996
X=4	0.0004	≈ 1
X=5	0.00002	≈ 1

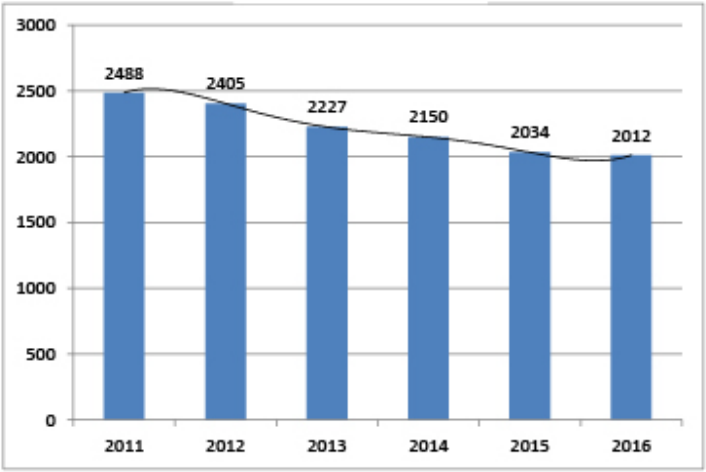
表一

$$P\left(X<6\right)=P(X=0)+P(X=1)+...+P(X=4)+P(X=5)$$
$$P\left(X<6\right)\approx 1$$

阮老師：那麼現在香港的消防局進入癱瘓狀態是接近零，也就是說一般情況下消防局都可以自行解決附近的消防事故。但是其實香港的消防局同時肩負者救護服務的重任。我們同時也要考慮消防局的救護服務。

情況二

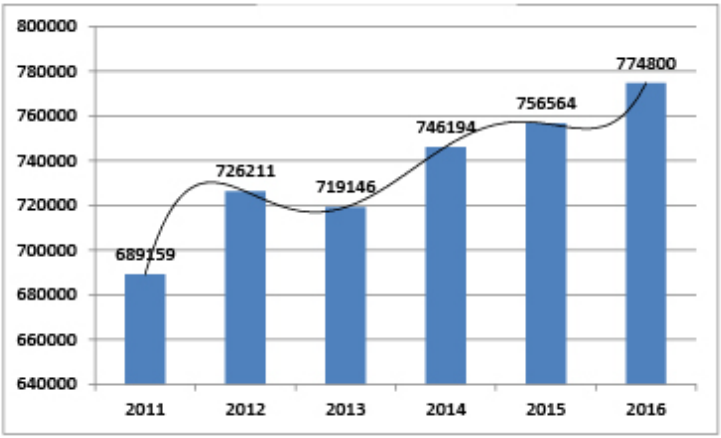
每年每部救護車處理
召喚的平均數目



圖四

潘：現在救護總區共有368輛救護車，香港共有39所救護站和29所救護外局，平均每個救護站或者救護外局有5輛救護車。平均每輛救護車一年處理召喚的平均數目是2219（如圖四）。平均每4個小時處理召喚1次。

每年救護車緊急召喚
和轉院召喚的數目



圖五

劉：平均每年救護車被緊急召喚的數目為735346次，也就是每個救護站每天的救護車平均會被緊急召喚29次。根據泊松分佈（Poisson Distribution），一個救護站每4小時出現緊急召喚的次數分佈為 $P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}, k=0,1,2,...$

λ是每4個小時每一間消防局會發生29/6次緊急召喚，平均每4

個小時一輛救護車處理召喚1次，所以最多可以在4小時之內同時出動5架次，而當$k>5$，這個消防局就會進入癱瘓狀態，以下消防局不會進入癱瘓的概率是（見下表）：

$$P\left(X<6\right)=P(X=0)+P(X=1)+...+P(X=5)$$

每4個小時消防局接到召喚次數（k）	概率（P）	累積概率
X=0	0.0080	0.0080
X=1	0.0385	0.0465
X=2	0.0930	0.1395
X=3	0.1498	0.2893
X=4	0.1810	0.4703
X=5	0.1750	0.6453

潘：救護車不會癱瘓的概率為 65%，怪不得救護員有時間在消防局內打排球！

情況三

劉：我覺得這種演算法可能太過單方面了，當消防處接收一般商住大廈的火警警報後，最接近火警現場的消防局會派出一輛泵車、一輛旋轉台鋼梯車(或梯台車)、一輛油壓升降台(或大型油壓升降台)、一輛細搶救車(或大搶救車)及一輛救護車，俗稱為「四紅一白」出動執勤。87所消防局共有402輛行

動車輛，39所救護站和29間救護外局(即有救護車駐守的消防局)共有386輛救護車，因爲救護外局是既有消防車又有救護車，所以除了每個救護外局有5輛消防車輛和5輛救護車之外，58所消防局和39所救護站各有5輛消防車輛和5輛救護車。

平均一年火警召喚總數35592宗，平均每天每1個救護站有1宗，在加上每天每個救護站被緊急召喚有29次，所以每天每個救護站會面對30個出勤車次。λ是每4個小時每一間消防局會發生5次緊急召喚，平均每4個小時一輛救護車處理召喚1次，所以最多可以在4小時之內同時出動5架次，而當$k>5$，這個消防局就會進入癱瘓狀態，以下消防局不會進入癱瘓的概率是(見下表)：

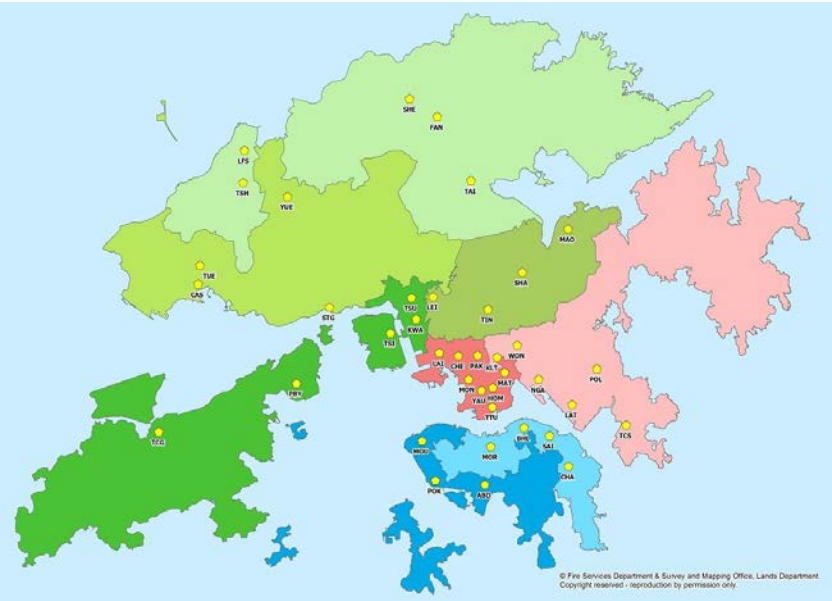
$$P\left(X<6\right)=P(X=0)+P(X=1)+...+P(X=5)$$

每4個小時消防局接到召喚次數（k）	概率（P）	累積概率
X=0	0.0067	0.0067
X=1	0.0337	0.0404
X=2	0.0842	0.1246
X=3	0.1404	0.265
X=4	0.1755	0.4405
X=5	0.1755	0.616

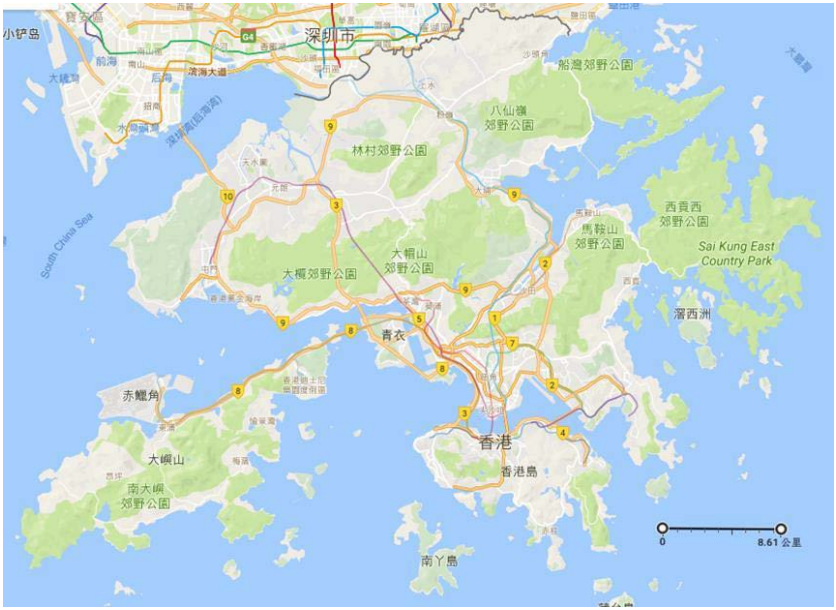
表三

潘：咦，反而這樣計算出來，不會癱瘓的概率超過了60%。不過消防局除了要有充足的消防車之外，同時還要求救護車輛要在指定時間內到達現場。

香港的緊急救護服務要求在12分鐘內到達現場，圖六就是香港18區68所救護站與救護外局的分佈圖。爲了可以準確得出這個地圖的實際距離，我們把它和Google Map(圖七)重疊在一起，得出圖八。



圖六

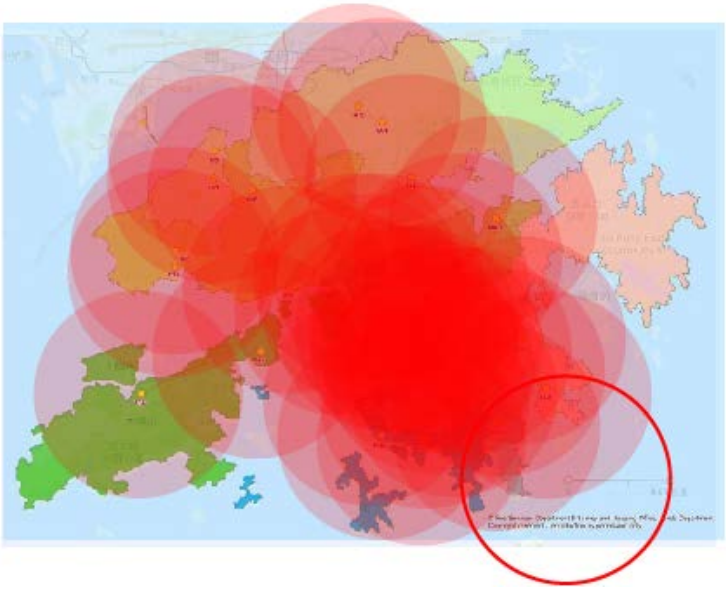

小铲岛
深圳市
前海
南山
后海
凌海大道
船灣郊野公園
八仙嶺
郊野公園
林村郊野公園
西貢西
郊野公園
馬鞍山
郊野公園
Sai Kung East
Country Park
大帽山
郊野公園
大欖郊野公園
South China Sea
青衣
滘西洲
赤鱲角
香港
香港島
大嶼山
南大嶼
郊野公園
南丫島
0
8.61 公里

圖七

小铲岛
深圳市
前海
南山
后海
凌海大道
船灣郊野公園
八仙嶺
郊野公園
林村郊野公園
西貢西
郊野公園
馬鞍山
郊野公園
Sai Kung East
Country Park
大帽山
郊野公園
大欖郊野公園
South China Sea
青衣
滘西洲
赤鱲角
香港
香港島
大嶼山
南大嶼
郊野公園
南丫島
0
8.61 公里

圖八

救護車在緊急救援服務中的平均時速為 47km/hr，假設救護員在接到通知的1分鐘後出發前往現場，在11分鐘內救護車大概可以到達8.61公里的地方。以救護站為圓心，8.61公里為半徑作圓。(如圖九)



圖九

劉：這是眼力測試嗎？九龍半島和香港島的地圖都淹沒在這片紅海底下了。

潘：雖然全香港有68所救護站，但是有些偏僻的地方沒有被覆蓋，也就是說有些地方當發生意外的時候，救護車不能在12分鐘內到達，難道緊急救援在12分鐘內到達現場的概率達不到100% 是因爲這些地方嗎？

阮老師：修建消防局和救護站就像修建樓房一樣，有些地方可能規劃署已經規劃了其用途，所以就不能修建消防局和救護站。另外，沒有被覆蓋的地區大多是郊野公園，即使發生意外，救護車也不能上山救人，可能交給飛行服務隊處理。

潘：地圖上越紅的地方也是越多人會去的地方，油尖旺、中環、金鐘、銅鑼灣，這些地方都同時被多個救護站所覆蓋。

劉：人口密集的地方，發生意外的概率也會更大，多個救護站就能更好地應變突發情況。

阮老師：人口密集的地方，車流量也會大，交通狀況很難預計，經常會遇上塞車等狀況，再加上紅綠燈密集，這些都會延遲救護車到達現場的時間，其實一個救護站的服務半徑可能小於8.61公里。同時有很多人爲的因素會影響，例如是司機對路面的熟悉程度。

潘：經過這次的專題文章的經歷，我發現原來消防員的工作也十分繁忙和辛苦。

阮老師：你們要認真讀書啦！做哪一行都是需要付出才有收穫。

(字數：2430)

參考資料:

1. 消防處2016統計資料

http://www.hkfsd.gov.hk/chi/source/statistics_2016chi.pdf

2. 消防處2011年報

http://www.hkfsd.gov.hk/images/review_11/tc_eng/appendix.pdf

3. 维基百科，自由的百科全书-香港消防處

https://zh.wikipedia.org/wiki/%E9%A6%99%E6%B8%AF%E6%B6%88%E9%98%B2%E8%99%95 -.E8.BB.8A.E8.BC.9B

優異作品: 特首寶座，花落誰家？

學校名稱：保良局何蔭棠中學

學生姓名：陳嘉敏 曹紹棠 麥素瑤

指導教師：陳智仁

特首寶座，花落誰家？

引言：

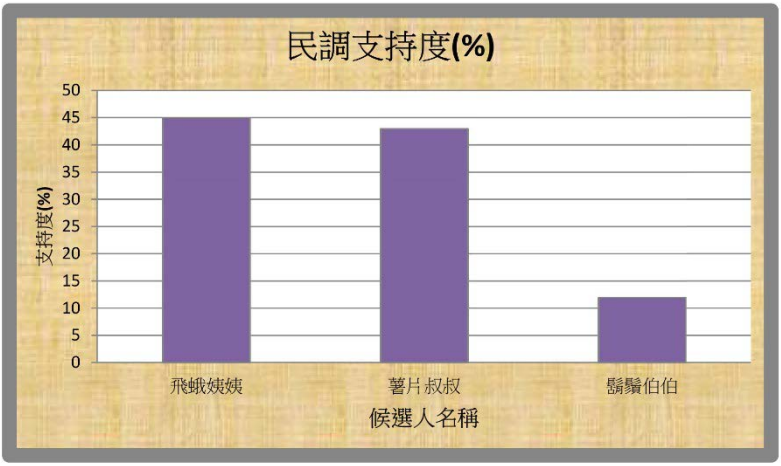
特首選舉將至，無論在網上還是普通市民，飯後討論的話題都是圍繞三位特首候選人(飛蛾姨姨，薯片叔叔，鬍鬚伯伯)，特首的寶座究竟花落誰家，當然成為全城熱話。大家平時討論時只存在主觀意見，但為何不用統計學計算一下他們的當選率呢？究竟當選率跟支持率有甚麼關係？結果令人萬眾期待！

****

特首選舉將至，無論網上還是普通市民的飯後討論話題都是圍繞三位特首候選人（飛蛾姨姨，薯片叔叔，鬍鬚伯伯）小陳，小曹，小麥亦正在談這個城中熱話...

小陳：嘿!小曹，小麥，三位候選人當中你們認為誰當選的概率最大呢？

小曹：根據非官方調查顯示，飛蛾姨姨目前的民調支持率達是三位候選人中支持率最高的一位，應該是她的當選概率最大吧！



圖表 1

小麥：有道理！過往經驗都是支持率越高，當選的機會越高。

小陳：嗯，若只觀察民調支持率，飛蛾姨姨的當選概率的確最高，但是不要忘記不是得票最多的候選人就能當選，而是要過半票數才行啊！

小曹：對哦！選委人數共120人，若要過半票...(計算中)...那麼當選人至少要獲得61票，但各候選人當選的概率是多少呢?不如我們動手計計各候選人當選的概率吧！

小麥：應該與民調支持率一樣吧!當選率一定與支持率成正比的。

小陳：嗯...(思考)...未必...

小曹：不如我們動手計計各候選人每種票數的概率吧！
(他們開啟excel 開始計算)
（飛蛾姨姨，薯片叔叔，鬍鬚伯伯各當選概率圖表）

小麥：但是這些圖表只是顯示每種票數的概率，我們只能看到每種票數的概率都很少，應該如何分析呢？

小陳：首先，我們應該計算他們過半數當選的概率。方法是求他們得超過 60 選票的所有概率。

飛蛾姨姨： $\sum_{r=61}^{120} C_r^{120} (0.45)^r (1-0.45)^{120-r} \approx 11.67\%$

薯片叔叔： $\sum_{r=61}^{120} C_r^{120} (0.43)^r (1-0.43)^{120-r} \approx 5.09\%$

鬍鬚伯伯： $\sum_{r=61}^{120} C_r^{120} (0.12)^r (1-0.12)^{120-r} \approx 3.91\times 10^{-23}\%$

小曹：咦！為甚麼他們的當選概率總和不是1呢？
$11.67\% + 5.09\% + 3.91\times 10^{-23}\% \neq 1$

小麥：因為你忘記了額外條件—過半數支持人數才能當選啊！在此項條件下，即使其中一位候選人獲得最高票數亦未

必能夠當選呢!因此概率總和不是1 是合理的。

小陳：嗯，說得沒錯...咦!(驚訝)從數據看來，即使飛蛾姨姨的民調支持率最高，但是能夠過半票數當選的概率也很低只有11.67%！

小曹： 那是因為她的支持率還未過50%呢。不過原來過半票數當選率與支持率可以相差這麼多呢。

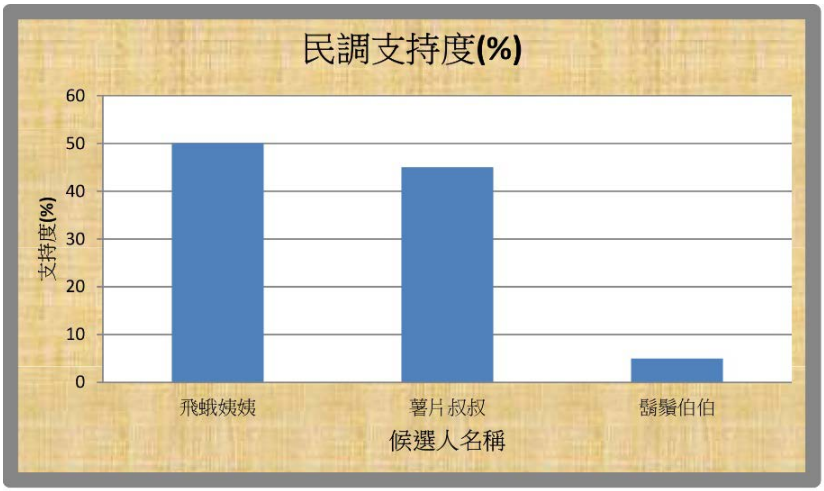
小麥：支持率經常都是浮動不停，這個民調支持率也許只是暫時性的。

小曹：對啊！不如我們分析一下其他情況吧！

小陳：我們已經分析了飛蛾姨姨獲45%支持率。此次的額外條件是需要過半票數，即是超過50%，我們可以分析當她的支持率剛達致50%，及當她的支持率超過50%的情況。

小曹：好，那我們先分析她的支持率剛剛達致50%的情況吧。

小麥：不如我們先假設各候選人的支持率先吧。



圖表2

小陳：在這個情況下，飛蛾姨姨的當選率達50%比起真實支持率45%多了5%，但當選率由11.37%大幅提升到46.37%，足足多了34.7%呢！你們留意一下之前的數據。

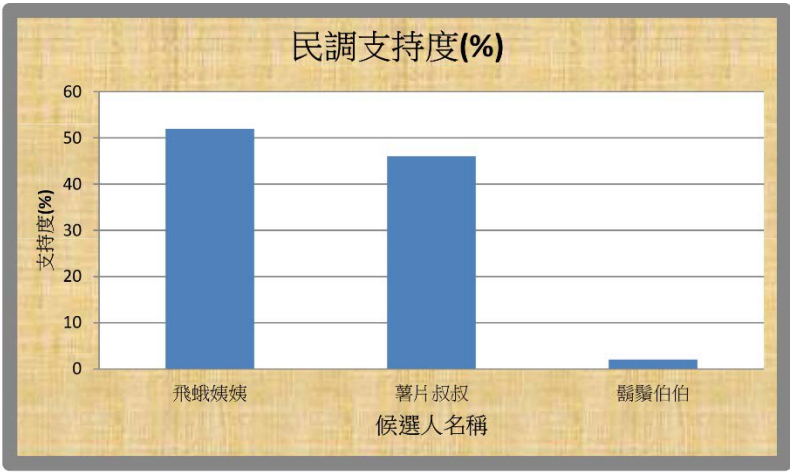
支持率	過半票當選率
43%	5.09%
45%	11.67%
50%	46.37%

看來當支持率愈大 當選率的上升速度的愈快呢！

小麥：嗯，但大家有沒有發現即使支持率達到50%，當選率亦未達一半。

小陳：看來若要當選率超過50%，支持率一定要超過50%，那麼我們計算一下支持率超過50%個情況吧！

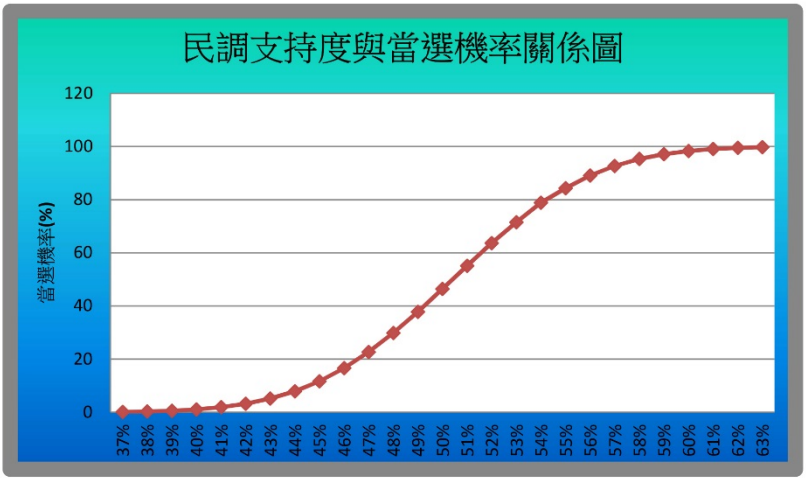
小曹：我們先假設飛蛾的支持率達52%。



圖表3

小麥：哇！在支持率52%情況下，她的當選率竟超過63.61%

小曹：原來支持率上升2%，當選率可以上升這麼多啊！但到底當選率與支持率的關係是什麼呢?不如我們製作一個關係圖吧。



圖表4

小陳:看來當選率雖然與支持率成正比,但是當支持率在50%前後,只要再增加幾個百分點,當選率就能大大提升。

小麥：對啊！剛開始時我們還以爲飛蛾姨姨的支持率是三位候選人中最高，就一定過半票數當選概率很高。但計算後得出原來支持率達45%時當選率只有11.67%。

小曹：嗯，三位候選人若想提高過半票數當選，都要在餘下的時間盡力爭取支持率方能取得最後勝利。

小麥：話說回來你們認為誰能當選呢？

小陳：應該是...應該是他們三個其中一個吧！

小麥、小曹：噓.....

參考資料:

https://theinitium.com/article/20161215-hongkong-run-for-ce/

(字數：1687 字)

附錄:

- 民調支持度及得票概率比對

  參考附件檔案: 計算方法

- 表4 之數據

	民調支持度與當選概率關係圖		民調支持度與當選概率關係圖
37%	0.1368	51%	55.106
38%	0.2843	52%	63.6091
39%	0.5596	53%	71.4975
40%	1.046	54%	78.84744
41%	1.8595	55%	84.3557
42%	3.15	56%	89.0786
43%	5.094	57%	92.6896
44%	7.8771	58%	95.3165
45%	11.6673	59%	97.1331
46%	16.5809	60%	98.3264
47%	22.6483	61%	99.0699
48%	29.7882	62%	99.5089
49%	37.7979	63%	99.7541
50%	46.3658		

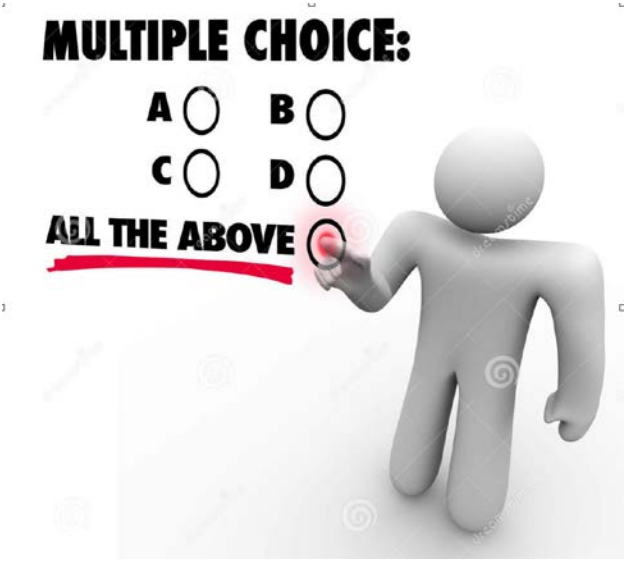
優異作品: 「盲」衝直「撞」

學校名稱：保良局何蔭棠中學

學生姓名： 趙穎詩，陳琳，葉俊銘

指導教師： 陳智仁教師

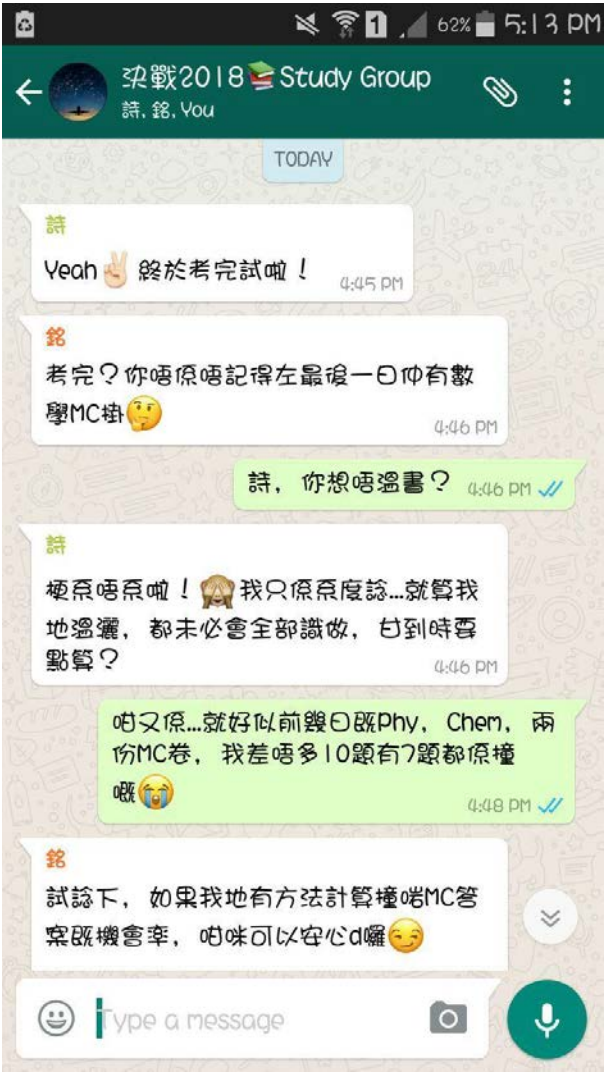
「盲」衝 直「撞」



引言

在我們的考試生涯中，每一科試卷上總會看見多項選擇題的蹤影，它們往往都令我們絞盡腦汁而又無法解決。當同學遇上不懂作答的問題，這個時候，你們又會怎樣做來增加自己答對的概率呢？就讓我們探討一下在四選一情況下答對的概率。

選擇題(Multiple choice question)是一種題型或方法，指的是從一組選項中，選出一個或多個作為正確(或較合適)的答案。而多項選擇題在考試中一般是指單選題(只選一個選項)，若是複選題(可選多個選項)，題目後面則會標上「可選一個或以上答案」。


62% 5:13 PM
決戰2018 Study Group
詩, 銘, You
TODAY
詩
Yeah 終於考完試啦！
4:45 PM
銘
考完？你唔係唔記得左最後一日仲有數學MC掛
4:46 PM
詩，你想唔過書？ 4:46 PM
詩
梗係唔係啦！我只係係度諗...就算我地過關，都未必會全部識做，白到時要點算？
4:46 PM
咁又係...就好似前幾日既Phy，Chem，兩份MC卷，我差唔多10題有7題都係撞嘅
4:48 PM
銘
試諗下，如果我地有方法計算撞啱MC答案既機會率，咁咪可以安心d囉
Type a message

詩：如果我們在考試前沒有溫習的話，而又想在不同科目如中文、物理、化學、數學考試中的多項選擇題中取得及格的成績，我們猜中答案的概率會是多少？

銘：我們先假設答案選項平均分佈。若在中文科中，題目一共有5 題，而問題總共有三個選項，

$$\sum_{x=3}^{5}\frac{5!\left(\frac{1}{3}\right)^{x}\left(1-\frac{1}{3}\right)^{5-x}}{x!(5-x)!}=\frac{17}{81}$$

我們答中的概率將會是 ≈ 0.2099

琳：那麼我們猜中的概率還挺高的！

銘：你先看看其他科目的概率吧！肯定會讓你驚訝的！在物理科中，問題總共有四個選項，而題目一共有33 題，我們答中的概率將會是：

$$\sum_{x=17}^{33}\frac{33!\left(\frac{1}{4}\right)^{x}\left(1-\frac{1}{4}\right)^{33-x}}{x!(33-x)!}=\frac{35084231326971293}{36893488147419103232}$$

$\approx 9.5096\times 10^{-4}$

詩：概率怎麼一下子小了這麼多！

銘：是啊，這世界是沒有免費午餐的，哪有甚麼不溫習就能及格的方法呢！所以你們還是好好回家溫習吧。

琳：這是不可能的，我不相信，要不我們再看看化學、生物和數學這三科吧！

銘：在化學和生物科中，問題都共有四個選項，而題目分別都有36題，我們答中的概率將會是：

$$\sum_{x=18}^{36}\frac{36!\left(\frac{1}{4}\right)^{x}\left(1-\frac{1}{4}\right)^{36-x}}{x!(36-x)!}=\frac{2521639568585725457}{3361183241434822606848}$$

$\approx 1.0680\times 10^{-3}$

銘：而在數學科中，問題亦共有四個選項，而題目一共有45題，我們答中的概率將會是：

$$\sum_{x=23}^{45}\frac{45!\left(\frac{1}{4}\right)^{x}\left(1-\frac{1}{4}\right)^{45-x}}{x!(45-x)!}=\frac{2292353343941091868648 7}{1547425049106725343623905 28}$$

$\approx 1.4814\times 10^{-4}$

詩：概率好像越來越小了，這是不是代表我們猜中的概率也越來越小呢？

銘：不錯，所以我們還是要在考試前努力溫習才能獲得好成績，不要總想著走捷徑。

詩：我們求學當然是要求分數！ 如果我們在考試中還想獲得更高的分數，那麼我們會猜中的概率是會比合格的較高還是較低？

銘：以化學科和數學科為例，在化學科中，假如題目數量依然是 36 題，選項數量仍然是 4 個，我們想取得百分之七十五的分數，我們猜中的概率將會是：

$$\sum_{x=27}^{36}\frac{36!\left(\frac{1}{4}\right)^{x}\left(1-\frac{1}{4}\right)^{36-x}}{x!(36-x)!}=\frac{517833291607}{1180591620717411303424}$$

$$\approx 4.3862\times 10^{-10}$$

銘：而在數學課中，假如題目數量依然是45題，選項數量仍然是4個，而我們想取得百分之七十五的分數，我們猜中的概率將會是：

$$\sum_{x=34}^{45}\frac{45!\left(\frac{1}{4}\right)^{x}\left(1-\frac{1}{4}\right)^{45-x}}{x!(45-x)!}=\frac{125342941784017}{7737125245533626718119526 4}$$

$$\approx 1.6200\times 10^{-12}$$

琳：那如果我們在每題中都有一至兩個選項可以肯定地排除，那麼猜中的概率又會變成多少？

詩：這個我知道，讓我以數學科爲例吧！假如題目數量依然是45題，選項數量仍然是4個，但我們能排除當中的一個選項，若我們依然是追求及格的分數，那我們猜中的概率將會是：

$$\sum_{x=23}^{45}\frac{45!\left(\frac{1}{3}\right)^{x}\left(1-\frac{1}{3}\right)^{45-x}}{x!(45-x)!}=\frac{10144827514025439857}{984770902183611232881}$$

≈ 0.01030

詩：而如果我們能排除當中的兩個選項，我們依然是追求及格的分數，那我們猜中的概率將會是：

$$\sum_{x=23}^{45}\frac{45!\left(\frac{1}{2}\right)^{x}\left(1-\frac{1}{2}\right)^{45-x}}{x!(45-x)!}=\frac{1}{2}$$

琳：咦，概率又變得越來越大了，那麼就是我們能排除的答案越多，答中及格的概率就越來越高？

銘：你說對了！

琳：現在文憑試的選擇題模式與從前會考年代的完全不同，那麼我們在哪種考試中會比較容易取得及格的分數呢？

詩：這個我也不是很清楚，不如就讓我問問我們的數學老師陳老師吧，他應該可以解答我們的疑問。

詩：我已經請教過陳老師了，現在讓我來解釋一下吧。同樣地，我們先假設答案選項平均分佈。在會考年代的中文科中，題目一共有5題，而問題總共有四個選項，我們答中的概率將會是：

$$\sum_{x=3}^{5}\frac{5!\left(\frac{1}{4}\right)^{x}\left(1-\frac{1}{4}\right)^{5-x}}{x!(5-x)!}=\frac{53}{512}$$

≈ 0.1035

詩：而在物理科中，問題總共有四個選項，題目一共有45題，我們答中的概率將會是：

$$\sum_{x=23}^{45}\frac{45!\left(\frac{1}{4}\right)^{x}\left(1-\frac{1}{4}\right)^{45-x}}{x!(45-x)!}=\frac{2292353343941091868648 7}{1547425049106725343623905 28}$$

$\approx 1.4814\times 10^{-4}$

詩：在化學科中，問題總共有四個選項，而題目一共有50題，我們答中的概率將會是：

$$\sum_{x=25}^{50}\frac{50!\left(\frac{1}{4}\right)^{x}\left(1-\frac{1}{4}\right)^{50-x}}{x!(50-x)!}=\frac{3882606262993817814171259 3}{3169126500570573503741758013 44}$$

$\approx 1.2251\times 10^{-4}$

銘：剩下的生物科和數學課就讓我來解釋吧！在生物科中，問題都共有四個選項，題目有60題，我們答中的概率將會是：

$$\sum_{x=30}^{60}\frac{60!\left(\frac{1}{4}\right)^{x}\left(1-\frac{1}{4}\right)^{60-x}}{x!(60-x)!}=\frac{4435346737611804531201387219071}{1661534994731144841129758825350430 72}$$

$\approx 2.6694\times 10^{-5}$

銘：在數學科中，問題亦共有四個選項，而題目一共有54題，我們答中的概率將會是：

$$\sum_{x=27}^{54}\frac{54!\left(\frac{1}{4}\right)^{x}\left(1-\frac{1}{4}\right)^{54-x}}{x!(54-x)!}=\frac{2696817447852195038174837441}{4056481920730334084789450257 2032}$$

$\approx 6.6482\times 10^{-5}$

琳：哦，原來在不同科目中，在會考和文憑試猜中而又能讓我們及格的概率是這麼不同的，讓我試試來作一個總結吧！

琳：在中文科中，因爲 $0.2099>0.1035$，因此我們在文憑試中較在會考中更容易猜中且及格。

琳：在物理科中，因為 $9.5096 \times 10^{-4} > 1.4814 \times 10^{-4}$，因此我們在文憑試中較在會考中更容易猜中且及格。

琳：在化學科中，因為 $1.0680 \times 10^{-3} > 1.2251 \times 10^{-4}$，因此我們在文憑試中較在會考中更容易猜中且及格。

琳：在生物科中，因為 $1.0680 \times 10^{-3} > 2.6694 \times 10^{-5}$，因此我們在文憑試中較在會考中更容易猜中且及格。

琳：在數學科中，因為$1.4814\times 10^{-4} > 6.6482 \times 10^{-5}$，因此我們在文憑試中較在會考中更容易猜中且及格。

銘：做得好！但你有沒有發現它們有一個共同點？

詩：我知道，就是在各科中，除了中文科以外，在會考中，各科的題目數量都比文憑試的要多，這是不是就代表在文憑試中較容易及格呢？

銘：你說得沒錯，但在我們的研究中沒有考慮到會考和文憑試中的題目深淺難度，所以我們不可以只靠這研究結果而妄下判斷，會考是設計予中五同學應考的而文憑試則是設計予中六同學的，二者所涵蓋的課題更是不同，或許同學在會考所懂得的題目比文憑試的更多呢？但是，只要你在考試中不是臨陣磨槍，在考試中能儘量把選項排除到最少，那麼你猜

中的概率也會變得更高。

詩： 你還遺漏了一點，就是如果我們想在同一科考試中獲得更好的成績，獲得百分之七十五的分數的概率會比獲得及格分數的低出很多。所以想在考試中脫穎而出，靠自己的努力才是上乘之策。

參考資料:

1. https://zh.wikipedia.org/wiki/%E9%81%B8%E6%93%87%E9%A1%8C (選擇題)

2. http://www.wolframalpha.com/widget/widgetPopup.jsp?p=v&id=d983db47634e1936eb568b79884012ac&title=Summation%20Calculator&theme=blue&i0=i%5E2&i1=10&i2=11&podSelect=&includepodid=Result&showAssumptions=1&showWarnings=1

   (Summation Calculator)

3. https://www.dreamstime.com/royalty-free-stock-photos-quiz-multiple-choice-choosing-best-answer-test-words-several-answers-person-selection-as-his-image32875758

(字數：2053 字)

優異作品：Usage of Priority Seats in Hong Kong

School Name: Belilios Public School

Name of Student: LAI Yuen- ting Tiffany, LI Victoria,

WONG Hoi-shun Jessica

Supervising Teacher: Ms. Yip.SL

Usage of Priority Seats in Hong Kong

Photo from: http://www.td.gov.hk/en/priority_seats/index.html

Introduction

With regard to the soaring concern towards the traditional Chinese belief filial piety, priority seats have been promoted in buses, MTR and many more. This measure has excited a hot debate among the community as many people doubt the effectiveness and limitations of the priority seats in enhancing the sense of filial piety in Hong Kong. Hence, we are interested in this issue and decided to investigate the current situation and give some possible solutions to the problem. We have conducted a sampling to collect data about the usage of priority seats in the MTR for consecutively two weeks and the data collected is analysed.

Story

After a long school day, Tiffany, Victoria and Jessica are travelling back home by the MTR together. Jessica and Victoria are quite exhausted after the school day and decided to take a seat.

Jessica: What a tiring day!

Victoria: Yes indeed. The exercises we did in today's PE lesson is really exhausting! I am worn out!

Jessica: It’s easy! Let’s take a seat!

Tiffany: Oh wait! Are you guys trying to sit on the priority seats?

Jessica, Victoria: The seat is vacant now. Why not?

Tiffany: You guys should be aware that these seats are only for those who are in need, which includes the disabled, the pregnant, the elderly and those parents with kids.

Victoria: But I don’t think everyone preserved this principle as much as you do. People do have many controversies towards this issue.

Jessica: So why don’t we conduct a sampling to find out the usage of priority seats in Hong Kong?

Victoria, Tiffany: Sure. What a good idea!

After two weeks…

Victoria: Let’s look at what we found out!

Table 1: Distribution of age group in peak and non-peak hours

Timeslot	Peak hours*			
Total number	77			
Age group+	children	teens	working	elderly
Number of people using the seats	9	2	42	24

Timeslot	Non-peak hours			
Total number	77			
Age group+	children	teens	working	elderly
Number of people using the seats	20	3	17	37

*Peak hours refers to the period of 7:00a.m. to 9:00 a.m. and the period of 6:00p.m. to 7:30p.m. during Mondays to Fridays.

+ Children are those below 13. Teens are those with the age between 13 and 19 years old. Working are the people between 20 and 59 years old. Elderly are those with the age above 60 years old.

Table 2: Frequency of yielding seats to those in need

	0 times	1 times	2 or more times
Peak hours	56	17	4
Non-peak hours	68	4	0

Table 3: Distribution of yielding of seats to those in need

	Children	Teens	Working
Peak hours (times)	1	1	19
Non-peak hours (times)	0	1	3

Jessica: Dividing the timeslots into two categories can help us compare the usage of the seats at different time.

Tiffany: Dividing the people into different age groups can let us know what is the age distribution of the people that are using the seats.

Victoria: Let’s discuss the distribution and proportion of the age group that use the seats first!

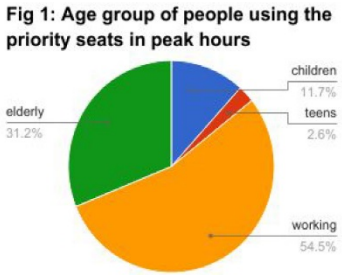


Fig 1: Age group of people using the priority seats in peak hours

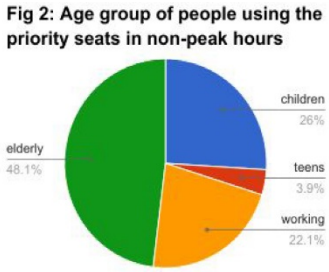


Fig 2: Age group of people using the priority seats in non-peak hours

Victoria: I have chosen to use pie charts (i.e. Fig 1 and 2) to show you all my data as long as it helps show the distribution of age groups. The percentage of each age group can be represented using the following formula

$$\text{Percentage}=\frac{\text{Number of people in the particular age group}}{\text{Total number of people}}\times 100\%$$

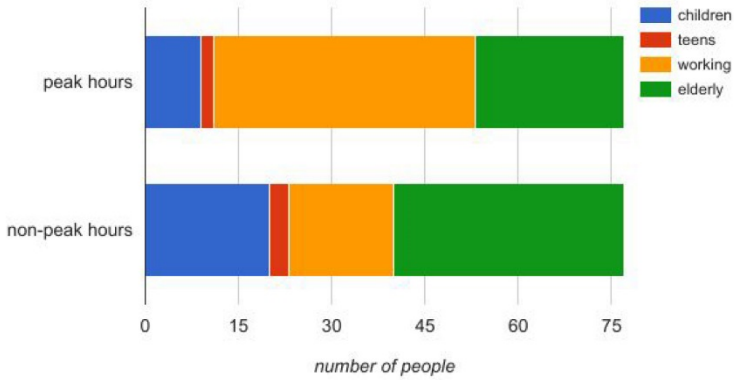
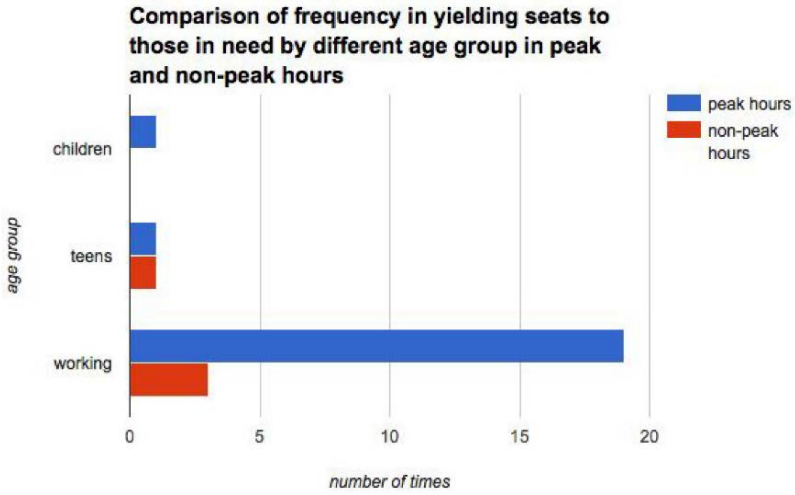


Fig 3: Proportion of different age group in peak and non-peak hours

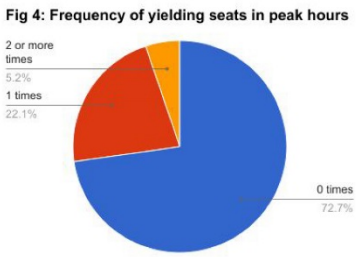
Jessica: And let's move on to the part of proportion of different age groups in peak as well as non-peak hours. From the data (i.e. Fig 1, 2 and 3), we can see that the working population owes the largest percentage of 54.5% and in the non-peak hours, the elderly owes the largest percentage of 48.1%. The usage of priority seats is not distributed fairly and does not fit the rationale behind setting up priority seats.

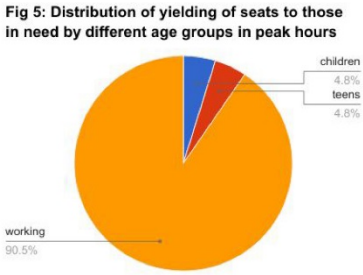
Tiffany: As the base number is the same (i.e. 77 people), it is good for us to make use of a stacked bar chart (i.e. Fig 3) to compare the proportion of different age group in peak and non-peak hours. We can see that the category of 'teens' still has the lowest proportion in both peak and non-peak hours but the proportion of the category of 'working' has decreased in non-peak hours, comparing to that in the peak hours.



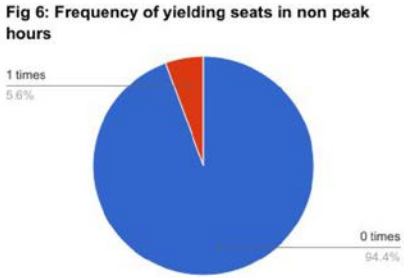
Jessica: Here are the results of the distribution of yielding of seats to those who are in need by different age groups. I have some stunning and startling findings. Shall we have a look in it?

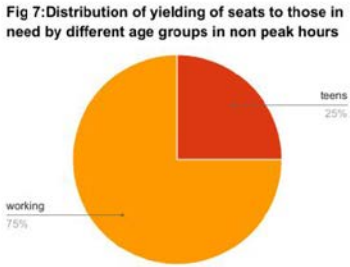
Victoria: Sure!





Jessica: From the pie chart concerning the peak hours (i.e. Fig 4 and 5), we can see that there was only 4.8 percent of children and students are yielding seats respectively which is only two in total out of 77,while for the remaining 90.5 percent, they are owned by the working population in Hong Kong. The figure really surprised me a lot in which the rationale behind setting up priority seats is not well-fitted.





Tiffany: On top of that, for the pie chart concerning the non-peak hours (i.e. Fig 6 and 7). I was stunned by the fact that there was no children yielding seats as well as only four people had yield their seats. The working population has hold a majority of 75 percent while the remaining 25 percent are owned by the teenager group.

Victoria: Wow! I was undoubtedly startled when I have heard about the low yielding seats rate. May I ask what phenomenon does the percentage shown?

Table 2

Frequency of yielding seats to those in need	0 times	1 times	2 or more times	Total number
Peak hours	56	17	4	77
Non-peak hours	68	4	0	77

Jessica: Sure. You have mentioned the phenomenon of a low yielding seat rate. According to the data we collected (i.e. Table 2), 56 out of the 77 people in the survey did not yield the seat to the people that meet the criteria of using the priority seats. This holds the majority of 72.7 percent. Meanwhile, only 37.3 percent of the passengers would yield their seat. And for the distribution of yielding of seats to those who are in need by different age group, it shows a relatively low rate too. From this, we can see that most of the Hong Kong people have a relatively low sense of filial piety and sense in showing respect to those in need.

Tiffany: Apart from peak hours, the results for non-peak hours are similar to that of peak hours, and even more serious than that! According the pie chart (i.e. Fig 6) showing the frequency of

yielding seats in non-peak hours, 94.4 percent of the people sitting on the priority seats did not yield their seats. The percentage is even higher than that of peak hour. Compared the results of non-peak hour to that of peak hours, it is not difficult to find out that most teens have a weaker desire to yield their seats to the people in need as only one of them in each timeslots had yield the seat. And the yielding of seats by the working population in peak hours is fourfold to that of non-peak hours. But why?

Victoria: Erm… In my opinion, maybe the working population also wear out after a long day of work and run out of battery. Just like us! (Tiffany and Jessica are laughing)

Jessica: Yes, I can't say that you are wrong. But back on the track, there are many conflicts towards the usage of priority seat. And these conflicts that have emerged over the issue are mainly due to different moral standards. According to news from AM730, 40% of the elderly cannot received seats when they travel on public transport. They are ignored by the people sitting on the seats that are focusing on their mobile phones.

Tiffany: Oh! I often heard news about the conflicts towards the

usage of priority seat. I remembered that they argue with each other and even fight because of it. The issue of priority seat has raised a lot of disputes.

Victoria: In order to ease the discontent towards this issue in Hong Kong, I think more priority seats are needed as there is not enough priority seats. Many people who are in need don't have a seat and they are forced to stand for a long period of time.

Tiffany: Yes! As a matter of fact, the concept of yielding seat to the needy in Hong Kong is not yet deep-rooted in Hong Kong people's heart. Therefore, the government should hold more talks or other activities at schools to educate the younger generation. Hence, the civil awareness of the youngsters can be heightened.

Victoria: In addition to this, I think more promotion on the usage of priority seat is required. For instance, posting more advertisements on the public transports by using colourful posters and interesting educational clips through the television, etc.

Tiffany: So would you guys pay more attention to the surroundings and yield seats to those in need starting from now?

Victoria, Jessica: Absolutely yes!

Conclusion

After the investigation towards the usage of priority seats in Hong Kong, we found that working group is mainly the group that sit on the priority seats in peak hours. And elderly are the principal group that sit on priority seats in non-peak hours. In the meantime, with regard to the data we collected, what we found is that the frequency of people yielding priority seat to people in need is relatively low, which is below 50%. These show that the concept of yielding seat to people in Hong Kong is not enough that more promotion is needed. Education also plays a pivotal role that children should learn to yield seat too.

Reference:

1. 怕網絡公審73%中學生不坐關愛座(Retrieved on 18 January 2017)
   http://hk.apple.nextmedia.com/news/art/20170118/19900615

2. A matter of priority: educate travellers on seats for most needy (Retrieved on 17 November 2013)
   http://www.scmp.com/comment/letters/article/1358043/matter-priority-educate-travellers-seats-most-needy

3. Train Improvement Works - Priority Seats
   https://www.mtr.com.hk/en/corporate/projects/projects_ops_improve_train_seat.html

4. Seating plan You must have noticed the MTR's priority seats - 星島教育 (Retrieved 18 January 2016)
   http://stedu.stheadline.com/sec/sec_news.php?aid=16337&cat=2&subcat=9

優異作品:

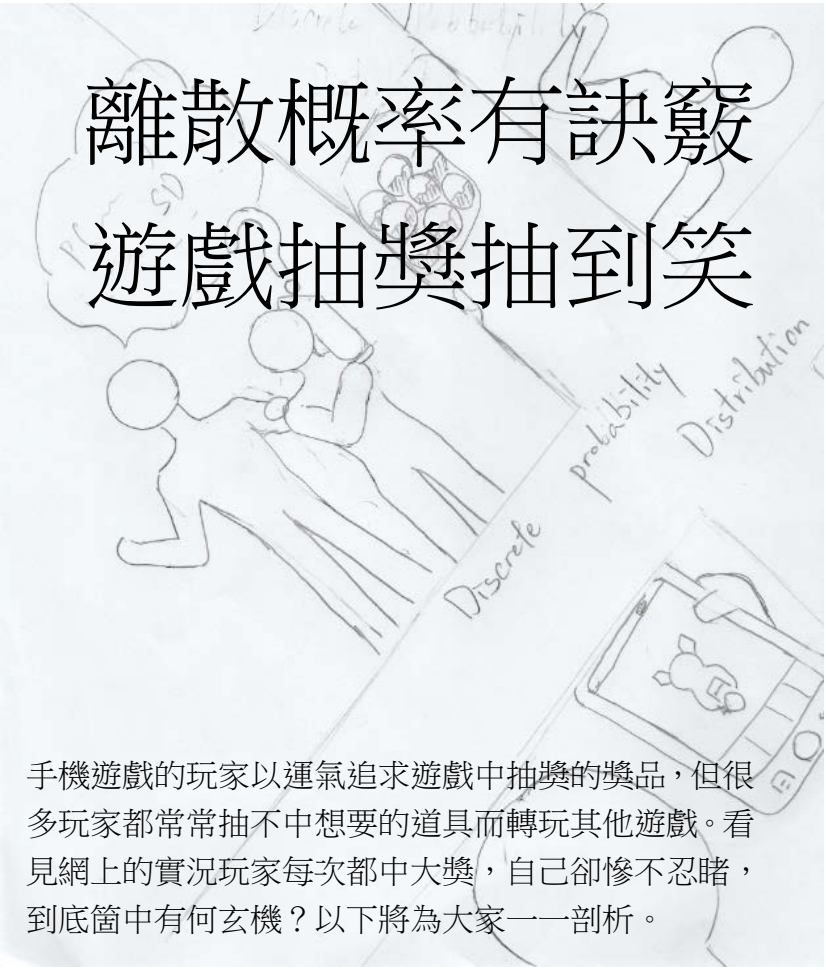
離散概率有訣竅，遊戲抽獎抽到笑

學校名稱：聖芳濟書院

學生姓名：吳泳航 鄭俊灝 王志超

指導教師：陳柏基老師

離散概率有訣竅
遊戲抽獎抽到笑



手機遊戲的玩家以運氣追求遊戲中抽獎的獎品，但很多玩家都常常抽不中想要的道具而轉玩其他遊戲。看見網上的實況玩家每次都中大獎，自己卻慘不忍睹，到底箇中有何玄機？以下將為大家一一剖析。

小明：老師好！ 最近我們一起玩的那個遊戲推出了新人物了！ 我應不應該買一些鑽石來抽獎呢？

龍大師：哎，玩玩就好，何必浪費金錢呢？再加上你不是學會了離散概率分佈了嗎？何不自己算算呢？

小明：我當然知道！不過我都已經把這課還給你了呢，所以才有勞你呢！我連**變量**是甚麼都忘記了。

龍大師：原來醉翁之意不在酒。讓我再教你一次吧！首先要明白離散概率分佈的特別之處。這個概念只限可數的獨立事件或物件的發生概率與趨勢，即分佈。比如說抽中某同學答問題，因爲每個同學都是一個個體，可以數出他的數量。相反，像是一些速率，例如不同種類雪糕融化的速度，身高變動這些便是不能數的，所以不是離散。

小明：就是說抽 10 次角色，抽中 1 隻 4 星的事件是離散概率分佈的原因是他們可以被數？

龍大師：沒錯！每一個時段巴士經過的數量，打牌抽到同花順的概率都可以用離散概率分佈來分析。弄清離散概率分佈的限制後，就要逐一項目解釋了。讓我弄一個表：

概念詞	解釋
變量 (random variable)	即在一個事件中，指定情況所出現的次數
二項分佈 (The binomial distribution)	即用來計算事件的概率，中位數，差距等的分析項目
幾何分佈 (The geometric distribution)	

概念詞	解釋
P(變量) (probability of 變量)	即該變量在整個事件中的發生概率 $(0 \le P(\text{變量}) \le 1)$
獨立事件 (independent event)	即該項事件不會影響到另一事件的發生概率

就用我們的例子先計算，然後再解釋吧!整個抽人物的池分有 2 星，3 星，4 星人物。有 18 隻 2 星，30 隻 3 星，和 16 隻 4 星。現在就將 X 當成一個代表在 10 次抽人中抽到多少隻 4 星角色的變量，即 $0 \le x \le 10$。那麼根據概率 $p = \dfrac{\text{預期結果的數量}}{\text{總結果的數量}}$。

每 1 次抽人抽中 1 隻 4 星角色的概率爲 $p = \dfrac{16}{(18+30+16)} = \dfrac{16}{64} = \dfrac{1}{4} = 0.25$。我們利用**二項分佈**來算一下 X 的數值不同，它的概率又有何改變？

$$X \sim B(10, 0.25)$$

$$P(X = x) = C_x^n p^x (1-p)^{n-x}$$

$$P(X = 0) = C_0^{10} (0.25)^0 (0.75)^{10} = 0.0563$$

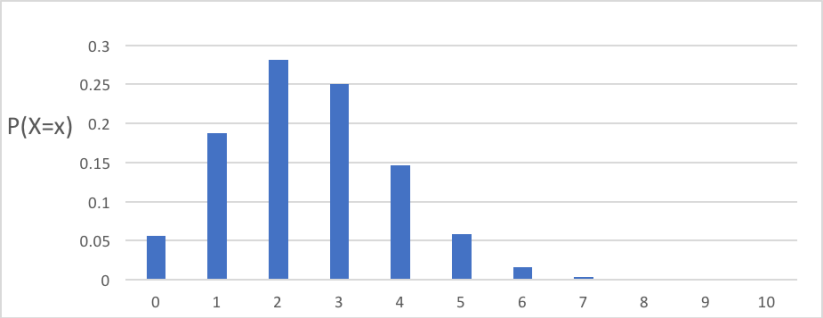
$$P(X = 1) = C_1^{10} (0.25)^1 (0.75)^9 = 0.1877$$

……

$$P(X = 10) = C_{10}^{10} (0.25)^{10} (0.75)^0 = 0.000001$$

x	0	1	2	3	4	5
$P(X=x)$	0.0563	0.1877	0.2817	0.2503	0.1460	0.0584

x	6	7	8	9	10
$P(X=x)$	0.0162	0.0031	0.0004	0.00003	0.000001



在上述例子中

$P(X=0)$代表在 10 次抽人中了 0 隻 4 星人物的概率，

$P(X=1)$代表在 10 次抽人中了 1 隻 4 星人物的概率。

如此類推。可看出其實 10 抽中 2 隻 4 星角色的概率是最大的，這也是離散概率分佈的其中一個算式和結果，可看出在 10 次抽人物的話，抽中 2 隻 4 星的概率分佈得較多。

小明：那麼我抽 10 次最高概率抽中 2 隻 4 星，即是 X 的期望值(Expected Value)是 2 嗎？

龍大師：你還記得期望值是離散的其中一環，但你算錯了！在二項分佈的情況下期望值 $E(X) = np$。即 10 抽預期會中 $10 \times 0.25 = 2.5$ 隻 4 星角色。

小明：為甚麼是 np 呢？

龍大師：其實期望值是把所有變量的數值乘以它的出現概率 $\sum xf(x)$，才能計算它平均能得到多少 X。如果好像平常找平均值一樣，把所有數加在一起然後除總數的話，就忽略了概率這因素。因為如果其中一個 X 的數值的出現概率較低卻用平常的方法，就會高估了該數值，而導致期望值會與現實不符。所以每一個 X 的數值乘了出現概率就能得到一個正確的比例。給你看一些外星符號吧。

小明：哈...哈...真後悔我沒認真上課。我剛找到了一條公式，是標準差 σ，當中 $\sigma^2 = np(1-p)$，好像是想找出 X 各個數值與中位數的差距，派考卷時常看到，可除了這個還有何用途呢？

龍大師：這正可以計算抽 4 星人物的事件。剛看到 $p = 0.25$，那麼標準差 $\sigma = \sqrt{10 \times 0.25 \times (1-0.25)} = \sqrt{1.875} = 1.369$，代表

各數值與平均值$E(X)$平均地相差 1.369。不過 1 個 SD 並不能表達甚麼，如果比較其他 SD，哪一方的 SD 較小就代表那組數據較接近平均值，較易估計，少誤差。所以標準差其實是一個比較各數據的準則來的。如果標準差越大，代表每個數值跟平均數的差距越大，在成績來說可代表班裡的差距大喔！

小明：哇！今天可真是上了幾個月的課了^_^。

龍大師：那好，你現在還有疑問嗎？

小明：當然有！ 我發現有些遊戲中的扭蛋機有限制蛋的數量，情況就您剛才舉的例子一樣。有一些卻沒有限制扭蛋機中蛋的數量。請問這兩種情況如何影響我的中獎概率呢？

龍大師：你能受現實啟發，想到這問題，實在是不枉我一直以來的教導。好吧，我就把畢生功力都傳給你，助你在遊戲裡屢創高峰！

小明： 就以這個高達遊戲作為例子吧！

扭蛋機中的蛋（道具）	概率
SFXC 高達（價值高） X_1	0.1
渣古（價值中） X_2	0.2
炮灰形吉姆（價值低） X_3	0.3
輔助道具（價值最低） X_4	0.4

假設我抽 10 次扭蛋機，我現在會利用**幾何分佈**算出在第 x 次($1 \le x \le 10$) 抽到 SFXC 高達的概率：$X \sim \text{Geo}(0.1)$

$$P(X=x)=(1-0.1)^{x-1}(0.1)$$

$$P(X=1)=(1-0.1)^{1-1}(0.1)=0.1$$

$$P(X=2)=(1-0.1)^{2-1}(0.1)=0.09$$

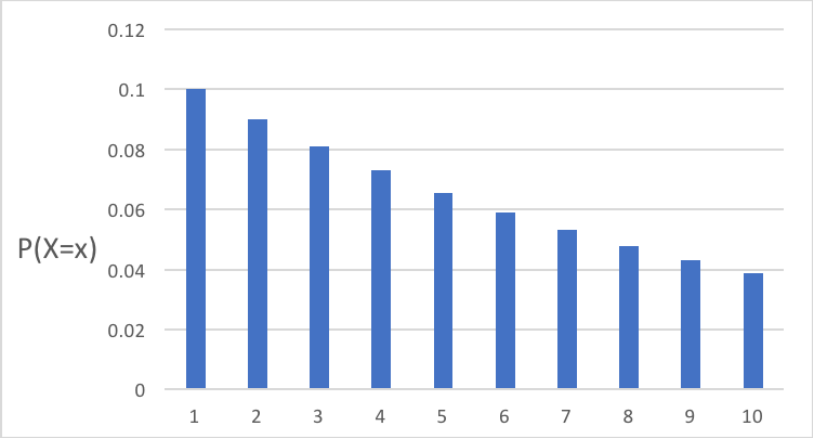
$$P(X=3)=(1-0.1)^{3-1}(0.1)=0.081$$

$$P(X=4)=(1-0.1)^{4-1}(0.1)=0.0729$$

$$P(X=5)=(1-0.1)^{5-1}(0.1)=0.06561$$

…

X	1	2	3	4	5
$P(X=x)$	0.1	0.09	0.081	0.0729	0.0656
X	6	7	8	9	10
$P(X=x)$	0.0590	0.0531	0.0478	0.0430	0.0387



小明：咦？等等！為甚麼我抽的次數愈多，抽到 SFXC 高達的概率就愈低呢？ 難道第四次數學危機會因我而爆發？

龍大師：咳咳！小明你冷靜一點，別再幻想了！我不知道未來會否發生第四次數學危機，但我卻知你犯了個致命的錯誤，以致得出了一個不合理的結論。你剛才利用幾何分佈計算出在第 x 次$(1 \le x \le 10)$ 抽到 SFXC 高達的概率的想法是好的，計算過程也沒有錯，那為何看似抽的次數愈多，抽到 SFXC 高達的概率就愈低呢？真相只有一個 — 你忽略了 $P(X=x)$ $(1 \le x \le 10)$的結果之間的關係了。$P(X=1)$ 至 $P(X=10)$

的概率是互不相干的，比如 $P(X=5)$指的是在正好第五次抽到 SFXC 高達，並不包括在第一次就抽到 SFXC 高達的情況。若你想知道在 x 次$(1 \le x \le 10)$ 以內抽到 SFXC 高達的概率，你可以用 $P(X=x) + P(X=(x-1)) + \ldots + P(X=1)$計算：

$P(X=1)=0.1$

$P(X \le 2) = 0.1 + 0.09 = 0.19$

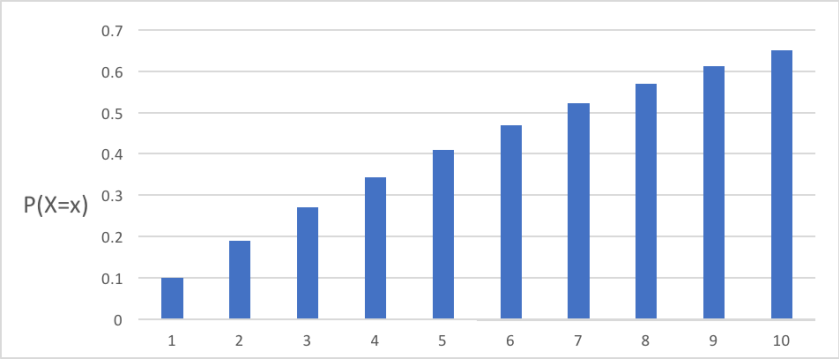
$P(X \le 3) = 0.19 + 0.081 = 0.271$

$P(X \le 4) = 0.271 + 0.0729 = 0.3439$

$P(X \le 5) = 0.3439 + 0.06561 = 0.40951$ 如此類推

現在的概率分佈便正常的多了，抽得愈多，抽到 SFXC 高達的概率便愈大。事實上，若你把 $P(X=1) + P(X=2) + P(X=3) + \ldots + P(X=\infty)$ 加在一起，便會發現它們的和正好是 1 (Total probability)！因為它們是**獨立事件（Independent events）**，這也是所有離散概率分佈都包含的特點。如此類推：

X	1	2	3	4	5
$P(X \le x)$	0.1	0.19	0.271	0.3439	0.4095
X	6	7	8	9	10
$P(X \le x)$	0.4686	0.5217	0.5695	0.6126	0.6513



龍大師：小明你可以根據以上計算作出預測，估計抽到 SFXC 高達的概率，再具體一點，我們會發現當你抽 7 次扭蛋機時，抽到 SFXC 高達的概率便已經超過一半了！

小明：咦？等等！我又有問題了，明明扭蛋機只有十件道具，應該每抽一件便少一件才對，但為何我抽 10 次扭蛋機時，抽到 SFXC 高達的概率不是 100% 而是約 65.1%呢？

龍大師：小明，你真有慧根！重點來了，我們剛才運用幾何分佈計算時，是假設每抽一次扭蛋機後均會重置扭蛋機內的道具，令扭蛋機內道具的數量回復原狀，也就是說這是**有重置（with replacement）**的情況。每一次抽蛋都是獨立事件，影響不到下一次抽蛋。如果你想知道在不重置扭蛋機道具的

情況下，在某一次抽蛋抽到 SFXC 高達的概率，便不能用幾何分佈，而需要用更複雜的方法進行計算了。

小明：噢！真是感謝你呢！其實明天 有個小測驗，我只是想向你請教請教而已，哈哈。

龍大師：你...算了！幫到你就好！加油！

（字數: 2495 字）

參考資料：

1. Effective Learning Mathematics Module 1: Calculus and Statistics by Chung Tai Educational Press
2. https://en.wikipedia.org/wiki/Probability_distribution
3. https://en.wikipedia.org/wiki/Geometric_distribution

邀請作品：

人口普查與人口登記冊

人口普查

相信大家對於人口普查並不陌生。世界各地很多國家／城市都有進行人口普查／中期人口統計。自 1961 年起，香港每 10 年進行一次人口普查，並在兩次普查中間，進行一次中期人口統計。最近一次就是於 2016 年 6 月至 8 月期間進行的 2016 年中期人口統計，所得的結果亦在 2017 年陸續發放。

人口普查／中期人口統計搜集有關人口的廣泛社會及經濟特徵資料。由於樣本數量大，細小地區的統計數字能較準確估算。這些數字不單對政府在規劃和制定政策十分重要，亦是學者和其他機構研究社會民生議題的重要基礎。私人公司亦會參考人口統計數字制定合適的公司政策。因此，人口普查／中期人口統計對每個社會的發展都十分重要。

然而，亦正因為每次人口普查／中期人口統計所訪問住戶數目眾多，人力、時間和金錢均花費不菲。以香港 2016 年中期人口統計為例，調查期間訪問全港約十分之一的屋宇單位，

即約 30 萬個住戶，需要增聘超過 6000 名臨時外勤工作人員以應付龐大的訪問工作，整個統計調查花費超過 3 億元。人口普查更需訪問全港數百萬個屋宇單位，2011 年人口普查期間增聘了約 1 萬 8 千名臨時外勤工作人員，整個普查花費超過 5 億元。隨著人口膨脹、家庭數目增長和物價上升，意味著未來的人口普查／中期人口統計所需的花費將會更大。單單是 2 萬名訪問員訪問時使用的平板電腦的費用就有點嚇人吧？！

另一方面，由於人口普查／中期人口統計是一個難得的機會訪問大量的市民，調查問卷一般涉及多方面的問題。香港 2016 年中期人口統計的問卷就有 43 條題目，包括房屋例如房屋類型、房間數目；住戶例如住戶人數、居所是租是買、租金、房屋按揭供款；人口例如年齡、性別、婚姻狀況、教育、工作情況、收入、國籍、種族、慣用語言等等。一個 3 至 4 人的住戶就需要約 50 分鐘的面談時間完成問卷。隨著社會進步，政府乃至不同機構團體皆希望更深入和透徹了解社會現狀，對人口普查／中期人口統計的數據需求亦隨之增加。數據需求和受訪者的負擔(respondent burden)兩者更難維持平衡。增加受訪者的負擔更會影響回應率(response rate)和問卷答案的質素。

面對以上的問題，是否沒有解決方法呢？事實上，一些國家例如丹麥、芬蘭、冰島、挪威和瑞典已利用人口登記冊(population registers)全面取締人口普查。其他國家例如荷蘭、德國和新加坡則利用行政紀錄計算人口數目，再把其他住戶統計調查的結果合併，令訪問成本和受訪者負擔大幅減少。

人口登記冊

甚麼是人口登記冊(population registers)呢？以下是聯合國的定義：

“The Population register is a mechanism for the continuous recording of selected information pertaining to each member of the resident population of a country or area, making it possible to determine up-to-date information about the size and characteristics of the population at selected points in time.”

簡單來說，人口登記冊是一個地方的名冊，名冊內有這個地方居住的所有人，每個人用一個獨有的號碼分辨（類似身份證號碼）。名冊內亦有每個人基本而不會改變或不經常改變的資料，譬如出生年份、性別、國籍和種族。而這本名冊內的資料會定期更新，例如加入新生嬰兒和外來移民，亦標記已經死亡和移民外地的人。倚靠這本名冊，就可以知道某年

某月某日，在這個地方的人數有多少、他們的年齡分佈、性別比例等等的基本特徵。

人口登記冊還可以與政府其他現有的行政紀錄系統及住戶統計調查的結果合併。以荷蘭為例，他們的人口登記冊與勞工保險登記系統(Employee Insurance Schemes Registration System)、社會援助福利行政系統(Social Assistance Benefits Administration)和稅務系統等結合，從而合併政府部門和政府相關機構現有的行政資料。再結合一些統計調查(Survey on Employment and Earning, Labour Force Survey)的結果，就可得到全面的人口統計數字。

人口普查最原始的目的就是準確點算人口數目，了解他們的年齡、性別分佈。有了人口登記冊，人口普查的基本目的就達到，發展成熟後就可以以中期人口統計取締人口普查。甚至單靠定期進行的住戶訪問，就能獲得需要的統計數字。

使用人口登記冊可以大幅減低受訪者的負擔、節省成本及無須等待 5 年一次的人口普查／中期人口統計才能編制需要的統計數字。更多數字由整個地方的人口資料編制，而非由一個統計調查訪問部分人而估算，得出的統計數字亦較準確。

事實上，人口普查／中期人口統計問及的資料，很多都是政府現有的。例如從稅務系統、社會福利系統有市民的工作及其他收入，從婚姻及離婚系統可以知道市民婚姻狀況，從學生紀錄可以知道學生上學地點和教育程度。然而，各政府部門的資料系統一般是獨立的。基於私隱考慮，各地法例亦通常規定政府部門不能發放收集的個人紀錄予其他機構，包括其他政府部門。部門一般只能提供合計的統計數字作參考。

那有甚麼方法解決私隱問題呢？

荷蘭就使用一套複雜的系統去儲存人口登記冊內的資料，從而保護個人私隱。

如果個人資料的變量，例如身份證號碼、出生日期、姓名、地址等標示為 I，其他統計變量，例如收入、教育程度等標示為 S。那人口登記冊內的資料就是

[I, S]

最簡單的方法就是把 I 和 S 拆開。

[I, S] =>[I] [S]

這樣就不可能追查到統計變量是屬於那一個人。但是，由於

個人資料和統計變量的連結已經完全斷開，我們無法再把人口登記冊連結到其他行政系統或統計調查結果。達不到應有的效果。

另一個常用的方法是再加入一套新的識別號碼(identifying number)。例如香港市民現有的識別號碼是身份證號碼，我們可以用電腦隨機產生另一組與身份證號碼相對應、每個人都不同的識別號碼 N。以 N 替代 I 作為登記冊的識別號碼。

$$[I, S] \Rightarrow [N, I, S] \Rightarrow [N, I][N, S]$$

這樣在一般情況只需要處理[N, S]編制統計數字，過程就無法辨認某個紀錄的真實身份。當有需要合併其他系統時，就以 N 找回 I。

你可能會說，這樣也是很容易破解呀。只要持有[N, I]和[N, S]的人合作就可以了。沒有錯，所以荷蘭最終使用的方法是再加入另一個識別號碼及使用加密技術。首先個人資料和統計變量各自有一個識別號碼。

$$[I, S] \Rightarrow [N_I, I, N_S, S] \Rightarrow [N_I, I][N_S, S]\ [N_I, N_S]$$

再把[N_I, N_S]加密，

$$[N_I, I][N_S, S]\ [N_I, N_S] \Rightarrow [N_I, I][N_S, S]\ [E(N_I), E(N_S)]$$

這樣變更加密的方法時，只需要處理[$E(N_I)$, $E(N_S)$]。N_I和 N_S獨立的時候只是一堆無意義的數字。加密過程沒有牽涉統計變量，就會更安全。再加上其他保安措施，例如統計部門的內聯網不能連接外界的互聯網、負責處理個人資料變量的員工不能發放有附件的電郵等等。務求全面保護個人資料不會外泄。

除了私隱問題，使用人口登記冊還有很多技術環節需要考慮。這些環節複雜，在這裡就先不討論。使用行政紀錄進行統計工作是很多統計先進的地方朝進的方向，而以人口登記冊編制統計數字有很大的發展潛力，香港或許可以向這些地方借鏡。

參考資料

1. 政府統計處 2016 年中期人口統計網站 - 常見問題
   http://www.bycensus2016.gov.hk/tc/faq.html

2. 政府統計處 2011 年人口普查網站 - 常見問題
   http://www.census2011.gov.hk/tc/faq.html#q9.7

3. Website of United Nations Statistics Division
   https://unstats.un.org/unsd/demographic/sources/popreg/popregmethods.htm

4. Statistics Netherlands, “Data security, privacy and the SSB”, Netherlands Official Statistics Volume 15, Summer 2000
   https://www.researchgate.net/profile/Paul_Van_Der_Laan2/publication/269675136_Integrating_Administrative_Registers_and_Household_Surveys/links/5492d1d30cf2302e1d0742d4/Integrating-Administrative-Registers-and-Household-Surveys.pdf

5. Frank Linder, Statistics Netherlands, “The Dutch Virtual Census 2001: A new approach by combining Administrative Registers and Household Sample Surveys”, Austrian Journal of Statistics Volume 33 (2004), Number 1&2, 69-88
   http://www.ajs.or.at/index.php/ajs/article/view/vol33%2C%20no1%262%20-%204

邀請作品：

如何提高隊伍在「中學生統計創意寫作比賽」當中獲獎的機會？對過往三屆參賽作品的統計分析

「中學生統計創意寫作比賽」已經舉辦了八屆。每一年評審委員會都收到很多優秀的作品，作品題目形形色色，而且每一份作品都是同學們認真努力下得出的心血。然而作品衆多、獎項有限，究竟參賽隊伍如何能夠提升作品得獎的機會？

例如：

- 多人隊伍比一人隊伍容易創作好作品？
- 高年級同學比起低年級同學作品較優秀？
- 字數越多越容易得獎？
- 用中文還是英文參賽好？
- 有過往參賽經驗的隊伍更有優勢？

我們整理了最近三屆參賽的 123 份作品的行政數據，並嘗試分析有可能影響作品得獎與否的因素。

近 3 屆作品的基本統計

參賽作品

近 3 年我們每年均收到 30 - 50 多份的作品。參賽同學約 40%是中五或中六的學生；而且超過 70%同學是以 2 人或以上隊伍形式參加。大約 13%的隊伍當中有同學是再次參賽。收到的作品方面，約 70%作品是以中文創作，另外 30% 作品則由英文創作。作品的字數平均約 1 900 字，但由最少字數至最多字數的範圍甚廣（675 字至 4 227 字）。

得獎作品

每年我們均會有 14–16 個不等的得獎隊伍（冠、亞、季、優異作品和入選作品）[1]。在得獎隊伍當中，近半(47.7%)是中五或中六的學生；大約 9%的隊伍當中有同學是再次參賽。作品方面，超過 80%作品是以中文創作。作品的字數平均約 2 300 字。

詳細的統計數字可參閱***表一***。

[1] 比賽還設有「香港大學統計及精算學系最佳專題寫作獎」以及「理大香港專上學院最佳文章演繹獎」。過去三屆比賽，這兩項大獎的得獎隊伍均同時獲得三甲或其他獎項，因此已被包括在得獎隊伍當中。

表一：「中學生統計創意寫作比賽」參賽和得獎作品統計，2014-2017

比賽年度		所有比賽年度（2014-17）		2014-15		2015-16		2016-17	
		參賽作品	得獎作品	參賽作品	得獎作品	參賽作品	得獎作品	參賽作品	得獎作品
總計		123	44	45	16	43	14	35	14
級別	中三	41 (33.3%)	10 (22.7%)	14 (31.1%)	5 (31.3%)	19 (44.2%)	3 (21.4%)	8 (22.9%)	2 (14.3%)
	中四	34 (27.6%)	13 (29.5%)	12 (26.7%)	4 (25.0%)	9 (20.9%)	5 (35.7%)	13 (37.1%)	4 (28.6%)
	中五／中六	48 (39.0%)	21 (47.7%)	19 (42.2%)	7 (43.8%)	15 (34.9%)	6 (42.9%)	14 (40.0%)	8 (57.1%)
隊員數目	1	33 (26.8%)	13 (29.5%)	12 (26.7%)	7 (43.8%)	16 (37.2%)	4 (28.6%)	5 (14.3%)	2 (14.3%)
	2	32 (26.0%)	8 (18.2%)	13 (28.9%)	3 (18.8%)	6 (14.0%)	1 (7.1%)	13 (37.1%)	4 (28.6%)
	3	58 (47.2%)	23 (52.3%)	20 (44.4%)	6 (37.5%)	21 (48.8%)	9 (64.3%)	17 (48.6%)	8 (57.1%)
參賽經驗	首次參賽	107 (87.0%)	40 (90.9%)	39 (86.7%)	14 (87.5%)	36 (83.7%)	13 (92.9%)	32 (91.4%)	13 (92.9%)
	再次參賽	16 (13.0%)	4 (9.1%)	6 (13.3%)	2 (12.5%)	7 (16.3%)	1 (7.1%)	3 (8.6%)	1 (7.1%)
作品語言	中文	87 (70.7%)	36 (81.8%)	32 (71.1%)	12 (75.0%)	30 (69.8%)	13 (92.9%)	25 (71.4%)	11 (78.6%)
	英文	36 (29.3%)	8 (18.2%)	13 (28.9%)	4 (25.0%)	13 (30.2%)	1 (7.1%)	10 (28.6%)	3 (21.4%)
作品字數	平均值	1919.1	2275.3	1901.6	2275.9	2003.0	2401.9	1838.3	2147.9
	最小值	675	925	675	925	1 139	1 139	1 000	1 126
	最大值	4 227	4 227	3 518	3 483	4 227	4 227	2 842	2 842

註：
1. 括號內數字表示數字佔所有作品的百分比。
2. 得獎作品包括冠、亞、季、優異作品和入選作品。
3. 如同一隊內包括不同級別同學，則以所有隊員當中的最高級別作為該組別的級別。
4. 如同一隊內包括首次參賽和再次參賽的同學，則該隊會被歸類為「再次參賽」。

邏輯迴歸分析（Logistic Regression Analysis）

為了更進一步分析不同因素對於得獎概率的影響，我們利用了近三年的數據進行了**邏輯迴歸分析**。簡單而言，邏輯迴歸分析可以了解不同因素對於一個二元結果（是與否；生與死；染病與沒有染病；勝與負）的「勝算比」（Odds Ratio，或者 OR）。

舉一個簡單的例子，根據上面的統計數字，中文和英文參賽作品的得獎分佈如下：

作品的語言	**獲獎**	**沒有獲獎**	**所有作品**
中文	36	51	87
英文	8	28	36
所有作品	44	79	123

使用中文參賽而得獎相比使用英文參賽而得獎的 Odds Ratio 如下：

$$\frac{N_{\text{中文，得獎}} / N_{\text{中文，沒有得獎}}}{N_{\text{英文，得獎}} / N_{\text{英文，沒有得獎}}} = \frac{36/51}{8/28} = 2.471$$

換句話說，使用中文的獲獎勝算為使用英文的獲獎勝算的 2.5 倍。

利用上面的數據，我們可以建立一個**邏輯迴歸模型**如下：

$$\log (p / (1-p)) = \boldsymbol{ß_0} + \boldsymbol{ß_1 X_1}$$

而當中 $\boldsymbol{ß_0}$ 是 y 截點(y-intercept), $\boldsymbol{X_1}$ 代表作品語言（中文為 1, 英文為 0），而 $\boldsymbol{ß_1}$ 則為中文作品(相對於英文作品)的 log(Odds Ratio)〔換句話說，exp($\boldsymbol{ß_1}$)就是中文作品（相對於英文作品）的 Odds Ratio〕。

運用電腦統計軟件分析數據，我們得出以下的結果：

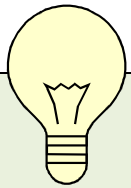
	ß	Z 值	P 值	Odds Ratio (=exp(ß))	置信區間(C.I.)	
					2.5 %	97.5 %
(Intercept)	-1.2528	-3.125	0.00178 **	0.2857	0.1215	0.5982
作品語言 - 中文作品 [1]	0.9045	1.983	0.04740 *	2.4706	1.0460	6.3724

[1] 相比起以英文作為作品語言。

從上表可見，作品語言的 Odds Ratio 的置信區間為[1.046, 6.372]。在 95%置信區間之下，中文作品相比英文作品有統計上明顯較高的勝算。

但是，除了作品所採用的語文，我們也有興趣了解其他因素對於得獎與否的影響。因此，我們將其他各個因素一併放進邏輯迴歸模型當中：

$$\log(p/(1-p)) = ß_0 + ß_{作品語言} X_{作品語言} + ß_{學生級別} X_{學生級別} + ß_{隊員數目} X_{隊員數目} + ß_{參賽經驗} X_{參賽經驗} + ß_{作品字數} X_{作品字數}$$



<u>為甚麼不為各個因素分別進行邏輯迴歸模型，而要將各因素一併放進同一個模型作單一運算呢？</u>

單一模型有兩個主要好處：

1. 在模型中每一個因素都是 Conditional Odds Ratio，即是假設其他情況不變之下該因素的 OR。這樣可以了解到各個因素在摒除其他影響之後的實際影響；
2. 每進行一次統計運算，均有相當機會因為「運氣」而誤將實際上無效的差別當作統計上的顯著差別（即 Type I error）。如果為每一個因素分別進行運算，那麼因為「碰運氣」而意外視作明顯差別的機會便會提高（即 inflated Type I error）；相反，如果使用單一模型，只需要一次計算，「碰運氣」的機會便會被研究員所設定的置信區間所控制，減少「誤中副車」的情況。

電腦統計軟件的運算結果如下：

	ß	Z 值	P 值	Odds Ratio (=exp(ß))	置信區間(C.I.)	
					2.5 %	97.5 %
(Intercept)	-1.4685	-1.9220	0.0546	0.2303	0.0483	0.9855
學生級別 - 中四 [1]	0.6717	1.1190	0.2630	1.9576	0.6063	6.4856
學生級別 - 中五/六 [1]	0.8493	1.5330	0.1253	2.3380	0.7993	7.1313
隊員數目	-0.0804	-0.2890	0.7725	0.9228	0.5319	1.5954
作品字數	0.0015	4.0740	0.0000***	1.0015	1.0008	1.0023
作品語言 - 中文作品 [2]	0.7783	1.5180	0.1289	2.1779	0.8233	6.2771
參賽經驗 - 再次參賽 [3]	-1.2886	-1.8100	0.0702	0.2756	0.0608	1.0367

[1] 相比起中三同學。

[2] 相比起以英文作為作品語言。

[3] 相比起首次參賽。

從運算的結果可以看到，當將不同的因素一併放進邏輯迴歸模型後，情況便跟先前結果有所不同了：

(1) **作品語言**不再是統計學上明顯的因素。

其中一個可行的解釋是，大部份作品語言對於勝算的影響，其實被另一個因素——作品字數——更好地被解釋（例如英文作品字數較少）。因此當各因素被同時間放進模型內被控制不變的情況下，作品語言所顯示出的影響便減弱不少了。

(2) 只有**作品字數**是一項統計上明顯影響作品勝算的因素。

那麼我們如何理解**作品字數的影響力呢？**從其 Odds Ratio 可以看到，作品每多一字，作品的勝算便多出 0.0015 ；換句話說，作品每多 1000 字，作品的勝算便是原來的 1.5 倍(0.0015*1000)。

其中一個可能的解釋，是用心和有內容深度的作品往往會採用較多字數；相反，字數太少的話可能會局限了同學闡述作品理念的空間。

(3) 其餘的因素（如**學生級別、參賽經驗、隊員數目**）與勝算均沒有顯著關係。

值得留意的原本其中一個假設，是較高年級的同學的數學課程較高深，因此作品可能相比中三同學較為優秀。然而，這個迴歸模型卻未能確立這個假設。事實上，不少中三學生的作品均非常優秀，絲毫不比高年級同學遜色。另一方面，近屆所訂立的比賽專題需要同學在數學課程以外作更多研習，因此亦收窄了不同級別之間同學的差別。

結論

以上的統計和分析利用過去三屆的行政數據作一次邏輯迴歸模型的簡單示範。事實上，在評審工作期間，上述統計模型所包括的因素（學生級別、過往參賽次數、隊員數目、作品語言等）均不在考慮之列，而上述的分析亦與評審的實際

做法相當一致。相反，評審時所考慮的創意、統計概念運用能力、表達技巧等因素卻甚難於行政數據上反映。因此，同學不論級別、曾否參賽、隊員數目、作品語言，均歡迎積極參與來屆的「**中學生統計創意寫作比賽**」。

邀請作品：

統計的學與教

暑假期間，本人整理物件，無意中發現一篇三十年前由數學教育界同工所寫的文章，提及一群教育學者對當時香港中學統計學教育的一些觀點，簡述如下:

1. 中學的統計學課程只局限於描述統計學的範疇，而絲毫不涉及統計推理，所以內容較易為學生明白及掌握

2. 教材與學生日常生活沒有密切關係

3. 教授方法過於刻板

4. 教學重點錯誤地放在計算技巧而不在概念上

5. 統計學中的非數學部分每被忽略

教育界同工建議應鼓勵學生多參與統計專題的探討和研究，克服以上的缺點。早於1986 年，香港統計學會首次舉辦的《中學生統計習作比賽》，至今已舉辦三十一屆。學生參與

統計習作需要把各科的知識結合來運用，以研究和解決生活上的實際問題，十分有教育意義，值得推廣。香港統計學會成員於這數十年間為統計知識普及化及統計教育方面下了不少功夫，值得讚許。

三十年過去了，上述文章內所提及的香港中學統計教育狀況是否還存在呢? 大多數同學及教師是否仍把統計學視為較次要的項目? 數學教師本身對統計學的認識又有否加深? 羅馬非一日建成，樂觀地看，我們朝著好的方面發展。中學數學課本已加入了不少與統計學相關的生活化示例，當中包括統計的應用及誤用。同時，新進的數學教師們亦於大學時選修了一定的統計課程，統計知識增加了不少。

我們應怎樣向學生教導統計知識較好呢? 有沒有一個正確的方法? 根據統計教育(GAISE)報告綱要的建議，教學應強化學生的統計素養和發展他們的統計思維能力。統計教學應採用真實的數據、強調知識的理解而不只是教授運算程序。統計教學可運用科技幫助學生理解和分析數據，培養學生學習的興趣。這些建議十分值得我們深思。隨著互聯網的普及，科技的發展，數據產生和收集變得容易，教學資源亦豐富了很多。單以香港的網站為例， 政府統計處的網站內(https://www.censtatd.gov.hk/hkstat/sub/sc150_tc.jsp)有很多香

港的統計資料供我們使用。同時，政府統計處的人員亦於2016年完成中期人口統計，並把很多有用的資料放於網頁(http://www.bycensus2016.gov.hk/tc/bc-index.html)，包括主要統計表、刊物和互動圖解。教師可以很容易採用這些資料於課堂內，令同學們明白統計與他們的生活息息相關。

很多公司負責人於傳媒的訪問中表示他們透過數據分析、預測，令決策更精確。同學們可能對整個決策過程感興趣。跟著，我嘗試舉一個現實情境中的例子以說明如何運用中學的統計知識幫助我們決策。

我相信大部分同工均有電郵戶口，每日無可避免會接收一定數量的垃圾電郵。垃圾電郵是指大批寄出、內容相同、不請自來的郵件。若要每一封電郵均開啟檢查，相當費時，有沒有想過電子郵件過濾器如何把它們找出來，原來是經由統計的數據，從字詞的出現，推算一份郵件是垃圾郵件的概率，做法如以下程序所述:

1. 先找出一定數量的已知的正常郵件和垃圾郵件，例如，各5000份。數量越多，效果越佳。

2. 數一個字詞在正常郵件出現的次數，和在垃圾郵件所出

現的次數，例如，「免費」、「中獎」、「成人」和「匯款」等詞。

3. 如果所選取的字出現的次數過少，例如五次，我們便不以這個字作參考。

4. 假設垃圾郵件、正常郵件中的字詞出現概率如下:

字詞	垃圾郵件	正常郵件
免費	0.88	0.23
中獎	0.97	0.11
成人	0.92	0.27
匯款	0.85	0.05

當我們收到一封電郵，若充滿“嫌疑度”較高的字詞，這是一封垃圾電郵的概率越高。

5. 電子郵件過濾器又會如何做呢? 當電腦收到一封電子郵件，它會根據先前收集字詞的數據，估計假如這個字詞在一份電子郵件裡出現的時候，這封郵件是垃圾郵件的概率，這個概率稱為「罪証概率」，我們可以試用 p 代表它。從以上的數據為例，若電郵出現「中獎」這字詞，它是垃圾郵件的概率為多少呢?

我們運用以下公式計算字詞的罪証概率:

$$\frac{\text{這個字詞在垃圾郵件出現的概率}}{\text{這個字詞在垃圾郵件出現的概率+這個字詞在正常郵件出現的概率}}$$

以「中獎」這個字詞為例，其罪証概率$=\frac{0.97}{0.97+0.11}=0.90$

根據以上的方法，我們便建立了一個字詞的罪証資料數據庫。當一封電郵傳送過來，我們先選出當中出現最多的字詞，數量假設為10個，我們從數據庫把這10個字詞的罪証概率找出來，分別以p_1 ，p_2,……,p_{10}表示之。若所選的字詞並沒有出現在罪証概率的數據庫，我們可假設這個字詞的罪証概率為 0.4 或 0.5。這個假設亦很算合理吧。根據以下公式，我們便能找出這封郵件是垃圾郵件的可能性 p:

$$p=\frac{p_1\times p_2\times p_3\times p_4\ldots\times p_{10}}{p_1\times p_2\times p_3\times p_4\ldots\times p_{10}+(1-p_1)\times(1-p_2)\times(1-p_3)\times(1-p_4.)..\times(1-p_{10})}$$

這可以看成是集體的罪証概率，如果電子郵件集體的罪証概率大於預先的設定值，例如 0.8，電腦就自動判斷它是垃圾電子郵件。

運用以上的公式時，同學們可留意我們假設各字詞的出現是互相獨立。同時，罪証概率數據庫隨著新的資料而調整，當我們收到電子郵件時，回覆為正常郵件或垃圾郵件，字詞的

罪証概率數據庫也會更新、完善。同學們可發現程序運用了「貝葉斯定理」。適當於統計課堂內注入生活中的概率趣事能增強統計教學的效能，教師們可多作嘗試。

參考資料:

1. ICOTS 2, 1986: P.H. Cheung, K. Lam, M.K. Siu and N.Y.Wong

2. 劉炯朗 (2017)。 數學的魔法:生活中無處不在的數學智慧。北京:團結出版社。

邀請作品：

Probability and Prime Numbers

The short article is about an interesting probability puzzle taken from the lovely little book *A View from the Top* by *Alex Iosevich*[1]. It may be a bit difficult for secondary school students, but be courageous and try to follow the heuristic arguments below as much as possible, and forget about any technical details related to any limiting procedures.

The puzzle is like this:
Suppose that we pick two positive integers randomly, how likely is it that they are relatively prime? Do you think it is greater than 1/2, or less than 1/2?

Let's understand what the question means, and solve the puzzle step-by-step.

(1) Two positive integers are relatively prime if they do not have

[1] *Iosevich, A.*, 2007. *A View from the Top*, Student Mathematical Library Volume 39, American Math-ematical Society.

any common prime factors. For instance, 18 and 37 are relatively prime, 20 and 27 are relatively prime.

(2) How about the meaning of "pick two positive integers randomly"? How to understand this? We may imagine that a very very large positive integer N (say N = 1, 000, 000, 000, …, 000, with many many zeros) is fixed. Then we pick an integer from the list 1, 2,…, N randomly, each integer being picked with a chance of $1/N$. Then we let $N \to \infty$ later on.

(3) What is the chance that a randomly picked integer from the list 1,2,…, N is divisible by 2 (that is, a multiple of 2)? If N is an even number, the chance is $(N/2)/N=1/2$;

if N is an odd number, then there are $(N-1)=2$ even numbers in the list, and so the chance is $((N-1)/2)/N=1/2-1/2N \approx 1/2$ (when N is very very large, $1/2N$ is very very small).

By the same argument, that a randomly picked integer from the list 1,2,…,N is divisible by any prime number p is approximately $1/p$.

(4) What is the chance that both randomly picked integers are divisible by a prime p? Assume that the choice of two integers are independent, then chance is $(1/p)(1/p)=1/p^2$

(5) From (4), the chance that at least one of the two randomly picked integers is not divisible by p is $1-1/p^2$

(6) Two positive integers are relatively prime if for every prime number p, at least one of the them is not divisible by p. Therefore, the chance that two randomly picked integers are relatively prime equals

$$\left(1-\frac{1}{2^2}\right)\left(1-\frac{1}{3^2}\right)\left(1-\frac{1}{5^2}\right)\left(1-\frac{1}{7^2}\right)\left(1-\frac{1}{11^2}\right)\ldots,$$

in which the product should include all prime numbers less than N. When $N \to \infty$, it becomes an infinite product, including all prime numbers. Now the problem is: can we estimate the value of this infinite product?

(7) A typical factor in the above product takes the form $1-p^{-2}$. Its reciprocal equals

$$\frac{1}{1-p^{-2}}=1+(p^{-2})^1+(p^{-2})^2+(p^{-2})^3+\cdots\sum_{k=0}^{\infty}p^{-2k}$$

by the standard geometric series formula. Therefore, our answer to the puzzle is the reciprocal of

$L=(\sum_{k=0}^{\infty} 2^{-2k})(\sum_{k=0}^{\infty} 3^{-2k})(\sum_{k=0}^{\infty} 5^{-2k})(\sum_{k=0}^{\infty} 7^{-2k})(\sum_{k=0}^{\infty} 11^{-2k})$

…

(8) What is the value of L? In fact,

$$L=\frac{1}{1^2}+\frac{1}{2^2}+\frac{1}{3^2}+\cdots=\sum_{k=0}^{\infty}\frac{1}{n^2}$$

To see why this is true, expand, as a little experiment, the product $(\sum_{k=0}^{2} 2^{-2k})(\sum_{k=0}^{3} 3^{-2k})(\sum_{k=0}^{2} 5^{-2k})$ and see what it looks like! Here, you also need to know the basic fact that every positive integer can be factorised into a product of prime numbers in a unique way (e.g.$18=2\times3^2$, $36=2^2\times3^2$).

(9) Putting all these together, we now know that the answer equals $1/L$

Is it larger than 1/2? Let's do some simple estimation:

$$L=\sum_{n=1}^{\infty}\frac{1}{n^2}=1+\sum_{n=2}^{\infty}\frac{1}{n^2}$$

$$\leq\sum_{n=2}^{\infty}\frac{1}{n(n-1)}$$

$$= \sum_{n=2}^{\infty} \left(\frac{1}{n\text{-}1} - \frac{1}{n} \right)$$

$$= 1 + \left(\frac{1}{1} - \frac{1}{2} \right) + \left(\frac{1}{2} - \frac{1}{3} \right) + \left(\frac{1}{3} - \frac{1}{4} \right) + \left(\frac{1}{4} - \frac{1}{4} \right) + \ldots = 2,$$

so the answer is greater than 1/2 !

Exercise 1: Can you provide a better lower bound? (Hint: mimic the last step by including more terms of the summation before applying the telescoping technique.)

Exercise 2: Google search the term \Riemann zeta function", and see how it relates to the summation$= \sum_{n=1}^{\infty} \frac{1}{n^2}$

邀請作品：

Bayesian Criminal Tracking

Stephen .M.S. Lee

Department of Statistics and Actuarial Science, the University of Hong Kong

1. Introduction

Broadly speaking, statistical inference refers to the process of inferring knowledge about some unknown parameter θ, taken to represent the true nature, from a sample of data $(Z_1,\cdots,Z_n)$ observed in real life. Traditionally there have been two main approaches of inference, namely the frequentist and the Bayesian, each having its own pros and cons when applied to real-life problems.

In this article we focus on a particular problem to which the Bayesian approach has found an especially appealing application. Consider a perpetrator who has continually committed crimes at different locations of a region and is still at large. The police desperately want to track him down, and should obviously do it as soon as possible lest the perpetrator commits a new crime. At

the beginning when the perpetrator had just committed the first crime, the police's job was the most difficult due to inadequate information. When more and more crimes had been committed by the same perpetrator, more and more information would have been available to the police, making their job easier. Surely the police cannot just sit there, waiting for a sufficient amount of information before taking any action. In other words, whenever a new crime has been committed by the perpetrator somewhere in the region, the police need to act immediately upon this new piece of information to revise their current action plan. It is in such a desperate situation as faced by the police where the Bayesian approach provides a very useful tool for planning fast and effective actions.

We define from now on $\theta=(\theta_1,\theta_2)$ to be the coordinates of the perpetrator's home base, unknown to the police. For each $i=1, 2,\cdots,$ the observation $Z_i=(X_i,Y_i)$ denotes the coordinates of the location of the ith crime committed by the perpetrator. By the time when n crimes have been committed by the perpetrator, the police's immediate goal is to infer the most likely value of θ based on the available observations $(Z_1,\cdots,Z_n)$.

2. Modelling

Before we start our Bayesian journey, two important ingredients need to be made ready: (1) a model depicting the attack behaviour of the perpetrator, and (2) a model depicting the police's initial opinion about the unknown parameters that characterise the model (1) proposed above.

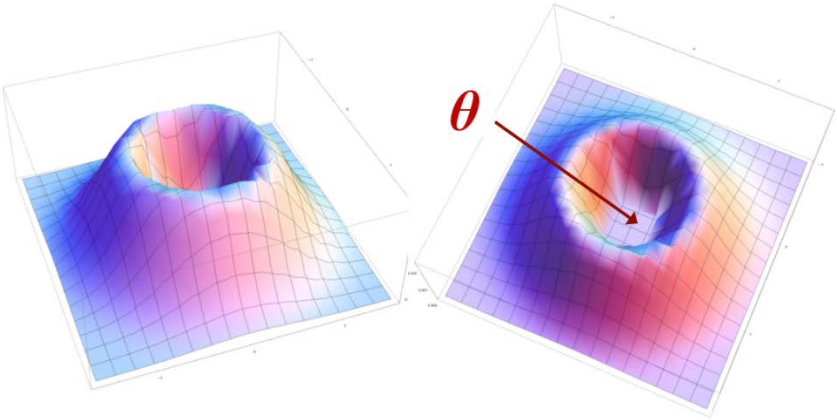
2.1 Perpetrator's attack behavior

For (1), we base construction of our model on two premises which reflect the perpetrator's psychology. First, we assume that he never launches attacks too close to his home base, certainly not within a radius R of it. Secondly, he prefers launching attacks at near locations to distant locations, but has no preference for any particular direction. With these two premises in mind, we propose a model for the perpetrator's attack location $Z=(X,Y)$ given by the *density function*

$$f(z/\theta,R)=f(x,y/\theta,R)$$

$$=\begin{cases}\dfrac{0.2}{2\pi}\left(1-\dfrac{\mathrm{R}}{\sqrt{(x-\theta_1)^2+(y-\theta_2)^2}}\right)e^{-0.1\left(\sqrt{(x-\theta_1)^2+(y-\theta_2)^2}-R\right)^2}, & (x-\theta_1)^2+(y-\theta_2)^2>R^2\\ 0, & (x-\theta_1)^2+(y-\theta_2)^2\leq R^2,\end{cases}$$

which is plotted below in 3D:



Briefly speaking, the function $f(z/\theta,R)$ indicates how likely the location z is to be chosen by the perpetrator for his attack. It is clear that the most dangerous area lies in a ring at some fixed radius around the perpetrator's home base, as stipulated by the premises. The danger recedes when one moves away from the ring in either direction.

2.2 Police's prior belief

At the beginning, before the perpetrator commits the very first crime, the police have no firm idea above the true values of θ and R. They believe only that θ must lie within a square with its opposite corners defined by the coordinates (-10,-10) and (10,10), and that R can be any distance not exceeding 10. The above prior belief of the police can be conveniently depicted by

the *prior density function*

$$\pi(\theta,R)=\begin{cases}\frac{1}{4000}, & (\theta,R)\in[-10,10]\times[-10,10]\times[0,10]\\ 0, & (\theta,R)\notin[-10,10]\times[-10,10]\times[0,10].\end{cases}$$

The constant level $1/4000$ simply says that the police initially think that any point within the square [-10,10]×[-10,10] is equally probable to be the perpetrator's home base, and any number between 0 and 10 is equally probable to be the true radius R.

3. Bayesian tracking

We may now apply the Bayesian method to help the police track down the perpetrator's home base after each new attack. After the first attack at the location $z_1=(x_1,y_1)$, we *update* the police's initial belief, represented by the prior density $\pi(\theta,R)$, to a *posterior* density $\pi(\theta,R|z_1)$ using a standard formula akin to the Bayes' Theorem:

$$\pi(\theta,R|z_1)\propto\pi(\theta,R)f(z_1|\theta,R)$$

$$\propto\left(1\frac{R}{\sqrt{(x_1-\theta_1)^2+(y_1-\theta_2)^2}}\right)e^{-0.1\left(\sqrt{(x_1-\theta_1)^2+(y_1-\theta_2)^2}-R\right)^2}$$

for $|\theta_1|,|\theta_2|\leq 10$ and $R<min\{\sqrt{(x_1-\theta_1)^2+(y_1-\theta_2)^2},10\}$. Assuming that the locations of the perpetrator's attacks are

independently selected, we can update the above posterior density iteratively after each new attack, so that after the nth attack, the updated posterior density becomes

$$\pi(\theta, R|z_1, \cdots, z_n) \propto \pi(\theta, R) f(z_1|\theta, R) \cdots f(z_n|\theta, R)$$

$$\propto \prod_{i=1}^{n} \left(1 - \frac{R}{\sqrt{(x_i - \theta_1)^2 + (y_i - \theta_2)^2}}\right) e^{-0.1\left(\sqrt{(x_i-\theta_1)^2+(y_i-\theta_2)^2}-R\right)^2}$$

for $|\theta_1|, |\theta_2| \leq 10$ and $R < \min\left\{10, \min_{1\leq i\leq n}\left\{\sqrt{(x_i - \theta_1)^2 + (y_i - \theta_2)^2}\right\}\right\}$. Since the police are primarily interested in the home base θ rather than the radius R, we focus attention on θ in $\pi(\theta, R|z_1, \cdots, z_n)$ by integrating out the irrelevant parameter R, that is

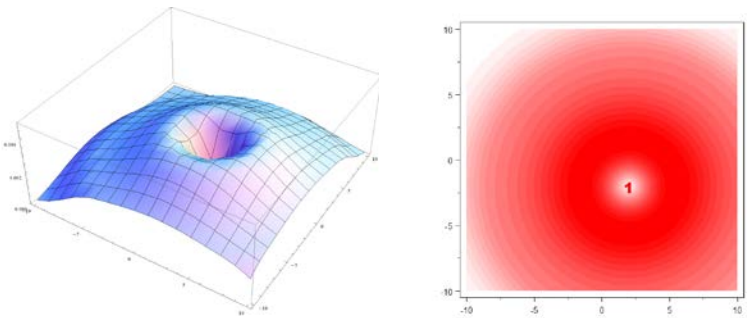
$$\pi(\theta|z_1, \cdots, z_n) = \int_0^\infty \pi(\theta, R|z_1, \cdots, z_n) dR$$

$$\propto \int_0^{min\left\{10, \min_{1\leq i\leq n}\left\{\sqrt{(x_i-\theta_1)^2+(y_i-\theta_2)^2}\right\}\right\}} \prod_{i=1}^{n} \left(1 - \frac{R}{\sqrt{(x_i - \theta_1)^2 + (y_i - \theta_2)^2}}\right) e^{-0.1\left(\sqrt{(x_i-\theta_1)^2+(y_i-\theta_2)^2}-R\right)^2} dR.$$

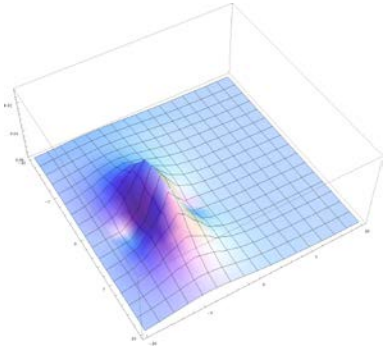
With the help of $\pi(\theta|z_1,\cdots,z_n)$, which is a function of θ alone, the police can identify the most likely location of the perpetrator's home base with the value of θ that maximises $\pi(\theta|z_1,\cdots,z_n)$, and revise their action plan accordingly.

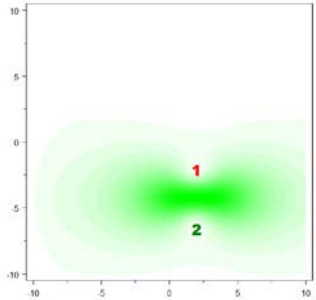
4. A real-life example

We illustrate our Bayesian approach with a real-life example. The incident began on 13 December, 2008, when an unidentified perpetrator threw a plastic bottle filled with corrosive liquid onto pedestrians in the Mongkok area, at the intersection between Sai Yeung Choi St. S. (西洋菜南街) and Shantung St. (山東街). The following shows a 3D and contour plots of $\pi(\theta|z_1)$. The label "1" indicates the location of the first attack. The darker regions in the contour plot indicate higher levels of $\pi(\theta|z_1)$, hence the more probable locations of the perpetrator's home base as inferred by the Bayesian method after observing one attack.

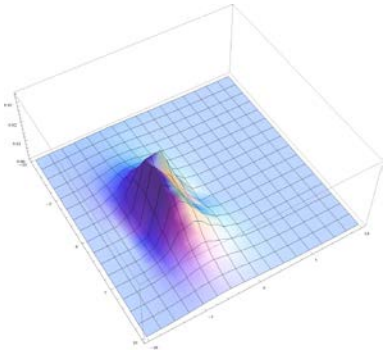


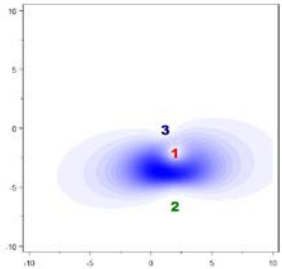
On 16 May, 2009, the perpetrator launched the second acid attack at the intersection between Sai Yeung Choi St. S. (西洋菜南街) and Soy St. (豉油街), which leads to a revision of the posterior density to $\pi(\theta|z_1,z_2)$:



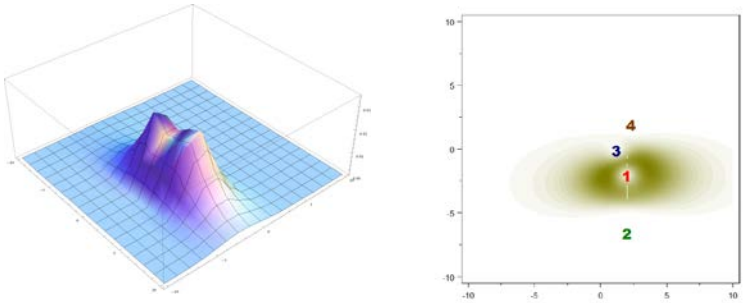


The third attack came on 8 June, 2009, at the intersection between Sai Yeung Choi St. (西洋菜街) and Nelson St. (奶路臣街), leading to yet another revision of the posterior to $\pi(\theta|z_1,z_2,z_3)$:

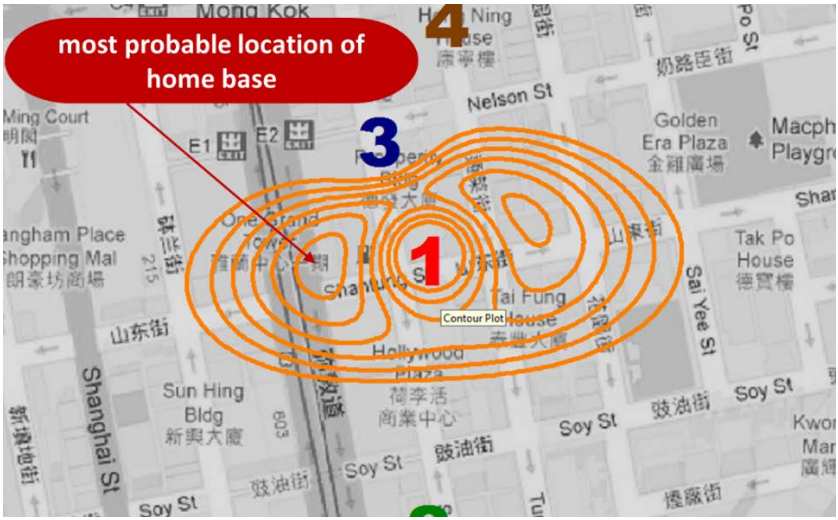




After the fourth attack, which occurred on 6 September, 2009 near Tung Choi St. (通菜街), the posterior density becomes $\pi(\theta|z_1,z_2,z_3,z_4)$:



By this time the search area of the police may be confined to a more restrictive zone indicated by high levels of $\pi(\theta|z_1,z_2,z_3,z_4)$. We see from the map below that $\pi(\theta|z_1,z_2,z_3,z_4)$ has its peak at the intersection between Shantung St. (山東街) and Nathan Rd. (彌敦道), which is suggested by our Bayesian method to be the most probable location of the perpetrator's home base after he had launched four attacks.



5. Conclusion

The Bayesian approach provides a powerful tool for making statistical inference, especially under circumstances when actions need to be revised and taken promptly as more and more information accumulate. We have illustrated an important application of the Bayesian approach with an example of criminal tracking. Similar applications can also be found to a wide range of problems, not least important of which is the search for a missing flight such as the recent Malaysia Airlines Flight MH370. For further examples of Bayesian inference, see McGrayne (2011).

Reference

1. McGrayne, S.B. (2011). The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy. Yale University Press.

二零一六至一七年度中學生統計創意寫作比賽的籌備委員會：

主席	楊良河博士，香港大學統計及精算學系
總評審主任	張家俊博士，香港大學統計及精算學系
籌委會成員	陳秀騰先生，教育局
	杜思縉先生，政府統計處
	袁耀波先生，政府統計處
	林曉峯先生，政府統計處
	李妍慧女士，政府統計處

數學百子櫃系列	作者
(一) 漫談數學學與教—新高中數學課程必修部分	張家麟、黃毅英、韓藝詩
(二) 漫談數學學與教新高中數學課程延伸部分單元一	韓藝詩、黃毅英、張家麟
(三) 漫談數學學與教新高中數學課程延伸部分單元二	黃毅英、張家麟、韓藝詩
(四) 談天說地話數學	梁子傑
(五) 數學的應用: 區像處理—矩陣世紀	陳漢夫
(六) 數學的應用: 投資組合及市場效率	楊良河
(七) 數學的應用: 基因及蛋白的分析	徐國榮
(八) 概率萬花筒	蕭文強、林建
(九) 數學中年漢的自述	劉松基
(十) 中學生統計創意寫作比賽 2009 作品集	
(十一) 從「微積分簡介」看數學觀與數學教學觀	張家麟、黃毅英
(十二) 2010/11 中學生統計創意寫作比賽作品集	
(十三) 2011/12 中學生統計創意寫作比賽作品集	
(十四) 數學教師不怕被學生難倒了! — 中小學數學教師所需的數學知識	黃毅英、張僑平
(十五) 2012/13 中學生統計創意寫作比賽作品集	
(十六) 尺規作圖實例、題解和證明	孔德偉
(十七) 摺紙與數學	阮華剛、譚志良
(十八) 2013/14 中學生統計創意寫作比賽作品集	
(十九) 2014/15 中學生統計創意寫作比賽作品集	
(二十) 宇宙的尺度變異定律	龍振強

(二十一) 三次數學危機與勇闖無窮大 梁子傑

(二十二) 2015/16 中學生統計創意寫作比賽作品集