ISBN 978-988-8370-88-7
9 789888 370887
政府物流服務署印

數學百子櫃系列(二十六) 2018/19 中學生統計創意寫作比賽 作品集

數學百子櫃系列 (二十六)

2018/19 中學生統計創意寫作比賽 作品集

教育局
課程發展處數學教育組

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

數學百子櫃系列 (二十六)

2018/19 中學生統計創意寫作比賽作品集

教育局
課程發展處數學教育組

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

ISBN 978-988-8370-88-7

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《2018/19 中學生統計創意寫作比賽作品集》，是這個系列的第二十六冊。本書輯錄的文章，大部分是「2018/19 中學生統計創意寫作比賽」的優勝作品，由參賽的中學生撰寫。

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

本書共有 14 篇文章，第 1 至 10 篇為「2018/19 中學生統計創意寫作比賽」的冠軍、亞軍、季軍和優異作品。其餘 4 篇則為邀請作品，分別由政府統計處的統計師，香港大學統計及精算學系的教授和中學教師撰寫，供讀者們閱覽。本書的文章充滿趣味，期望讀者閱後能獲得啟發、不僅增加統計的知識，還能善用統計決策、解難。

此書得以順利出版，實有賴這次比賽的籌備委員會成員所付出的努力。在此，謹向撰寫作品的得獎隊伍、政府統計處的統計師、香港大學精算及統計學系的教授和朱吉樑老師致以衷心的

感謝。最後，更要多謝這次比賽的籌備委員會主席楊良河博士和總評審主任張家俊博士。兩位鼎力協助，審訂本書的內容，讓學生能夠閱讀更多有趣的文章，增加他們學習統計的興趣。

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

九龍油麻地彌敦道 405 號九龍政府合署 4 樓
教育局課程發展處
總課程發展主任(數學)收
(傳真: 3426 9265 電郵: ccdoma@edb.gov.hk)
教育局課程發展處
數學教育組

前言

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

中學生統計創意寫作比賽已舉辦了十年,是時候停下來作出檢討。回顧過去的參賽作品,我們看到同學對統計概念的認識深入了,並能正確地運用統計知識作解說。得獎作品的整體質素亦有所提升。本年度的比賽專題是「運動中的統計」,與同學的生活較相關,同學很容易便能找到題材,創作故事。本屆參賽作品約 70 份,數量與上屆相若。文章取材創新,趣味盎然;同學活用各種統計知識創作故事,分析條理分明,解說清晰,值得欣喜和嘉許。本書輯錄了本屆所有的得獎作品,藉此嘉許得獎同學所付出的努力,並希望同學能夠從創作或閱讀這些得獎作品中得到啟發,對統計知識有更深入認識。

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

2019 年 11 月 27 日

目錄

冠軍作品:

是否贏在起跑線才能贏到最後之

名牌幼稚園的迷思

學校名稱：保良局何蔭棠中學

學生姓名：黃穎璇，布嘉俐，曾淑瑜

指導老師：陳智仁

摘要：

「名校出狀元」等的報導，不斷鞏固了「要成功便要入讀名大學，要入讀名大學便要入讀名中學，要入讀中學便要入讀名小學，要入讀名小學便要入讀名幼稚園」的層層疊迷思，所以不少家長為了催谷年幼的子女「贏在起跑線上」，不但爭相讓子女入讀名校幼稚園，更為他們安排各種興趣班。然而國外多項研究卻發現，較遲入學的學生之學術表現比早入學好，即「贏在起跑線」並無數據支持，甚或會令子女有更多壓力。究竟這當中孰真孰假，入讀名牌幼稚園是否入讀名小學的入場券，又或者是以訛傳訛的美麗誤會呢？因此，本篇將會透過統計和概率分析，探究是否要贏在起跑線入讀某某名牌幼稚園，才能贏到最後。

今年文憑試的結果出爐了，穎璇在社交網站上看狀元究竟花落誰家，卻感嘆地說：你們看，今年的狀元又全是出自傳統名校，果然贏在起跑線這句話是沒有說錯的，他們的家長從小為他們鋪路，讓他們入讀名牌幼稚園和報讀各種的興趣班，讓他們能夠一條龍直升名牌小學和中學。假如我們的父母以前有好好地讓我們贏在起跑線，也許狀元的名單上也可

▲ DSE 2018 放榜，今屆誕生9名狀元。

DSE 2018放榜，今年共有9名7科5**DSE狀元，6男3女狀元們分別誕生於6間傳統名校包括女拔、男拔、聖保羅男女、德望、皇仁，以及喇沙，當中DSE首名奪9科5**的18星終極超級男狀元亦是來自喇沙。

能會有我們的一席之位吧！

淑瑜卻以另一個網頁的內容反駁說：我才不認為要成功必須要贏在起跑線，你看看大部分狀元也不是出自名牌幼稚園的，所以我們相信也未必差過那些贏在起跑線的同學。正當二人在激烈討論時，作為數學學霸的嘉俐開口了: 爭論是不會有結果的，我們何不去計算分析？要計算兩者的關係，我們課堂上學過的概率、樹形圖等，都是可以找出結論的。

名牌幼稚園名稱	直升其小學的百分率	直升其直屬聯繫中學的百分率	升讀大學百分率
低主教幼稚園	55%	50%	35.9%
香港假光中學幼稚園堅道	70%	72%	60%
香港假光中學幼稚園	90%		
聖減勒幼稚園	50%	49%	65%
聖保綠學校幼稚園*spk*	80%	90%	71%
聖保綠學校幼兒園*spn*	90%		

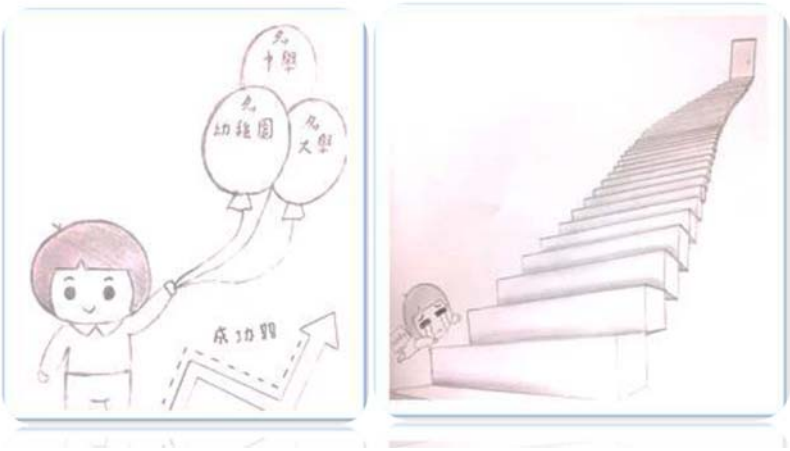
名牌幼稚園名稱	直升其小學的百分率	直升其直屬聯繫中學的百分率	升讀大學百分率
紋身書院幼稚園部	77%	55%	74%
培根小學附屬幼稚園	100%	80%	85%
協音小學附屬幼稚園	75%	80%	69.7%

很快，穎璇就從網上找到了一大堆名牌幼稚園升學率的資料，但由於資料數量的龐大，正煩惱著如何處理。然後嘉俐就建議說：要量度統計大量的數據，我們可以運用算術平均數，即是數學堂所學的 arithmetic mean

$$\overline{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{\sum_{i=1}^{n} x_i}{n}$$

名牌幼稚園名稱	直升其小學的百分率	直升其直屬聯繫中學的百分率	升讀大學百分率
平均率	76.3%	68%	65.8%

（註：假設每間名牌幼稚園的學生人數相等）



研究模型(1a)名牌幼稚園學生升大學的概率計算

(假設他們會由名牌幼稚園直升直屬小學，再直升到直屬中學)

名牌幼稚園　　直屬小學　　直屬中學

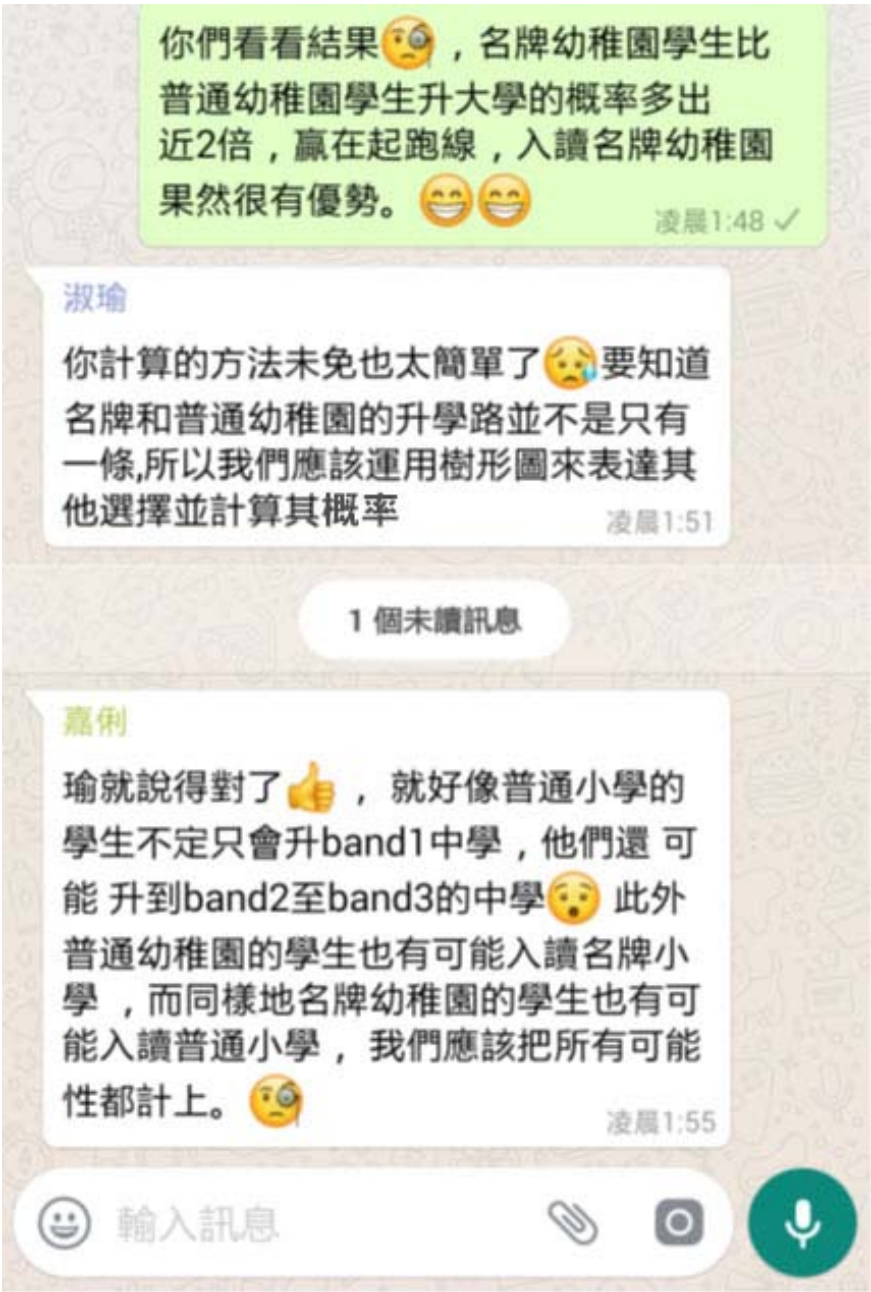
名牌幼稚園學生升大學的概率

76.3% x 68% x 65.8% = 34.14%

研究模型(1b)普通幼稚園學生升大學的概率計算

(假設他們由普通幼稚園升至普通小學再升到 Band1 中學)

階段	升讀百分率	解釋
幼稚園升小學	100%	全港的津貼小學，均以「小一入學統籌辦法」收生，「統一派位」會用隨機編號來分配學位，所以每個人入普通津貼小學的機率都是相同的。
小學升Band1中學	1/3 ≈ 33.3%	根據香港津貼中學的「中學學位分配辦法」，主要根據小六生自己的Banding及選校意願，加上隨機編號來分配中學學位，每個組別佔全港學生的三分之一。
Band1中學升大學	54.78%	
普通幼稚園學生升大學的概率	100% × 33.3% × 54.78% = 18.24%	

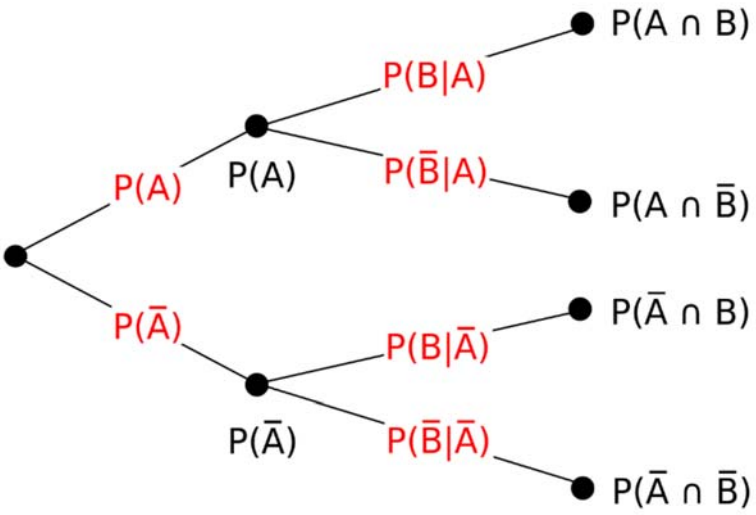

你們看看結果，名牌幼稚園學生比普通幼稚園學生升大學的概率多出近2倍，贏在起跑線，入讀名牌幼稚園果然很有優勢。
凌晨1:48
淑瑜
你計算的方法未免也太簡單了要知道名牌和普通幼稚園的升學路並不是只有一條,所以我們應該運用樹形圖來表達其他選擇並計算其概率
凌晨1:51
1 個未讀訊息
嘉俐
瑜就說得對了，就好像普通小學的學生不定只會升band1中學，他們還 可能 升到band2至band3的中學 此外普通幼稚園的學生也有可能入讀名牌小學 ，而同樣地名牌幼稚園的學生也有可能入讀普通小學， 我們應該把所有可能性都計上。
凌晨1:55
輸入訊息

條件概率（英語：conditional probability）就是事件A在另外一個事件B已經發生條件下的發生概率。條件概率表示為$P\left(A\middle|B\right)$，讀作「在B條件下A的概率」。

設 A 與 B 為樣本空間 Ω 中的兩個事件，其中 $P(B)>0$。那麼在事件 B 發生的條件下，事件 A 發生的條件概率為：

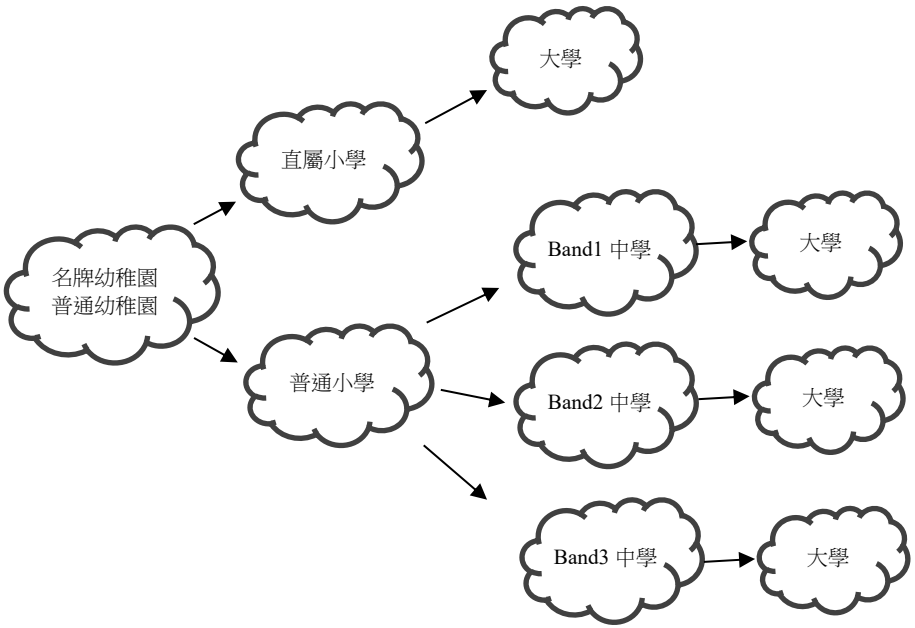
$$P\left(A\middle|B\right)=\frac{P\left(A\cap B\right)}{P\left(B\right)}$$

在概率論中，樹形圖（Tree Diagram）是用來表示一個概率空間。



	Band 1 中學	Band 2 中學	Band 3 中學
小學升中學的概率	33.3%	33.3%	33.3%
中學升大學的概率	54.78%	17.44%	5.424%
普通小學學生升大學的概率	33.3%×54.78% = 18.24%	33.3%×17.44% = 5.81%	33.3%×5.424% = 1.81%

研究模型(2)　名牌幼稚園和普通幼稚園學生升大學的樹形圖



研究模型(2a)　名牌幼稚園學生升大學的概率計算

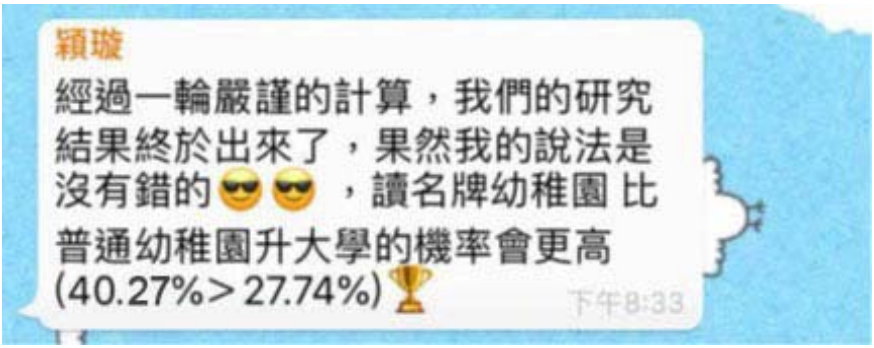
	直屬小學	普通小學
名牌幼稚園升小學的概率	76.3%	23.7%
小學學生升大學的概率	68% × 65.8% = 44.74%	18.24%+5.81%+1.81% = 25.86%
名牌幼稚園經不同中學升大學的概率	76.3% × 44.74% = 34.14%	23.7% × 25.86% = 6.13%
名牌幼稚園學生升大學的概率	34.14% + 6.13% = 40.27%	

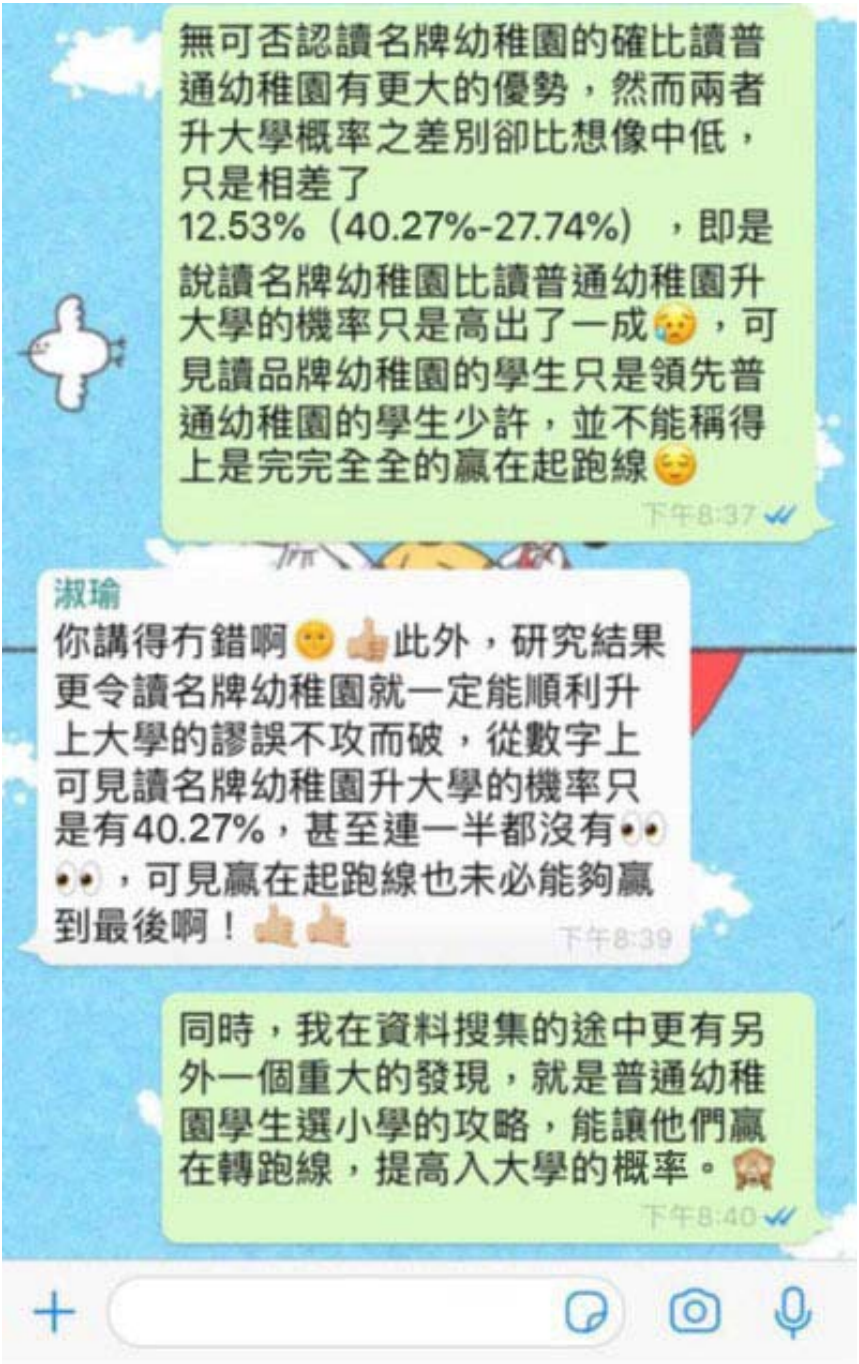
研究模型(2b)　普通幼稚園學生升大學的概率計算

(註:名牌幼稚園的直屬小學均屬於直資和私立。)

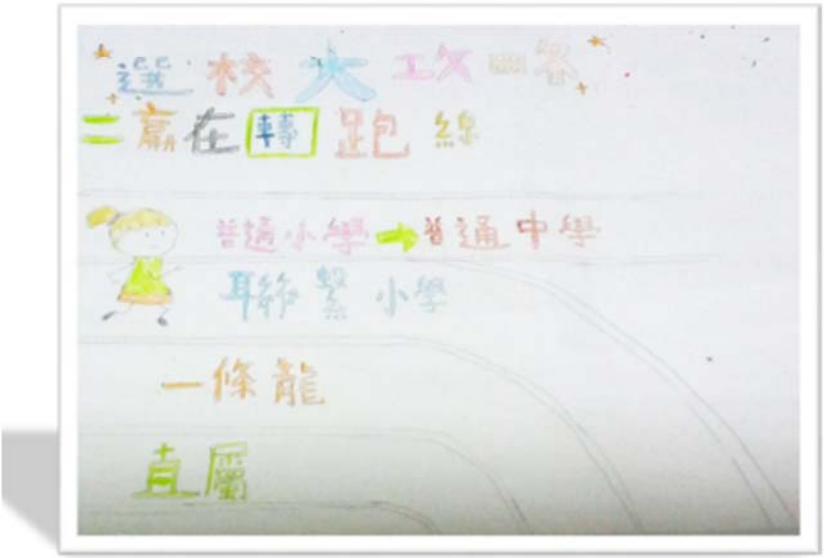
研究模型(2b) 普通幼稚園學生升大學的概率計算

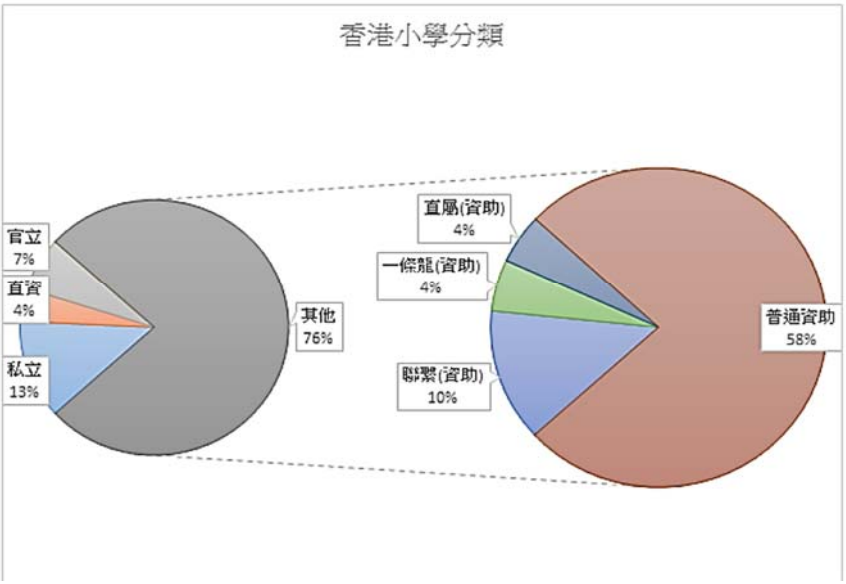
	直屬小學	普通小學
普通幼稚園升小學的概率	8%+2%=10%	100% – 10%=90%
小學升大學的概率	68%×65.8% =44.74%	18.24%+5.81%+1.81% = 25.86%
幼稚園經不同小學升大學的概率	10%×44.74% =4.47%	90%×25.86% =23.27%
普通幼稚園學生升大學的概率	4.47%+23.27% =27.74%	




無可否認讀名牌幼稚園的確比讀普通幼稚園有更大的優勢，然而兩者升大學概率之差別卻比想像中低，只是相差了12.53%（40.27%-27.74%），即是說讀名牌幼稚園比讀普通幼稚園升大學的機率只是高出了一成，可見讀品牌幼稚園的學生只是領先普通幼稚園的學生少許，並不能稱得上是完完全全的贏在起跑線
下午8:37
淑瑜
你講得冇錯啊 此外，研究結果更令讀名牌幼稚園就一定能順利升上大學的謬誤不攻而破，從數字上可見讀名牌幼稚園升大學的機率只是有40.27%，甚至連一半都沒有，可見贏在起跑線也未必能夠贏到最後啊！
下午8:39
同時，我在資料搜集的途中更有另外一個重大的發現，就是普通幼稚園學生選小學的攻略，能讓他們贏在轉跑線，提高入大學的概率。
下午8:40


教子女如何靠「聯繫」入Band1英中
聖公會林護紀念中學
(學額：21個)
浸信會呂明才中學
(學額：26個)
保良局百周年李兆忠
紀念中學(學額：21個)
聖公會仁立紀念小學
聖公會仁立小學
浸信會呂明才小學
浸信會沙田圍
呂明才小學
保良局莊啟程第二小學
保良局梁周順琴小學


選校大攻略
二扁在轉跑線
普通小學
普通中學
聯繫小學
一條龍
直屬



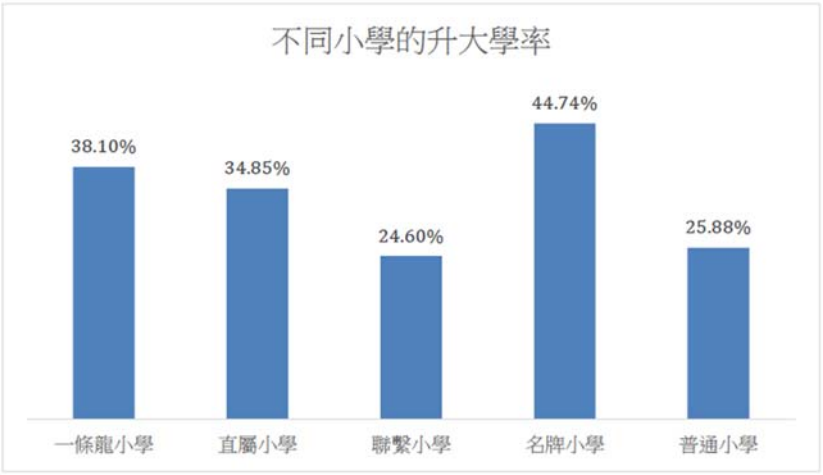
要「贏在起跑線」，未必一定要入讀直資或私立小學。教育局制度下不少名牌英文中學屬於津貼聯繫學校、一條龍、直屬學校。入讀這類小學，就更易升讀有聯繫的中學。

小學類別	定義	幼稚園升小學率
一條龍中小學	一條龍學校，大意是指小學和中學結合成「一條龍」，「龍」內的小六畢業生毋須考試或參加升中派位，便可全數直升「龍」內的中學。	$\frac{19}{537} = 3.56\%$
直屬中小學	直屬制度是整個統一派位制度的一部分。根據教育局制定的中學學位分配辦法，直屬中學經校方扣除重讀生及自行分配學位數目後，可最多保留餘額 85% 學位予其直屬小學的學生。	$\frac{20}{537} = 3.72\%$

小學類別	定義	幼稚園升小學率
聯繫中小學	聯繫制度是整個統一派位的一部分。根據教育局制定的中學學位分配辦法，聯繫中學經校方扣除重讀生及自行分配學位後，可最多保留餘額 25%學位予其聯繫小學的學生。	$\frac{51}{537} = 9.50\%$

研究(3) 對比不同小學的升大學率

小學類別	小學升中學率	中學升大學率	小學升大學率
一條龍中小學	100%	38.1%	100% × 38.1% =38.1%
直屬中小學	85%	41.0%	85% × 41.0% =34.85%
聯繫中小學	25% + 33.3% = 58.3%	42.24%	58.3% × 42.24% = 24.6%
名牌中小學	44.74%		
普通中小學	25.88%		



結論

雖然研究結果最後顯示名牌小學的入大學率都是最高，但是其實一條龍和直屬小學的入大學率都高於普通小學，甚至跟名牌小學相差無幾。可見即使同學輸在起跑線，沒有入讀名牌幼稚園，接下來也能夠入讀一條龍和直屬小學去提高自己的優勢。因此，家長根本不必為了讓子女贏在起跑線而費盡心思，這不但辛苦了自己,甚至會為孩子帶來沉重壓力，令他們失去快樂的童年。「看來這個社會是沒有起跑線的。」三人異口同聲地說，大家都對研究結果十分滿意。

(字數：2066 字)

參考資料：

[1] 基測百分百 HKDSE 優良率
http://blog.qooza.hk/ephmpntcj?eid=25924612&bpage=35
[2] 有直屬小學的中學-升學天地
https://www.schooland.hk/ps/feeder
[3]【DSE放榜】DSE 9狀元龍虎榜
https://topick.hket.com/article/2112573/
[4] 全港一條龍中小學名單
https://www.schooland.hk/ps/through-train
[5] 有聯繫中學的小學
https://www.schooland.hk/ps/nominated
[6] 龍媽自製龍校幼稚園升小攻略
https://topick.hket.com/article/1445702/
[7] 贏在轉跑綫？趙榮德：很多DSE狀元來自普通幼小
https://topick.hket.com/article/1780071/
[8] 平均數- 維基百科，自由的百科全書 – Wikipedia
https://zh.wikipedia.org/wiki/%E5%B9%B3%E5%9D%87%E6%95%B0
[9] 條件概率- 維基百科，自由的百科全書 – Wikipedia
https://zh.wikipedia.org/wiki/%E6%9D%A1%E4%BB%B6%E6%A6%82%E7%8E%87
[10] 樹形圖- 維基百科，自由的百科全書 –Wikipedia
https://zh.wikipedia.org/wiki/%E6%A8%B9%E5%BD%A2%E5%9C%96

亞軍作品:

"Hotel, Really Trivago?"

Discovering the Logic behind Hotel Selection of Trivago

School Name: HKUGA College
Name of Students: HUNG Lok Ching Emily,
LAM Kwan Yat Kyla, NG Hoi Tsing Tina
Supervising Teacher: Mr Timothy Ng

Abstract

Does Trivago actually recommend the cheapest prices of hotel choice? Did they fake the users? Did they live up to their claims? In this article, the writers are going to explore these questions using probability and statistics.

* **Lily** and **Joe** are watching TV at home. The TV was playing advertisements.*

TV : Hotel? Trivago.

Lily : Ugh, it's this advertisement again. Actually, does Trivago really works?

Joe : Well, Trivago claims that it "Compares the prices of over 600,000 hotels from over 200different websites" and " Makes it easy for you to find the ideal hotel for the best price".

Lily : Wow. If the claim was true, then it will be very impressive and effective to use. Let's try using it for the trip later this year. Our parents asked us to find the hotel for them, didn't them?

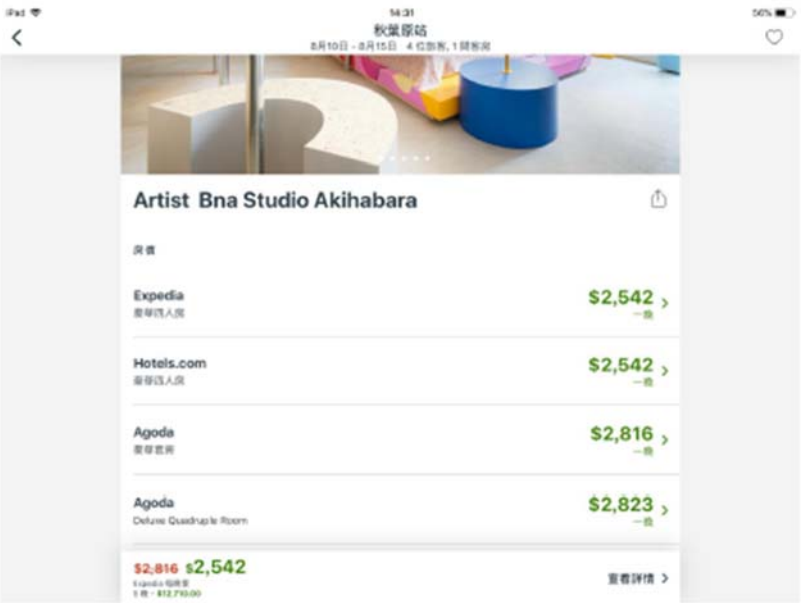
Lily : Oh right. Let's experiment using Trivago.

Lily : Here I have listed the requirements for this trip:

- **People: 2 Adults and 2 children (7 and 12 years old respectively)**
- **Dates: 10/8/2019 - 15/8/2019**
- **Destination: Tokyo, Japan**

Joe : I have just searched for a great hotel on trivago! It's called Arist Bna Studio Akihabara.

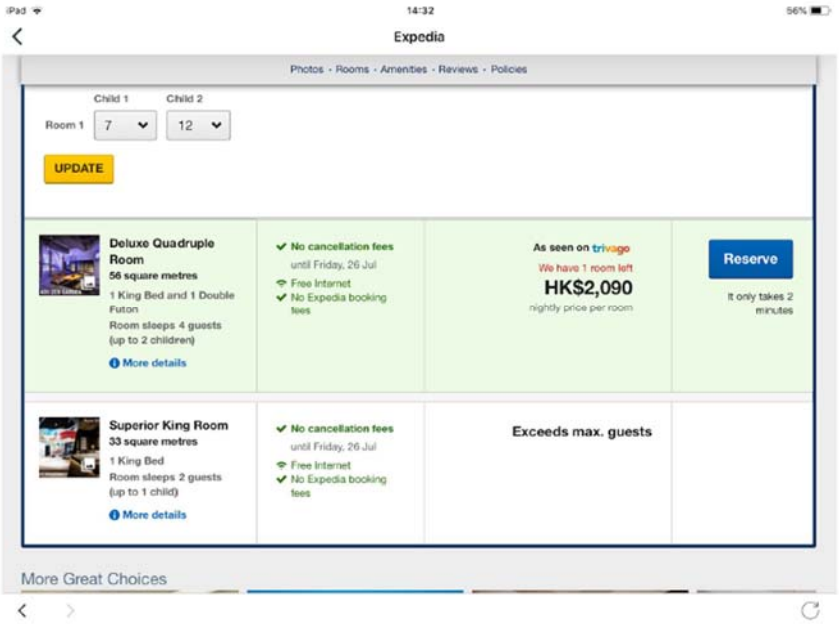
Take a look.



Lily : Great! I am going to have some more researches on this hotel.

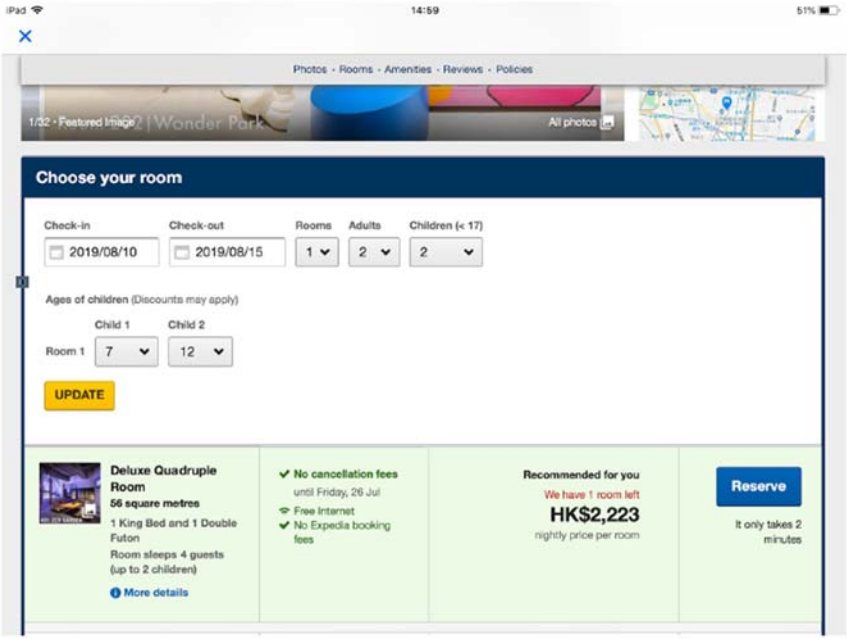
Joe : Let me click into the link of Expedia given by trivago. It shows that the price is $2542 per night. Hey, Lily. I remember that you have an Expedia app. I want to try to compare the prices to see if there are any differences. Can you search for this hotel on Expedia app for me, please?

Lily : Sure.



Joe : Wait, the price of the hotel shown on the Expedia link that trivago provides is $2090 per night. Why is it different?

Lily : I have just searched on the Expedia app, and it shows the price is $2223 per night. What happened?



Comparing table of the hotel deluxe room (4 people)

NETWORK	Trivago	Expedia link that trivago provides	Expedia App
PRICE (S)	2542	2090	2223

Joe : Look at the table above I have listed, the price varies between different platforms.

Lily : Wow. That's shocking. I think we should check the original price the hotel provides as well to compare.

Joe : Good idea. Let me do it now.

Comparing table of prices on different platforms

NETWORK	Trivago	Expedia link that trivago provide	Expedia App	Original Price Point from Hotel
PRICE (S)	2542	2090	2223	2445
The difference from the original price by the hotel (Corr. to 3 sig fig)	+3.97%	-14.5%	-9.1%	/

Joe : I have reorganised the table and compared the price points with the original.

Lily : Surprisingly, refer to this table, Trivago is actually even more expensive than the original price from the hotel. What's the point of using Trivago then?

Joe : Same thoughts here. I think we should dig deeper, I feel like may be Trivago is lying.

Lily : Here's the way this investigation to work. We have to collect mass data from different hotels and from different locations. The choices for the hotels must be random so it won't be biased and unfair. Then we can compare the price points and get a reasonable conclusion.

Joe : For the location selection, I think we can use cities that we normally visit.

Lily : You always have such creative ideas! Here is the list I found.

1. Tokyo, Japan
2. Seoul, Korea
3. Taipei, Taiwan

Joe : Time for me to work now. First of all, the platforms providing prices we are going to compare will be:

- The original price from the hotel
- Trivago
- Expedia
- Hotel.com
- Agoda
- Booking.com
- Trip.com
- Wington.travel

Joe : Then, to decide the hotels of different locations, I think we should do 10 hotels for each location, which will be 3 × 10 = 30 hotels needed.

Lily : Okay no more talking, we still have a lot of work to do, let's start now.

Joe : Here are the summary of steps I did.

1) Firstly, randomly select 10 hotels from Trivago website from the three cities: **Tokyo, Seoul and Taipei**. Fixed criteria: Stay at 10/8/2019 - 15/8/2019 for 2 adults, no breakfast.
2) Secondly, list prices recommended by Trivago and that of major hotel booking sites (6 of them) shown in Trivago

3) Thirdly, check prices from hotel official sites
4) Next, for hotels of Tokyo city selected, check the prices of major hotel booking sites **directly from their sites and compare them with that in Trivago**
5) Then, observe the relation of the prices recommended by Trivago and that of all prices available in Trivago
6) Lastly, observe the trends of the prices found

Lily : I appreciate the efforts you paid, now please present your findings!

Joe : Here it is! Take a look first!

Trivago shown prices

Location	Tokyo, Japan	Tokyo, Japan	Tokyo, Japan	Tokyo, Japan	Tokyo, Japan	Tokyo, Japan	Tokyo, Japan	Tokyo, Japan	Tokyo, Japan	Tokyo, Japan
Hotel Name	Mystays Premier Omori	Axas Nihonbashi	Richmond Tokyo Mejiro	Sun Members Tokyo Shinjuku	Monterey Ginza	Candeo S Tokyo Roppongi	Centurion Hotel Ikebukuro	Mimaru Tokyo Ueno East	Solaria Nishitetsu Hotel Ginza	Hilton Tokyo
The original price from the hotel	1,554	2,982	717	N/A	1,292	1,495	N/A	1,871	1,780	3,013
Trivago's recommendation:	Hotel	Expedia	Japanican	Expedia	Expedia	Expedia	Expedia	Agoda	Expedia	Amoma
Trivago	1,464	2,007	678	1,072	1,285	1,529	1,138	1,818	1,673	2,006
Expedia.com	1,464	2,007	not listed	1,072	1,285	1,529	1,138	2,158	1,673	3,255
Hotel.com	1,464	2,007	not listed	1,072	1,285	1,529	1,138	2,158	1,673	2,891
Agoda.com	1,465	1,852	848	1,080	1,081	1,523	1,131	1,818	1,781	2,979
Booking.com	1,606	2,669	847	1,382	1,285	1,685	1,131	2,158	1,780	3,227
Trip.com	1,455	not listed	679	953	1,288	not listed	1,429	N/A	2,815	not listed
Wingon Travel	1,455	not listed	679	990	1,287	1,626	1,139	Not listed	Not listed	1,931

Location	Seoul, Korea	Seoul, Korea	Seoul, Korea	Seoul, Korea	Seoul, Korea	Seoul, Korea	Seoul, Korea	Seoul, Korea	Seoul, Korea	Seoul, Korea
Hotel Name	Lotte Hotel Seoul	Hotel28 Myeongdong	Hotel Manu	InterContinental Seoul COEX	Sejong Hotel	Park Hyatt Seoul	Novotel Suites Ambassador Seoul Yongsan	The Splaisir Seoul Dongdaemun	Urban Place Gangnam	H Avenue Hotel Idae Shinchon
The original price from the hotel	1,863	1,252	559	1,921	655	2,161	1,262	956	619	675
Trivago recommended website	Expedia	Expedia	Expedia	Wing On	Wing On	Expedia	Trip	Hotel	Expedia	Expedia
Trivago	1,762	1,061	544	1,350	773	2,393	1,182	862	539	545
Expedia.com	1,762	1,061	544	1,923	904	2,393	1,489	862	539	545
Agoda.com	1,768	1,022	583	1,774	875	2,554	1,494	749	575	525
Booking.com	1,766	1,131	582	1,927	not listed	2,550	1,492	959	575	581
Trip.com	not listed	not listed	582	2,019	not listed	2,541	1,182	679	582	494
Wingon Travel	not listed	1,251	580	1,350	773	2,098	1,177	663	578	493

Location	Taipei, Taiwan	Taipei, Taiwan	Taipei, Taiwan	Taipei, Taiwan	Taipei, Taiwan	Taipei, Taiwan	Taipei, Taiwan	Taipei, Taiwan	Taipei, Taiwan	Taipei, Taiwan
Hotel Name	The Howard Plaza Hotel Taipei	Humble House Taipei	Eastin Taipei	Grand Mayfull Taipei	Green World - Grand Nanjing	Cho	Mandarin Oriental Taipei	Roaders	Hotel Royal Seasons	Beitou Hot Spring Resort
The original price from the hotel	1,031	2,108	603	1,941	625	625	3,378	594	660	1,039
Trivago's recommendation:	Amoma	Amoma	Expedia	Expedia	Booking	Expedia	Expedia	Wing on	Wing on	Expedia
Trivago	933	1,602	602	1,423	677	541	3,227	556	570	1,252
Expedia.com	973	1,789	602	1,423	677	541	3,433	537	630	1,252
Agoda.com	952	1,670	603	1,643	677	608	3,433	not listed	671	1,195
Booking.com	973	2,200	602	1,873	677	608	3,433	637	574	not listed
Trip.com	971	not listed	not listed	1,923	1,226	not listed	3,185	not listed	570	988
Wingon Travel	973	1,597	not listed	1,920	1,227	676	not listed	2,973	570	not listed

Lily : Oh, remember what the teacher taught us in school? We can calculate the mean and the standard deviation of these data.

Formula of standard deviation	Formula of mean
$SD = \sqrt{\frac{\sum\|x - \bar{x}\|^2}{n}}$	$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$

Joe : I agree. Since standard deviation means a quantity expressing by how much the members of a group **differ from the mean value for the group**, we can compare the prices in a mathematical way.

Lily : Let me calculate the standard score of Mystays Premier Omori by Trivago first. By calculation, I found that :

The **mean** is:
(1464+1464+1464+1465+1606+1455+1455) /7
=1481.86
For the standard deviation: 50.84
For the standard score of Trivago:
(Trivago's Price – Mean) / Standard Deviation
(1464 –1481.86) / 50.84
≈ – 0.3513

Joe : Let me do the rest of the calculation for all other hotels.

City/Country	Tokyo, Japan									
Hotel Name	Mystays Premier Omori	Axas Nihonbashi	Richmond Tokyo Mejiro	Sun Members Tokyo Shinjuku	Monterey Ginza	Candeo S Tokyo Roppongi	Centurion Hotel Ikebukuro	Mimaru Tokyo Ueno East	Solaria Nishitetsu Hotel Ginza	Hilton Tokyo
Standard Score of Trivago's recommended prices	-0.35	-0.35	-0.88	-0.13	0.40	-0.67	-0.39	-1.50	-0.55	-1.57

Location	Seoul, Korea									
Hotel Name	Lotte Hotel Seoul	Hotel28 Myeongdong	Hotel Manu	InterContinental Seoul COEX	Sejong Hotel	Park Hyatt Seoul	Novotel Suites Ambassador Seoul Yongsan	The Splaisir Seoul Dongdaemun	Urban Place Gangnam	H Avenue Hotel Idae Shinchon
Standard Score of Trivago's recommended prices	-1.09	-0.55	-1.79	-1.69	-1.13	-0.17	-1.08	0.63	-1.76	0.47

Location	Taipei, Taiwan									
Hotel Name	The Howard Plaza Hotel Taipei	Humble House Taipei	Eastin Taipei	Grand Mayfull Taipei	Green World - Grand Nanjing	Cho	Mandarin Oriental Taipei	Roaders	Hotel Royal Seasons	Beitou Hot Spring Resort
Standard Score of Trivago's recommended prices	-3.84	-0.79	-0.58	-1.52	-0.73	-1.22	-1.16	-0.60	-0.72	0.77

Lily : Then what does it actually mean when all the standard score of Trivago is negative? Is it a bad thing?

Joe : Well, remember standard score calculates how much the price is far away from the average price. So being a negative valued standard score, it means it is lower than average. That's not a bad thing, but a good sign. You always want a cheaper deal, right?

Lily : But most of the scores are between 0 and -1. So Trivago's prices are actually only a bit lower than other sites….Hang on! Some scores are even positive!

Joe: You are correct. **Trivago's recommended prices are most of the time not the lowest price**. With references to the table above, I've counted the total number of times while Trivago recommended price is not the cheapest. Each city I did for 10 hotels and mostly about 50% of the recommended prices are not the cheapest.

Location	Number of times trivago recommended price is not the cheapest
Tokyo	6
Seoul	5
Taipei	5

Lily : But doesn't Trivago claim to provide the cheapest prices to us? Why are the recommended prices not the cheapest? Then maybe Trivago is a scam!

Joe : Don't just jump into the conclusion. Trivago claims to provide the best prices. That doesn't necessarily mean the cheapest price, that's a misunderstanding. Based on the official website of trivago, it says:

The 'our recommendations' feature is based on a dynamic algorithm that shows you a range of attractive and relevant offers we think you're going to love. In the 'top position' we display in green the offer which our algorithm recommends as a great offer. Our algorithm takes into account a number of relevant factors, such as your search criteria (for example your location and stay dates), the offer's price, and its general attractiveness – for example, the experience we think you'll likely have on the displayed booking site. We also take into account the compensation booking sites provide us with when a user clicks on an offer.

Lily : Oh so the recommended prices are not necessarily based on only price but also based on the past experiences and compensations provided. Oh now I understand!

Joe : Secondly, prices shown in Trivago for the major booking sites are **different from that of individual site.**

	No. of times	**Reason**
Expedia.com	2	Trivago referral discount
Hotel.com	2	Trivago referral discount
Agoda.com	5	Trivago referral discount
Booking.com	1	Trivago referral discount
Trip.com	2	Wrong info in Trivago site
Wingon Travel	0	/

Lily: Why is that happening?

Joe: Major websites like expedia.com gave Trivago referral discounts. So if you click through trivago to expedia, the prices for the hotels may be

discounted. Meaning it may be cheaper. As you can see by the table, I have counted the number of times those major booking websites have trivago referral discounts and resulting to cheaper prices.

Lily : Wow. Agoda had 5 times of incidents, that's pretty common.

Joe : Thirdly, **hotel direct rates are always more expensive than the major booking sites** as you can see in the table.

Lily : Is there any exceptions? I have noticed one in Tokyo.

Joe : Sure there will be some exceptions. I have counted them below:

Location	**Number of times hotel direct rates are cheaper than major booking sites**
Tokyo	1
Seoul	0
Taipei	1

Lily : This is less than I expected. According to the finding, we may not order directly from the hotel anymore.

Joe : And for my last finding, I have noticed some of the websites constantly provide the same prices for hotels. I have also collected some interesting information. Actually, price pattern of the major booking sites reveals their ownership status.

Expedia and Hotel	Under same group and prices are always the same
Agoda and Booking	Under same group but prices are sometimes the same
Trip and Wing on	Under same group but prices are sometimes the same

Lily: Wow, I didn't know that. It's almost midnight now, I think we should conclude things and go to bed soon...

Joe: Yah sure.

Lily: So after we have summed up all the points, below is the conclusion after all the work we had done:

1)	Ignore Trivago's recommendation price.
2)	Comparing to booking directly with hotel, high chance to have a more favorable price by booking via hotel booking sites.
3)	Comparing to booking via hotel booking sites, high chance to have a more favorable price by going through Trivago, so enjoy the referral discount.

(Word count: 2480)

References:

[1] Trivago Advertisement (English Version):
https://www.youtube.com/watch?v=Zv9UbMFWxnM
[2]https://support.trivago.com/hc/zh-tw/articles/360016108153-trivago-%E6%88%91%E5%80%91%E7%9A%84%E6%8E%A8%E8%96%A6-%E6%98%AF%E5%A6%82%E4%BD%95%E6%8E%92%E5%BA%8F

季軍作品:

Safety Begins With Data

School Name: HKUGA College
Name of Students: Aggie Chow, Timothy Chau, Tiffany Lee
Supervising Teacher: Mr. Michael Yip
Word count: 2415 (including title)

Prologue:

Humans are soft beans, at any moment something as mundane as a car can kill us. While the thought of getting into a traffic accident wouldn't usually cross our minds, it's actually one of the major causes of death in developing countries. Of course, it's impossible and impractical for society to stop using cars. But in the spirit of road safety, we decided to do what we can. By studying when and where traffic accidents are most likely to happen, we can take extra precautions. Armed with this knowledge, perhaps our odds of surviving will increase…

A Story Based on a Real Event

Image reference: John Chan's Instagram post

This is a post from John Chan's Instagram, a secondary three student. He recently witnessed a car accident which caused him late for school. Here is a conversation between Sam and John, about the accident that happened this morning.

"Hey Sam! You know what? I think I almost died today,"

"You're over exaggerating, what happened?!" Sam asked

"This morning a car accident happened right in front of me! I had never thought that I would witness an accident first hand! It's a terrible experience! " John cried.

"Oh no! Were you okay?" Sam asked.

"I'm fine, but it was a little shocking to see it happened before my eyes. I hope that'll never happen to me, anymore," said John "so like, a thought popped into my mind: **how can I avoid getting into a car accident?**"

"***You can't! Unless you never step out your door!"*** Sam said.

Like the stubborn person he always was, John wanted to prove Sam wrong. He thought: ***if I analysed statistics about when and where traffic accidents are most likely to occur, I can avoid those places and time periods so that my chances of getting involved in one will be much smaller.***

So, let's start analysing!

WHEN are traffic accidents most likely to occur?

Thus, Sam and John started to discuss about which day and which time traffic accidents were most likely to occur in. They both had their own opinions.

"I think traffic accidents are most likely to occur on Sunday because most people would be free to spend their weekend outdoors to have fun," said John. "No way! In that sense, wouldn't there be **the largest amount of accidents on Friday?** Since more people would want to go out right after a long week of work" said Sam.

"Let's move on from this topic first and discuss about **which time period traffic accidents** are most likely to occur in," said John.

"Okay, may be there isn't a clear trend on which day traffic accidents are most likely to occur, but there must be a time where they usually happen?" asked Sam.

"I think it should be **around 6 p.m. to 7 p.m., because that's the time that most people finish their work and get back home. More vehicles,**

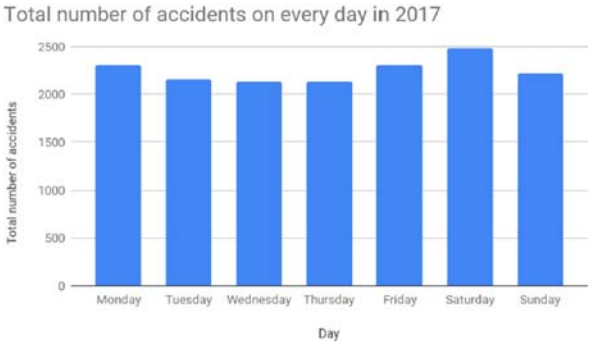
more car accidents, right?" said John.

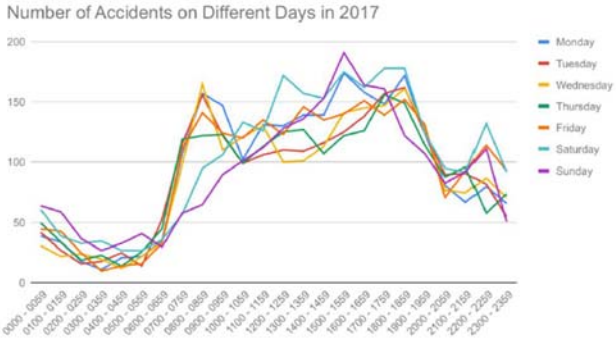
"Wait, I don't think so. Shouldn't it **be seven o'clock to eight o'clock in the morning? Both students and workers have to travel during that period so there should be more accidents**," said Sam.

"Hold up, we'll never come to a conclusion like this. Let's just have the data be the judge. Want to have a bet?" said John.

"Bring it on!"

Number of accidents in 7 days a week





After discussing with Sam, John found some data from the Transport Department of Hong Kong and analysed them using the chart above, by their **time period and days of week.**

"After getting this data, I found that the peak time of accidents is ***1500-1559 during Sunday***." said John.

"So you mean that we should avoid driving at 3 o'clock to 4 o'clock during Sunday?" said Sam.

"May be, but if we do, we should be extra careful during that time period," said John.

"Let's look at the other chart, it indicates that Saturdays had the highest number of accidents…" said Sam.

"1700-1759 and 1800-1859 share the same number of accidents on Saturday," said John.

"Wait, I'm confused. So should I not go out on Saturday or 1500-1559 during Sunday?" asked Sam.

"Just be careful on both will be fine," commenced John.

"How about the trend?" said John?

"The number of accidents happened between the time period 0000-0659 were the less among all the time periods, which makes sense," said Sam.

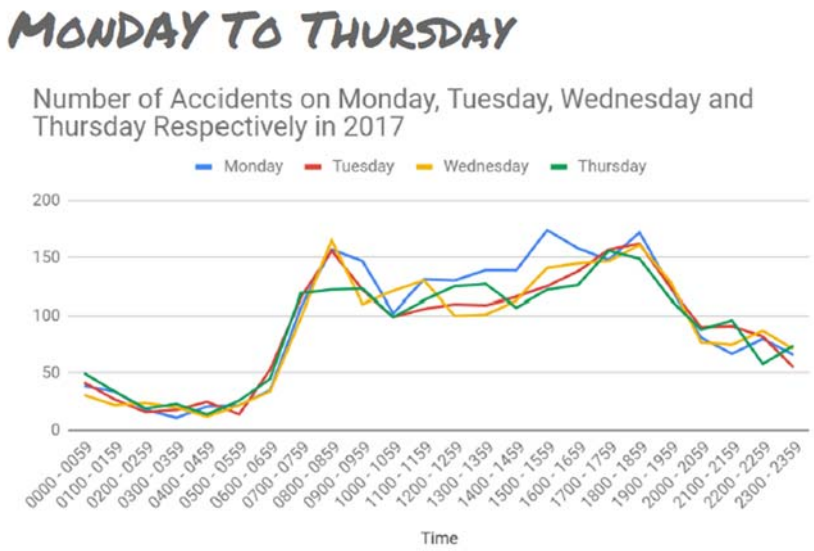
“The number of accidents increases drastically afterwards, **except for Sunday and Saturday**.” John said.

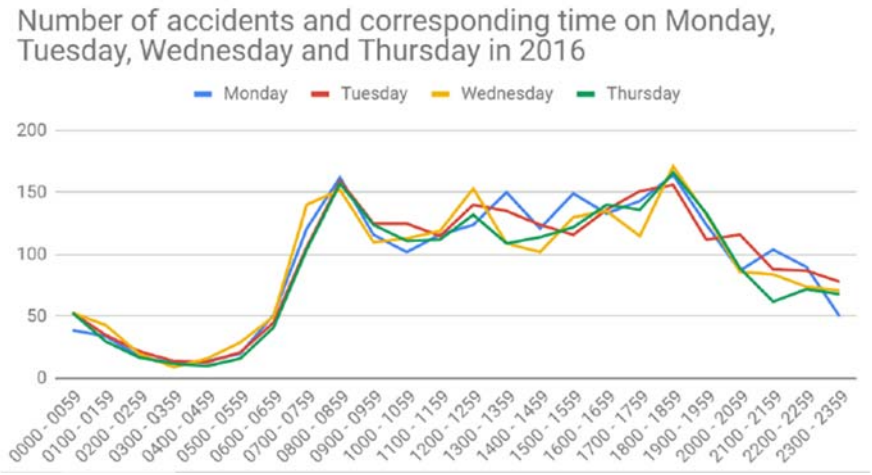
“The number of accidents starts to decrease in 0800-0859. During lunch hour, the number of accidents increases again. The amount of accidents fluctuated between 100 and 190 during 1100-1159 and starts to decrease from 1800-1859 until the end of the day.” Sam said.

“But why does the line representing Sundays and Saturdays have a big difference from the Monday to Friday one?” John said.

“Let’s separate them into two charts and see what’s going on.”

So Sam and John continued on their quest for knowledge, at the expense of neglecting the work that they were supposed to do.





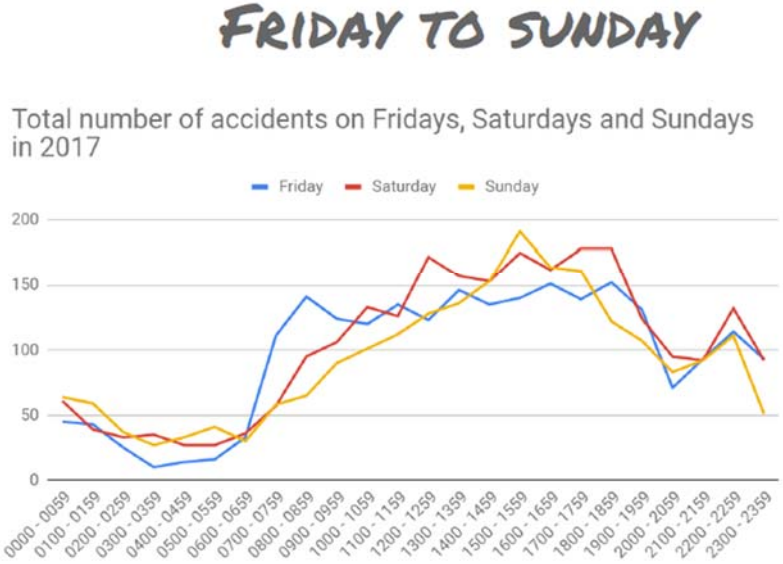
On Mondays to Thursdays, Traffic accidents **typically peak at eight to nine o'clock in the morning and at six to seven o'clock at night**. In those time periods many people are travelling from home to work and vice versa, so there are more cars on the roads and thus more traffic accidents are likely to occur.

There's a big difference between the number of accidents happening between 20:00 to 06:59 and between 07:00 to 19:59, when accidents are much more likely to happen.

However, out of these four days, the time period with the most accidents happening would be 15:00 to 15:59 on Thursday, Thursdays seem to have **slightly more accidents in general from 07:00 to 19:59 too**.

"I also spotted that the blue line which represents Mondays' data, is slightly different from the other 3 days," said Sam.

“This is probably because of the black Monday effect, which everyone were simply sleepy and don’t want to work. As a result, accidents are most likely to occur due to the low level of concentration,” John said, pretending to be an expert.



“According to the chart, I have found out that most of the accidents actually happened **during 1500-1559 on these three days**! Probably because most of the family events end at this time. ” John said.

“Interesting. Also, maybe since Friday is a working day, the number of accidents is typically higher than Sunday and Saturday during the time period of 0800-0859, which matches what we derived from the ‘Monday to Thursday’ chart!” exclaimed Sam.

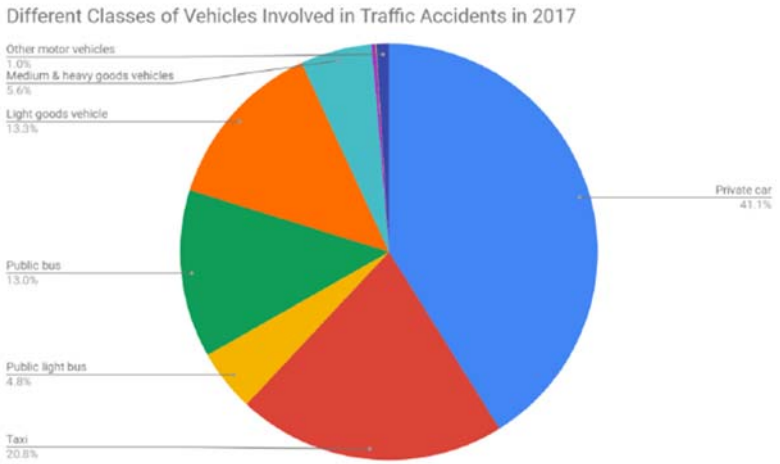
"We can conclude from the chart that the trends of Sunday and Saturday are similar as well. And the trend of Friday follows the trend of Monday to Thursday." said John.

"Oh, it was my dad driving this morning! May be the type of vehicle used is a factor too?" asked John.

"Let's look at this graph!" said Sam.

FACTORS AFFECTING THE ABOVE DATA

CLASSES OF VEHICLES



"Wow, this is amazing….."

Sam: "What? The number of medium and heavy good vehicles only consist 5.6% of all types of car involved in accidents?!"

John: “It’s still reasonable to see private cars in the first place.”

Sam: “Um, I think this may be because the drivers working in public transport are well trained enough and there are simply more private cars on the road!”

John: “Then judging by the graph, taking public transport is actually safer than driving my own private car!”
We hypothesize that one of the ways to minimize your risk of getting into a car accident is by taking public transport instead.

WHERE are traffic accidents most likely to occur?

“John, where did you see the accident?”, asked Sam.

“Cross Harbour Tunnel.” John answered.

“Oh well, I bet accidents always happen there, since so many cars pass through there every day,” Sam said.

“Probably, but car accidents always happen in Tseung Kwan O Tunnel as well.” John said.

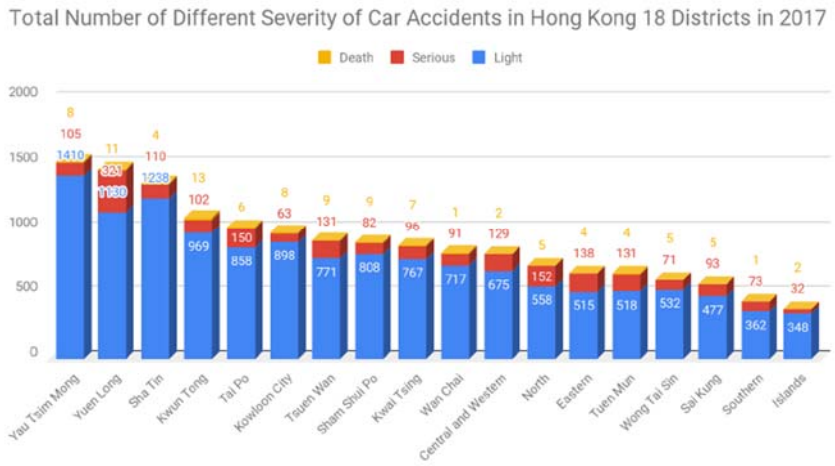
“Well, it seems that our opinions are not the same again. Then in which district do you think that car accidents are most likely to occur?” Sam asked.

“Um... let me think, Kwun Tong?”, John answered.

“I think it is Sha Tin, ” Sam said.

Their opinions differed once more, so they started to search for data to analyse.

Out of the 18 districts in Hong Kong, Yau Tsim Mong district has the highest amount of car accidents happening there overall.



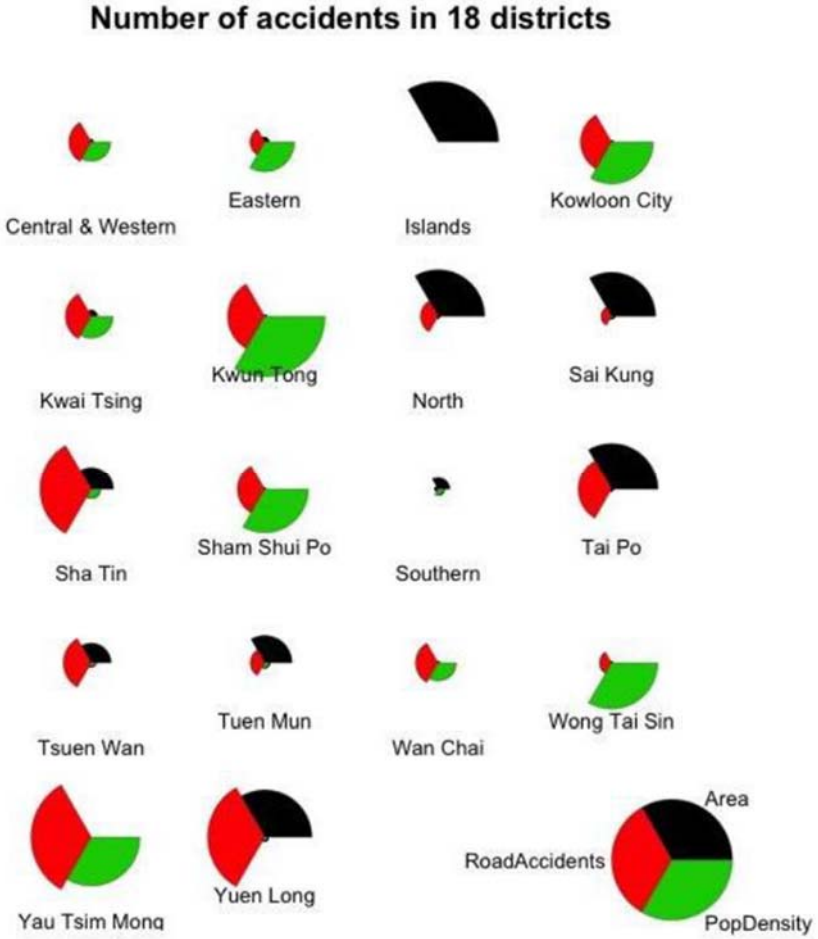
The number of car accidents happening in Yuen Long district and Sha Tin district are also noticeably higher. Yau Tsim Mong also has the highest amount of ‘light’ car accidents, followed by Yuen Long and Sha Tin once more. However, the district that has the highest amount of serious accidents, which is much higher than that of other districts, is Yuen Long. Other districts have more or less the same amount (except for the Islands,

which have much less). The amount of car accidents that involved deaths is constant across all districts.

“Hold on,” interrupted Sam, “Would some districts have a particularly higher amount of accidents because their area is bigger and the population density is higher?”

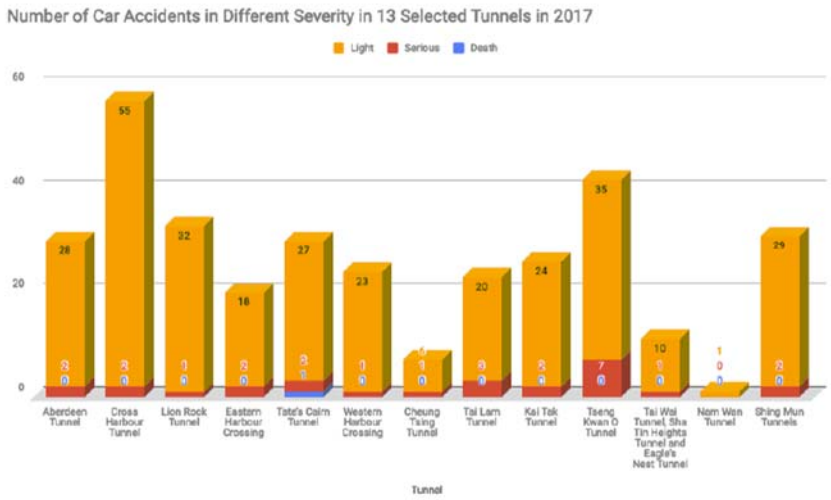
“Oh yeah… I didn’t think of that,” John admitted.

However, that may not always be the case. Here we have a star plot representing area, population density, and amount of road accidents each district has. The size of the sector is proportional to the respective value it represents. Large districts with low population density are expected to have fewer accidents, e.g. Islands, North and Sai Kung Districts. In contrast, Yau Tsim Mong and Kwun Tong districts have high accident rate while their size is small and the density is high. Yuen Long and Tai Po districts are an exception though. They are accident-prone despite its large area and low population density.



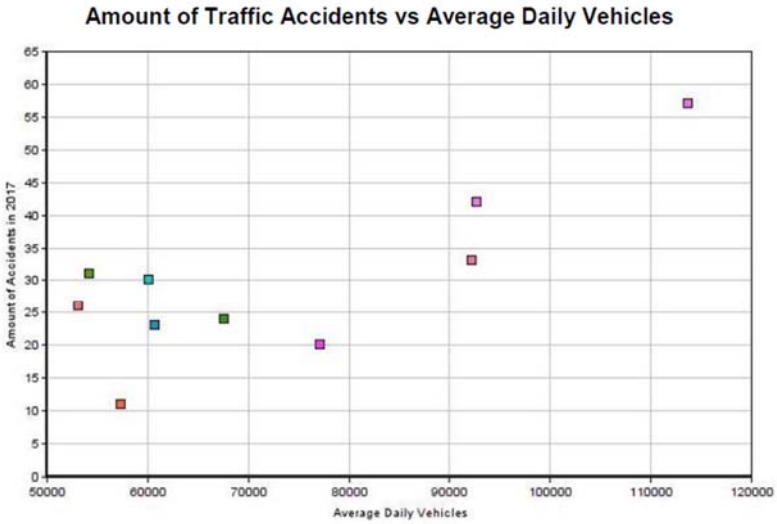
“Hmm… That’s still too many exceptions to call it a clear trend,” Sam said, “May be districts are too broad of an area to research on...”

“Let’s try looking at data on tunnels then!” suggested John.



Note- Area of the tunnels are defined by the area dictated in the law, which include the interior and exterior of the tunnels.

The tunnel with the most accidents was the Cross Harbour Tunnel, not surprising at all, considering that it is the tunnel with the most traffic in Hong Kong. However, the amount of accidents in a tunnel doesn't always have to do with its traffic conditions. Below is a scatter plot illustrating the relationship between the amount of accidents a tunnel has and the amount of cars that pass through it every day.



*Due to limited data, not all tunnels in the above list are used in the above graph

Typically the higher the amount of average daily vehicles, the more accidents a tunnel has. But again, there are multiple exceptions. We can't really reach a concrete conclusion since there isn't much difference between the amounts of accidents each tunnel has.

In conclusion, Yau Tsim Mong and Kwun Tong are especially dangerous districts to drive in, and the Cross-Harbour Tunnel requires attention as well. The busiest areas always have the highest amount of traffic accidents, so we should be especially vigilant when driving through busy areas like those. But the population density and the average amount of vehicles daily don't have a consistent relationship with the amount of car accidents an area has, so in other areas road conditions are probably a bigger factor.

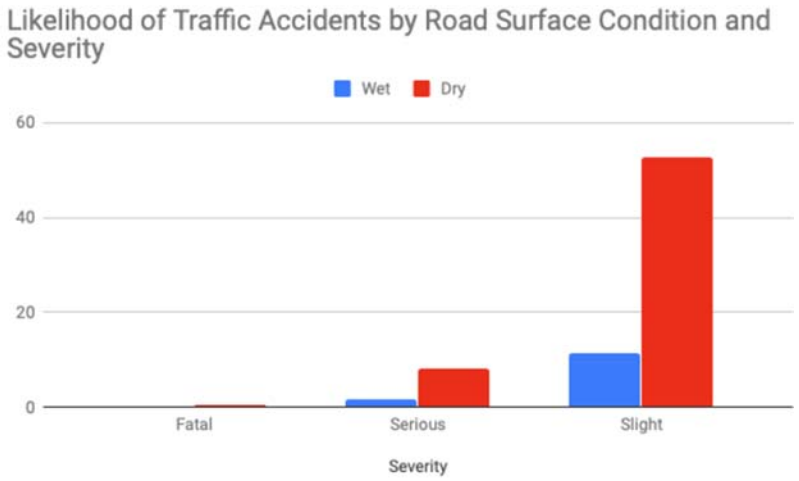
WHAT WEATHER CONDITIONS are Traffic Accidents more likely to occur in?

“I’m not sure how useful this information can be, but there is one thing I’m sure of. Traffic accidents are more likely to occur in the rain, right?” said Sam.

“But judging by what we learnt today, the answer to that might be surprising,” John challenged, “We should do some research before coming to conclusions.”

“No way, there’s no reason why wet roads wouldn’t be more dangerous than dry roads- It’s harder for cars to stop when the ground is slippery, that’s why accidents are more likely to occur in rainy days, everyone knows that!” Sam exclaimed.

“Then why don’t you see for yourself?” John said.



“What?! The probability of getting into an accident on a wet road is much lower than the chance of getting into an accident on a dry road?” exclaimed Sam, “But how can that be?”

“I don’t really understand either,” confessed John, “Logically speaking, it would be more dangerous to drive on a slippery road, wouldn’t it? Maybe it is because people are especially alert when driving in dangerous conditions and end up performing better, or maybe more people take the MTR instead on those days.”

“That makes sense,” said Sam, “I guess there are things that we just can’t know for sure until we do more research.”

******

In the end, Sam and John forgot about the research they spent their precious time on. But then again, even if they followed the information they gained from these graphs and altered their lifestyle to be incredibly inconvenient, it still wouldn’t be a foolproof plan. After all as Murphy’s Law goes - “Whatever can go wrong will go wrong”, so as long as cars exist, so will traffic accidents. The statistics shown today are a bit of extra knowledge, but please don’t follow them exactly. Ultimately, the best way to avoid traffic accidents is simply to be careful and stay alert at all times.

CONCLUSION

Collecting data and designing graphs were quite foreign tasks to us. After finishing the problem, we concluded that there isn’t a definite way to avoid

car accidents since there is an infinite amount of factors that can affect our topic, such as the state of the driver, problems with cars, or even a planned murder etc. While there may be a trend, unpredictable exceptions can always occur, especially when there are so many factors that lead up to such events, so it is best that we stay vigilant at all times.

Yet, we still want to do the best we can do to avoid car accidents and help readers of this project to be a safe person by analysing data related to road accidents.

In conclusion, we found out that using public transport can lower our probability for getting involved in car accidents, that we should avoid busy areas and rush hours, and that the weather doesn't necessarily indicate safer or more dangerous conditions.

Stay safe!

References:

[1] Data from the transport department:
https://www.td.gov.hk/tc/road_safety/road_traffic_accident_statistics/2017/index.html
[2] The census and statistics department:
https://www.censtatd.gov.hk/hkstat/sub/sp150.jsp?productCode=FA100096
[3] Hong Kong Weather Observatory:
https://www.hko.gov.hk/cis/statistic/rf_1_e.htm

優異作品:

運動成績與體格的關聯

學校名稱：梁文燕紀念中學（沙田）
學生姓名：方肇楠，方偉華，林均俊
指導教師：陳志文老師

摘要：

相信大家在中學生涯中都參加過陸運會，大家當時會否羨慕在頒獎臺上的同學呢？而你們又有沒有想過他們是憑甚麼獲勝呢？是出色的技術嗎；還是過人的體格？在本篇文章中我們會探討運動員的身高體重對其表現的影響。

左方為謙謙，右方為林林

在體育課上，體育老師問了我們一個問題：大家的身形究竟跟跑步有什麼關係呢？林林認爲越高大的人就跑得越快。謙謙認爲體重越重就跑得越慢。其實在醫學界有一個叫 BMI 的指標，BMI 的設計是一個用於公眾健康研究的統計工具。當需要知道肥胖是否為某一疾病的致病原因時，可以把病人的身高及體重換算成 BMI，再找出其數值及病發率是否有線性關連。BMI 指數可以通過以下公式計算：

$$\mathbf{BMI} = \mathrm{w} \div h^2$$

W 是重量（公斤），h 是高度（米）。

體重指標	類別
18.49 或以下	過輕
18.5 – 22.9	適中
23.0 – 24.9	過重
25.0 – 29.9	肥胖
30.0 或以上	極度肥胖

圖表(一)

於是，我們就連同體育科老師和數學科老師一起收集及分析了全校學生的體適能數據，並集合了全港中學生的體適能數據，再加以分析，我們得出了以下結果:

項目	林林	謙謙	本校同學平均水平	全港學生平均水平
一分鐘仰臥起坐(次)	30	10	20.3	33.07
坐地前伸(釐米)	40	9	23.8	34.26

項目	林林	謙謙	本校同學平均水平	全港學生平均水平
掌上壓(次)	35	8	13.5	19.43
九分鐘跑(米)	1700	1350	1473.1	1538.77
身高(米)	1.73	1.6	1.692	1.6966
體重(千克)	75	63.7	59.8	59.26

圖表(二)

經過計算，林林的 BMI 指數為 25.0，而謙謙的 BMI 指數為 24.8，看上去林林的身形高大，但是 BMI 指數卻和謙謙差不多。從圖表(二)中可看到，他們在長跑項目中相差 350 米的距離，但短跑又如何？

讓我們回顧 2016 年裡約奧運會男子 100 米的首八名：

2016 年夏季奧林匹克運動會田徑男子 100 公尺比賽	身高（米）	體重（公斤）	BMI 指數
尤塞恩·博爾特	1.95	86	22.6
賈斯汀·加特林	1.85	79	23
安德烈·德格拉塞	1.76	68	21.9
約安·布雷克	1.8	76	23.4
阿卡尼·西姆拜恩	1.74	67	22.1
本·約瑟夫·梅特	1.79	70	21.8
吉米·維科	1.86	83	23.9
特拉伊瑪·布魯梅爾	1.73	70	23.3
平均值	1.81	74.875	22.75

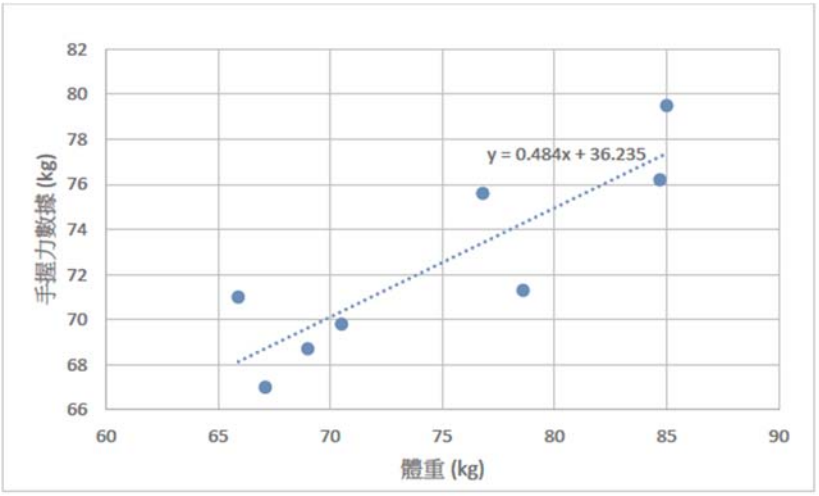
圖表(三)

經過計算，他們BMI指數的總體標準差為0.7228，不高。而他們身高和體重的總體標準差分別為6.892和6.7535。以上的數據顯示，雖然身高和體重有不同的分別，但是只要你有一個適當的BMI指數，即是身高和體重有一個一定的比例，你便能在短跑中取得佳績。我們在這裡也可以作出一個大膽的假設：林林和謙謙的短跑時間不會相差太遠。為了測試我們的假設是否正確，我們特別邀請了林林和謙謙進行一次短跑100米競賽，林林的成績為13.75秒；而謙謙的成績是14.19秒，當中只相差0.44秒，證明了林林和謙謙的BMI指數在短跑這項運動中表現相仿。

除了短跑，其實其他運動都有一個特定的BMI指數範圍，該範圍內的人取得勝利的機會很大。為了證明，我們也找到了馬拉松的世界紀錄的前三名的BMI指數，分別為埃利烏德·基普喬蓋：20.0, 鄧尼斯·基普魯托·基梅托：18.8，肯內尼薩·貝克勒：21.0。可以看到馬拉松這類長跑的選手BMI指數比較低，以上三位的BMI指數平均值為19.9。

謙謙和林林的 BMI 指數那麼高，有甚麼運動是適合他們的呢？2016 裡約奧運的鉛球冠軍里安·克勞澤身高 2.01 米，體重 132 公斤，BMI 指數 32.1，但這對他們來說還算是太高了，而且我相信對香港的大部分學童也不是十分適合的。此時我們的另一位朋友—小明走了過來參與我們的討論，他是我們學校鉛球比賽的記錄保持者，他剛剛做完手握力測試，於是他便問手握力的強度是否與推鉛球有關，爲了解答他的疑問，我們再次收集了我校鉛球比賽八強的手握力數據作出比較：

鉛球比賽八強的手握力	手握力數據 (千克) (左手+右手)	體重 (千克)
1.	79.5	85.0
2.	76.2	84.7
3.	75.6	76.8
4.	71.3	78.6
5.	71.0	65.9
6.	69.8	70.5
7.	68.7	69.0
8.	67.0	67.1
平均值	72.3875	74.7



圖表(四和五)

從以上圖表可得知，發現手握力的強度與推鉛球的距離(成績)真的有關聯。在圖表(四)中可看到第一名的手握力數據最高，有79.5千克，然後手握力的強度隨著參加者的名次下降，到67千克: 有越高的手握力能在推鉛球中取得更好的成績。

回到我們的主題，參加者的 BMI 指數與手握力(及推鉛球成績)又是否有關？我們根據推鉛球八強的體重及手握力數據繪畫了圖表(五)進行比較。我們發現兩者成正比，參加者體重越高，手握力的指數便越高。透過繪畫圖表後，我們便要尋找該條線性回歸的直線方程(最佳配適線)，這次我們只需要進行簡單計算便能夠計算出圖中線性回歸的直線方程。

A	B	C	D	E	F
x 體重(kg)	y 手握力數據(kg)	$x-\overline{x}$	$y-\overline{y}$	$(x-\overline{x})(y-\overline{y})$	$(x-\overline{x})^2$
85	79.5	10.3	7.1125	73.25875	106.09
84.7	76.2	10	3.8125	38.125	100
76.8	75.6	2.1	3.2125	6.74625	4.41
78.6	71.3	3.9	-1.0875	-4.24125	15.21
65.9	71	8.8	-1.3875	12.21	77.44
70.5	69.8	-4.2	-2.5875	10.8675	17.64
69	68.7	-5.7	-3.6875	21.01875	32.49
67.1	67	-7.6	-5.3875	40.945	57.76
體重的平均數$\overline{x}$	手握力的平均數$\overline{y}$			總和	總和
74.7	72.3875			198.93	411.04
		斜率(m)	0.483967		
		y 軸截距(c)	36.23513		

圖表(六)

在圖表(六)中，我們進行運算。該條直線方程的斜率只需將E欄總和除以F欄總和，便能得出0.484(斜率, m) 。而y軸相交點(c)則為36.235。

圖中線性回歸的直線方程是 $y = 0.484x + 36.235$ 。十項全能的冠軍羅曼·謝布爾勒的身高體重為1.86米，體重為88公斤，BMI指數為25.8，對他們來說非常適合。

在最後，如果你希望在運動中取得佳績，根據上述的研究，我們可以得出運動員的成績是與 BMI 指數直接掛鈎的。就如上述，謙謙和林林的 BMI 指數與奧運選手大致相約，因此，推論他們能獲得獎項，而事實證明他們在學校陸運會中取得獎項。而在鉛球比賽中，小明能取得佳績亦是和體重與手握力息息相關。體重和手握力越高，在比賽中取得的名次越高。因此，我們可以證明我們的推論是正確的。根據以上對鉛球和短跑的分析。我們可以大膽地假設運動成績是和身體質量指數成正比。我們根據這個猜想，搜集了上年度陸運會鉛球的冠軍和他的各項身體數據，見以下圖表(七):

sex	house	grade	totheats	heat_nu	lane_nc	position	record	remarks	finalpos
M	Y	B	1	1	2	0	8.84		1
M	R	B	1	1	3	0	8.63		2
M	P	B	1	1	28	0	8.49		3
M	Y	B	1	1	1	0	8.45		4
M	Y	B	1	1	16	0	7.24		5
M	Y	B	1	1	26	0	7.19		6
M	P	B	1	1	5	0	6.69		7
M	R	B	1	1	19	0	6.55		8
M	G	B	1	1	21	0	6.08		9
性別	社	級別				順序號碼	記錄		名次

圖表(七)

在名次一欄中，第一名的推鉛球紀錄是8.84米，再參考圖表(八)看看他的體重和身高，發現其BMI指數和手握力數據和紀錄保持者的相似，因此他在比賽中取得佳績。由此我們發現身體質量指數能幫助

我們在相應的運動項目中取得優勢，較易獲取好成績。

	記錄保持者	上任鉛球冠軍
體重	90.5 千克	95.1 千克
身高	170 厘米	180.5 厘米
BMI 指數	31.3	29.1
手握力	79 公斤（左手加右手）	76.5 公斤（左手加右手）

圖表(八)

因此，各位在選擇運動時可以參考一些運動員的身體質量指數來選擇一個適合自己的運動了。

參考資料：

[1] 維基百科: 2016年夏季奧林匹克運動會田徑男子100公尺比賽

[2] 梁文燕紀念中學（沙田）二零一八至二零一九陸運會數據

[3] 梁文燕紀念中學（沙田）體育科體適能數據

[4] 香港中學生體適能常模表：

https://cd1.edb.hkedcity.net/cd/pe/tc/rr/pfs/sec_09_10_c.pdf

優異作品:

罰中有序

學校名稱：中華傳道會安柱中學
學生姓名：陳澤聲、曾智翹、伍俊聲
指導教師：周恩隆老師

摘要：

NBA(National Basketball Association) –國家籃球協會，是不少籃球粉絲所喜愛和關注的熱門話題。當俊聲看到有關 James Harden 在主場的傑出罰球表現的新聞時，立即與自己的好朋友智翹和澤聲討論有關主場優勢的問題。他們嘗試運用統計學作分析，以 NBA 球星的罰球數據來尋找出箇中的「主場之利」。

一天，俊聲、澤聲和智翹看到以下新聞:

「香港時間 3 月 1 日，火箭在主場以 121-118 險勝熱火，本場比賽，占士·夏登出戰 44 分鐘 32 投 16 中，三分球 18 中 8，罰球 18 中 18，轟下 58 分 7 籃板 10 助攻 4 偷球 1 封籃。這是夏登生涯第 59 次單場至少罰進 15 球，高居 NBA 歷史第一。」

（新聞內容節錄）

James Harden

HOUSTON ROCKETS

俊聲：哇，James Harden又入這麼多罰球了！依我看，一定是因為「主場優勢」的緣故。

澤聲：主場優勢？甚麼是主場優勢呢？

俊聲：主場優勢，顧名思義，就是因為比賽在主場（自己球隊的體育館）舉行，所以球員在環境和心理上有優勢，例如更熟悉比賽場地及當地的氣候，以及由於支持者佔大多數，因而能夠得到更多的加油助威，並對對手施加更大的壓力。

澤聲：原來如此！這麼看來，James Harden 的罰球表現很可能就是因為主場才能這麼出類拔萃！

俊聲：要證實James Harden 的罰球表現是不是在主場較好，不如我們來統計一下？

澤聲和智翹：好！

俊聲：今天是2019年3月9日，那麼我們就只統計由開季(2018年10月17日)至今的賽事吧。同時，我們可以來主場和客場的罰球表現作對比。

智翹：截至 2019 年 3 月 9 日，James Harden 的罰球表現如下：

表一、2018-2019 James Harden 主客場罰球表現

	主場 罰球次數	主場 罰球進球數	主場 罰球命中率	客場 罰球次數	客場 罰球進球數	客場 罰球命中率
James Harden	375	336	89.6%	312	267	85.6%

智翹：我們可以計算**罰球命中率**來比較和清楚表達主客場的罰球表現。

罰球命中率公式

$$\text{罰球命中率} = \frac{\text{罰球進球數}}{\text{罰球次數}} \text{ x } 100\%$$

當主場或者客場罰球命中率越高，代表球員在主場或者客場罰球表現較好；反之，當主場或者客場罰球命中率越低，則代表球員在主場或者客場罰球表現較差。從中可見，James Harden 的主場罰球表現明顯比客場較佳，主場罰球命中率比客場罰球命中率約高 4%。

澤聲：看來 James Harden 確實擅長於主場。這麼說，球員在主場，罰球的表現是不是就較好呢？

智翹：也不能一概而論，不同球員在不同的地點有不同的能力，例如以下的一組球員，其中五人（除了 Lebron James）皆擔任後衛，而 Lebron James, James Harden 和 Stephen Curry 更曾獲得最有價值球員(MVP，Most Valuable Player)的殊榮，對各自球隊有著豐功偉績：


Damian Lillard


James Harden


klay thompson


Kyrie Thompson

Lebron James

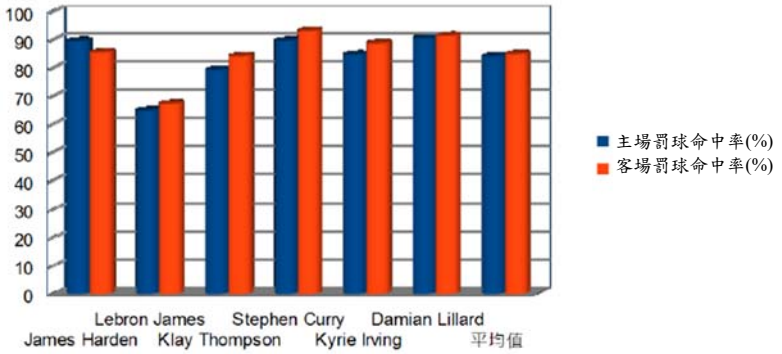

Stephen Curry

表二、2018-2019 球員主客場罰球表現

	主場罰球次數	主場罰球進球數	主場罰球命中率	客場罰球次數	客場罰球進球數	客場罰球命中率	命中率差(主－客)
James Harden	375	336	89.6%	312	267	85.6%	4%
Lebron James	183	119	65.0%	178	120	67.4%	-2.4%
Klay Thompson	83	66	79.5%	57	48	84.2%	-4.7%
Stephen Curry	107	96	89.7%	132	123	93.2%	-3.5%
Kyrie Irving	92	78	84.8%	98	87	88.8%	-4%
Damian Lillard	191	173	90.6%	233	213	91.4%	-0.8%
平均值	172	145	84.2%	168	143	85.0%	-0.8%

智翹：以棒形圖來看：

2018-2019 球員主客場罰球表現



（棒形圖：球員主客場罰球表現）

智翹：其中，平均值(Mean)是指統計對象的一般水平，也是描述數據集中趨勢的一種方法。我們既可以用它來反映一組數據的一般情況，也可以用它進行不同組數據的比較，以看出組與組之間的差別。要計算平均值，可以利用以下的公式。

平均值公式

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

(n =數據的數量)

智翹：在這裡，我們就用平均值比較了六位球員主客場的罰球表

現。客場罰球表現較佳，儘管受到 James Harden 這個極端值影響，客場罰球命中率比主場罰球命中率仍然高了 0.8%

澤聲：極端值？

智翹：極端值（Extreme values）是指在統計中，特別大或特別小的數值。在這裏 James Harden 的主場和客場發球次數為大約 300 以上，明顯比其他球員的 100 左右高出很多，是一個極端值。極端值對於數據的平均值影響很大，因此這時候我們可以利用更多球員的數據來減少極端值的影響，例如在我們這裏就利用了六個球員的數據。

澤聲：你真是人如其名的「智多星」！那麼表中出現的命中率差又是甚麼呢？

智翹：命中率差，顧名思義，是主場罰球命中率和客場罰球命中率的差異（主場罰球命中率-客場罰球命中率），用以反映主場和客場的罰球表現差距。當命中率差為正數，代表主場罰球表現比客場較為良好；若為負數，則代表客場罰球表現比主場較為良好。六個球員中，五個球員的命中率差都為負數，反映這些球員中，客場罰球表現普遍較好。

澤聲：命中率差和平均值反映這一組球員的客場罰球表現的都優於主場，看來真不可以一概而論。

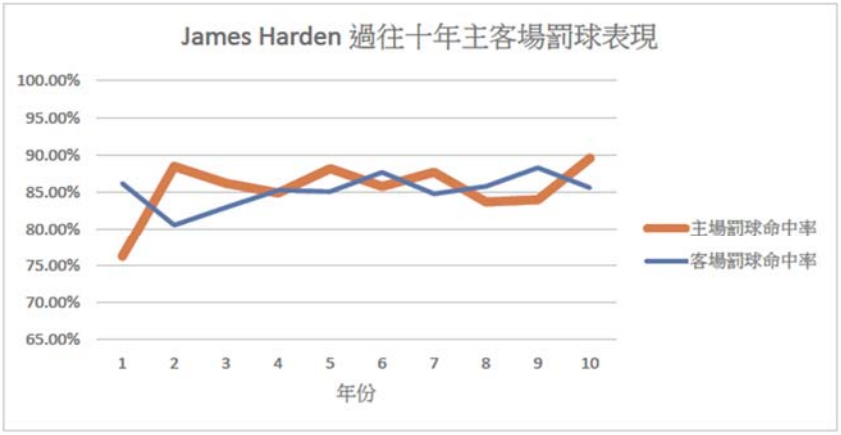
俊聲：要知道球員的罰球表現無法單憑一個賽季來下定論，不如我們再統計一下James harden過往的表現？

智翹：好，經過搜集網上資料統計，以下是 James Harden 過往十年，即 2009-2010 賽季至 2018-2019 賽季的主客場罰球表現：

表三、過往十年 James Harden 主客場罰球表現

年份	賽季	主場罰球次數	主場罰球進球數	主場罰球命中率	客場罰球次數	客場罰球進球數	客場罰球命中率	命中率差(主—客)
1	2009-2010	131	100	76.3%	109	94	86.2%	-9.9%
2	2010-2011	157	139	88.5%	186	150	80.6%	7.9%
3	2011-2012	181	156	86.2%	188	156	83.0%	3.2%
4	2012-2013	398	338	84.9%	394	336	85.3%	-0.4%
5	2013-2014	330	291	88.2%	335	285	85.1%	3.1%
6	2014-2015	408	350	85.8%	416	365	87.7%	-1.9%
7	2015-2016	406	356	87.7%	431	364	84.8%	2.9%
8	2016-2017	466	390	83.7%	415	356	85.8%	-2.1%
9	2017-2018	419	352	84.0%	308	272	88.3%	-4.3%
10	2018-2019	375	336	89.6%	312	267	85.6%	4%
	總數	3271	2808	85.8%	3094	2645	85.5%	0.3%
	平均值	327.1	280.8	85.8%	309.4	264.5	85.5%	0.3%

智翹：也可以參考下圖 James Harden 在過往十年的罰球表現：



（折線圖：James Harden 過往十年主客場罰球表現）

智翹：從資料可見，James Harden在過往十年的罰球表現中，主場表現稍微佔優：平均值方面，主場罰球命中率的平均值為85.8%，比客場罰球命中率85.5%高了0.3%；命中率差方面，主場表現仍然較好，中位數為1.25%，反映主場罰球表現較好。

俊聲：甚麼是中位數?

智翹：中位數（median）也是描述數據集中趨勢的一種方法，它是一組數據中位於中間位置的數字，因此不容易被極端值影響。計算中位數時要由小至大排列。要計算中位數，也可以利用以下兩種公式，分為偶數和奇數：

中位數公式

$$x = \text{第}\left(\frac{n+1}{2}\right)\text{項}\left(\text{若}n\text{為奇數}\right)$$

$$x = \frac{1}{2}\left(\text{第}n\text{項} + \text{第}(n+1)\text{項}\right)(\text{若}n\text{為偶數})$$

智翹：計算命中率差時，上表中有十個數據，屬於偶數，因此中位數是第五（-0.4%）和第六個數據（2.9%）的總和除以二，得出命中率差中位數為1.25%。

澤聲：主場罰球命中率在平均值和中位數上都比客場要高，可以看到主場優勢確實存在。同時以最大值（maximum）來看，主場罰球命中率曾經到達本年度 89.5%的歷史高位。而客場罰球命中率只到達

87.7%，比主場少了 1.8%，反映 James Harden 在主場罰球的潛力比客場要高。

智翹：沒錯，當我們比較球員主客場的表現時，除了考慮整體表現，也要考慮到球員在主客場的潛力，這個時候我們就可以用最大值比較。

俊聲：既然James Harden在主場的潛力比客場大，這是不是說明主場優勢較大？

智翹：是的，不過現在有些球員更加熟悉其他環境，主場優勢並不那麼顯著。正如表二部分球員在客場的罰球表現意料之外地比主場還要好。

澤聲：不過還是有球員，在主場優勢影響下，像 James Harden一樣在主場發揮較好。

俊聲：（不勝其煩地）你們慢慢討論，我先打一會籃球……

智翹和澤聲：豈有此理，等等我們啊！

(字數：2374 字)

參考資料：

[1] 球員數據

https://www.basketball-reference.com/

[2] 新聞選段

https://m.sina.com.hk/news/article/20190301/0/4/2/18%E7%BD%B018%E4%B8%AD%E5%A4%8F%E7%99%BB%E7%BD%B0%E7%90%83%E5%89%B5%E6%AD%B7%E5%8F%B2%E7%AC%AC%E4%B8%80-%E7%A7%91%E7%A5%96%E8%89%BE%E9%83%BD%E4%B8%8D%E5%A6%82%E4%BB%96-9837719.html

[3] 維基百科: 主客場制

[4] 球員照片

- https://zh.wikipedia.org/zh-tw/%E5%85%8B%E9%9B%B7%C2%B7%E6%B9%AF%E6%99%AE%E6%A3%AE
- https://zh.wikipedia.org/wiki/%E5%8B%92%E5%B8%83%E6%9C%97%C2%B7%E8%A9%B9%E5%A7%86%E6%96%AF
- https://zh.wikipedia.org/wiki/%E5%87%AF%E9%87%8C%C2%B7%E6%AC%A7%E6%96%87
- https://zh.wikipedia.org/wiki/%E9%81%94%E7%B1%B3%E5%AE%89%C2%B7%E9%87%8C%E6%8B%89%E5%BE%B7
- https://zh.wikipedia.org/wiki/%E6%96%AF%E8%92%82%E8%8A%AC%C2%B7%E7%A7%91%E9%87%8C
- https://zh.wikipedia.org/zh-tw/%E8%A9%B9%E5%A7%86%E6%96%AF%C2%B7%E5%93%88%E7%99%BB

優異作品:

一擊全中

學校名稱：香港培正中學

學生姓名：翟凱澄、焦采溢、 蔡卓霖

指導教師：梁偉雄老師

摘要 ：

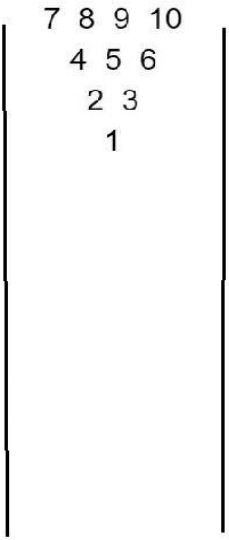
此專題透過海晴和教練的對話，介紹保齡球的各種知識，例如球瓶排列方法和計分方法，而且以水準分類，分析各類型的保齡球賽，包括學界、香港、亞洲和世界賽。透過計算每類比賽的程度，我們分析海晴最適合參加的組別。

五年前：

教練：歡迎你加入保齡球會，

海晴：請教練多多指導，我會努力學習。

教練：首先，你要清楚球瓶的排列號碼。排列方法是這樣的：



這裏的直綫是代表球道。你看，其實保齡球以三角形的形狀排列。當中，每個數字是代表每個球瓶的號碼。如果用這些號碼來溝通，這樣會更加具體，而且更加清楚打不中哪一個球瓶。

海晴：我知道了。原來球瓶也有編號的，真是十分方便。

海晴：那麼比賽是如何計分呢？

教練：讓我給你講解一下吧，保齡球比賽的計分方式：

原則

1 每一局共十格，依序完成每一格。

2 每格的分數將累計到下一格。

3 每一格以兩球以內將全部十個球瓶擊倒為原則。

海晴：教練真的謝謝你！我學會了很多。

教練：另外，還有不少比賽規則要注意的 … …

今天:

海晴加入保齡球會 5 年了，他學會了怎樣使用不同的技巧打保齡球。經過海晴努力的練習和教練耐心的教導，海晴逐漸提升自己的水準，甚至超越其他的隊員。上星期，教練認為海晴的水準良好，足以參加外面的比賽。

以下是他的一次練習記分：

局	1	2	3	4	5	6	7	8	9	10
失球號碼	10	X	X	4，7，10	7，10	X	X	10	X	X

'X' = 全中

數字代表失球號碼，例如當中第 5 局的失球號碼為 7，10: 即是只有 7 號及 10 號球瓶失中

教練：海晴，你這一局表現不錯呀! 你參加了保齡球會已經有一段日子了，而且你在眾多球員中是表現最好的一個。我邀請你參加 2019 年舉行的保齡球賽。

海晴：謝謝你的欣賞，我真的十分榮幸，但我如何知道自己的水準是否適合參加那類比賽呢?

教練：參加比賽前，應該了解一下不同比賽的水準。以下是近期一些保齡球賽的賽果，我們以比賽的種類而分類，例如香港賽、亞洲賽、世界賽等。我只是記錄了第一球的結果，因為我還未需要看球員怎樣打第二球。我以這些數據作參考，來找出最常出現的殘瓶組合，即是最容易剩下哪一個/幾個球瓶未能打中。

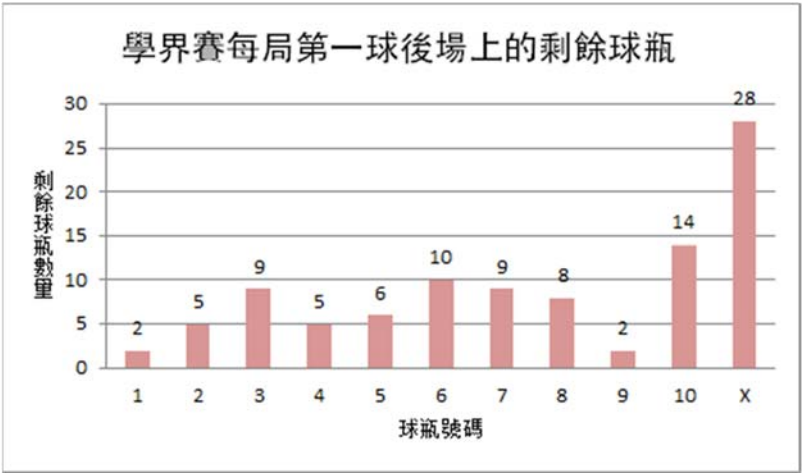
我們也會在每個類型的比賽數據之後加上棒形圖和分析，方便分析比賽的程度和最常打不中的球瓶。

學界賽：

	2016 全港學界保齡球公開賽美樂雙人賽初賽	2018 Junior Gold Boys U12 Final		2018 Sean Yonan Memorial Youth Championships		2014 Teen Masters Grand Championship	
失球號碼							
1.	10	3	7	X	2	X	3,6
2.	5,7	3,10	1,2,4,6,10	X	X	X	4,7,10
3.	X	X	X	X	7,10	X	7
4.	4,5,10	3,6,10	2	8	X	3,6	5,8
5.	5,6,8,9,10	5	3,6	X	X	X	4,6
6.	7	/	X	X	X	10	1,2,10
7.	5,8,9	/	/	X	X	X	7

	2016 全港學界保齡球公開賽美樂雙人賽初賽	2018 Junior Gold Boys U12 Final		2018 Sean Yonan Memorial Youth Championships		2014 Teen Masters Grand Championship	
8.	8	7	/	X	3,6	10	X
9.	X	4,10	X	8	2,8	3,6,10	3,6,10
10.	8	/	7	X	X	X	X

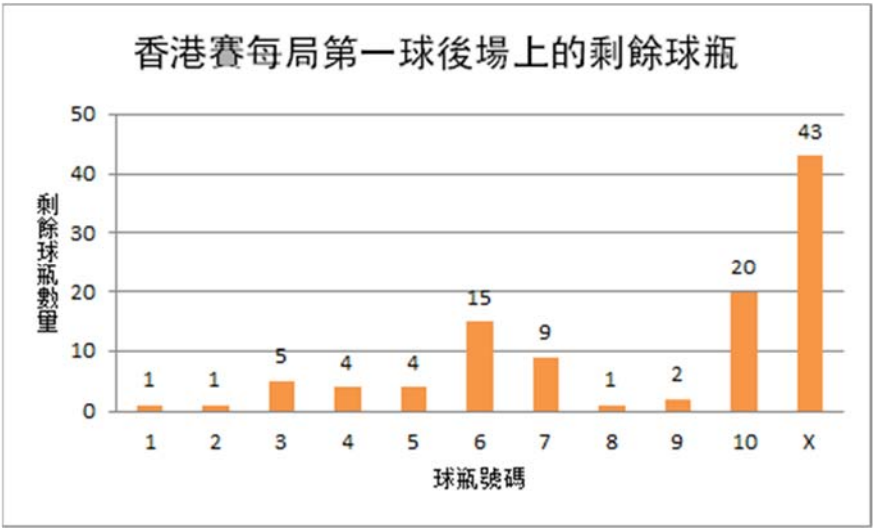
*由於 2018 Junior Gold Boys U12 Final 中沒有足夠的數據，有些失球號碼不能確定，所以有些位置漏空了。



從上圖可見，十號及六號瓶不被擊中的機會率是最高的，隨後就是三和七號瓶。選手一共發了 70 球，而全中一共出現 28 次，全中的概率為 = 28/70 = 0.4。 選手一共失了 14 次十號瓶，失球的概率是 = 14/70 = 0.2；其次是六號瓶，一共失球 10 次，失球的概率是 = 10/70 = 0.143。由於全中的概率很低，只有 0.4，所以水準一般，而且偏差。

香港賽：

	2016 Samsung 第 59 屆體育節保齡球錦標賽雙人賽		香港保齡球公開賽 2017 Bowling Men Double		2017ABF Hong Kong Men's Semifinal 02		2017 ABF Hong Kong Men's Final	
失球號碼								
1	10	X	4,6,7,10	X	6,10	10	3,6,10	4,6
2	X	10	X	10	X	3,6	X	X
3	6	9	X	X	X	X	10	7
4	7,8	X	X	X	X	7	X	6,10
5	10	X	X	6,10	X	X	7	X
6	X	X	3,6,10	X	6	X	X	X
7	X	6,9	X	X	4,6	X	6,10	7
8	6	X	X	3	X	10	X	10
9	X	5,6,10	10	X	X	X	X	X
10	5,10	5,7,10	X	1,2,3,5	10	10	4,7	X

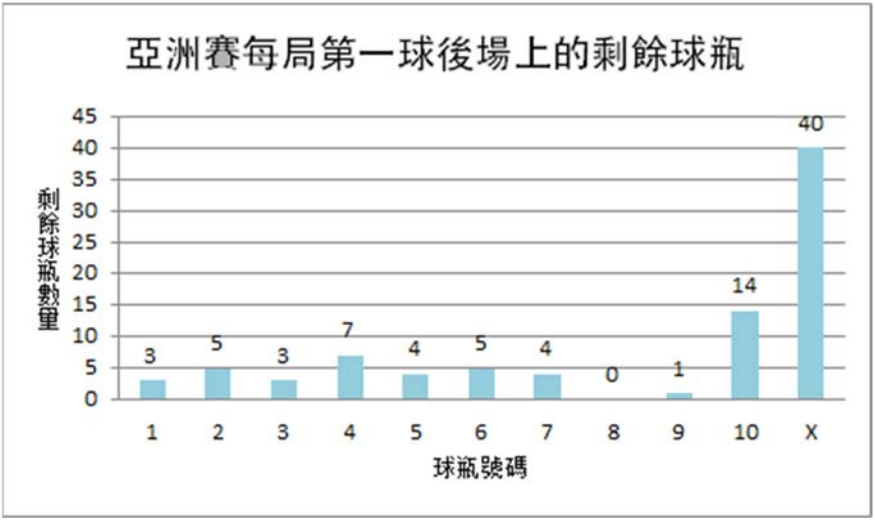


從上圖可見，十號瓶是最難擊中的，而隨後就是六號瓶。

選手們一共發 80 次球，而共有 43 次出現全中，因此全中的概率 = 43/80 = 0.538。所有選手共失了 20 次 10 號瓶，所以失球概率 = 20/80 = 0.25。其次是 6 號瓶，失球概率 = 15/80 = 0.188（準確致 3 個有效數字）。由於全中的概率只有 0.538，所以水準一般，但比學界賽的水準高。

亞洲賽：

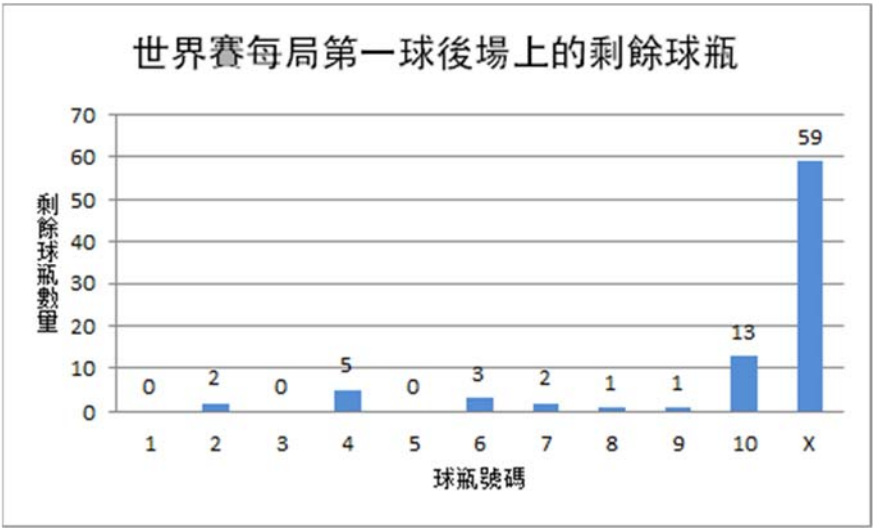
	2018 Macao-China International Open Tenpin Bowling Championships Women Final		2018 雅加達亞運男子保齡球菁英賽決賽 game2		16th Chinese Taipei International Open Men Final	44th Hong Kong International Open Men semifinal 02	
失球號碼							
1.	10	X	X	X	X	1,2,5	3,6,10
2.	4	X	X	X	5	X	X
3.	4,7	X	X	X	5,10	10	X
4.	10	X	10	X	10	X	X
5.	1,2,4,10	10	2	X	9	7,10	3,6
6.	X	4	X	X	10	3,6,10	2
7.	7	X	10	X	X	X	X
8.	1,2,4	4	4	6	X	X	X
9.	X	X	X	X	7	X	X
10.	X	X	10	X	5,6	X	X



在亞洲賽方面，選手們總共發了 70 次球，當中有 40 次全中，最常不能擊中的球瓶是 10 號，其次是 4 號。全中概率 ＝40/70＝0.571 (準確致 3 個有效數字)。選手總共失了 14 次 10 號瓶，失球概率 ＝14/70＝0.2。其他的失球號碼的分佈平均水準：由於全中的概率一般，只有 0.571，所以水準一般，比香港賽好。

世界賽：

	2017 World Bowling Men Double Final		2017 保齡球世錦賽－團體台北對美國 game 1		2018 Qubica AMF World Cup Final (Men)		2018 PWBA World Bowling Tour Championship Final	
失球號碼								
1.	X	X	2	X	X	X	X	X
2.	X	X	7,10	X	7,10	X	4	X
3.	X	X	X	X	4	X	X	X
4.	4	X	4,6	X	X	X	2	10
5.	X	X	X	X	X	10	10	10
6.	X	X	X	6,10	9	X	X	X
7.	X	X	X	X	X	X	X	X
8.	X	X	X	X	X	X	X	X
9.	6,10	10	X	X	X	X	X	10
10.	X	X	8	X	X	4,7,10	10	10



從上圖可見,十號不被擊中的機會率是最高的,隨後就是四和六號瓶。選手一共發了 80 球，而全中一共出現 59 次，全中的概率為 59/80 = 0.738。而選手一共失了 13 次十號瓶，失球的概率是 =13/80 = 0.163 其次是四號瓶，一共失球五次，失球的概率是 = 5/80 = 0.063。由於全中的概率較高，有 0.738，所以水準偏好，比所有球賽的水準好。

結論:

按水準排列：世界賽>亞洲賽>香港賽>學界賽

從上述的水準排列可見，世界賽的水準是相對最高，普遍人會認同世界賽的水準是最高的（符合結果），從我們的調查可以見到香港賽與亞洲賽的水準差不多，而也能可見當中失球號碼出現比較多是 10 號球瓶，佔 61/300 ≈ 0.2。

各賽事打不中球瓶數目的平均數：

	學界賽	香港賽	亞洲賽	世界賽	海晴
總共失球數量	70	61	46	27	7
共發球數量	70	80	70	80	10
失中一個球瓶的機率	1.00	0.763	0.657	0.336	0.7
全中機率	0.4	0.538	0.571	0.738	0.6

＊失中一個球瓶的機率和全中概率取至 3 位有效數字

從上述計算推論：
學界賽：每次發球都會幾乎有一個球瓶失中
香港賽：每發一球，就有 0.763 的機會失中一球瓶
亞洲賽：每發一球，只有 0.657 的機會失中一球瓶
世界賽：每發一球，就有 0.336 的機會失中一球瓶

如果以這種方法來看，這四項賽事當中，世界賽的水準比其他都高，因而無論從全中機會率或者失球球瓶數目百分比來看都是世界賽的水平是最高的。

結局

教練：海晴，你應該參加香港賽，因為失中一個球瓶的機率與香港賽相約，所以你的水準與這項比賽差不多，而且全中概率比香港賽好，有機會勝出，因而最適合你參加。

海晴：哦！我明白了！教練，你真精明！！我會努力參加更多不同比賽，汲取經驗，謝謝教練！

(2497 字)

參考資料：

[1] 香港保齡球總會
http://www.hktbc.org.hk/results/results-ch.htm
[2] 亞洲保齡球協會
http://www.abf-online.org/results/result.htm
[3] 2018 亞洲保齡球巡迴賽影片
http://www.tdm.com.mo/c_video/play_video.php?id=38101
[4] 2018 ABF Tour Macau Women's Final
https://youtu.be/kpDqf2AmI-s
[5] 2018 ABF Tour Hong Kong Men's Semifinal
http://www.csa.edu.hk/~pe/bowling/means.htm
[6] 保齡球計分方法 (新)
https://www.sportsroad.hk/archives/118529
[7] 2016 全港學界保齡球公開賽美樂雙人賽初賽
https://www.youtube.com/watch?v=y7TOqRwTHz
[8] 2018 Junior Gold Boys U12 Final
https://www.youtube.com/watch?v=xeN47OCOApM
[9] 2018 Sean Yonan Memorial Youth Championships
https://www.youtube.com/watch?v=p_m5nezm9sk
[10] 2014 Teen Masters Grand Championship
https://www.youtube.com/watch?v=tN0Vx_mGwt0
[11] 2018 Macao-China International Open Tenpin Bowling Championships Women Final
https://www.youtube.com/watch?v=Q6s0svkUK_g
[12] 2018 雅加達亞運男子保齡球菁英賽決賽 game2
https://www.youtube.com/watch?v=CKEXt_lqkEw
[13] 16th Chinese Taipei International Open Men Final
https://www.youtube.com/watch?v=DhguO4dXyHA

[14] 44th Hong Kong International Open Men semifinal 02
https://www.youtube.com/watch?v=qgACeXDYmXY
[15] 2016 Samsung 第 59 屆體育節保齡球錦標賽雙人賽
https://www.youtube.com/watch?v=EAT5P7RxZ7o
[16] 香港保齡球公開賽 2017Bowling Men Double
https://www.youtube.com/watch?v=O8T3gwXHRnA
[17] 2017 World Bowling Men Double Final
https://www.youtube.com/watch?v=GBjXw93Aa3Y
[18] 2017 保齡球世錦賽－團體台北對美國 game 1
https://www.youtube.com/watch?v=cVWAZJ3A-0s&t=770s
[19] 2018 Qubica AMF World Cup Final (Men)
https://www.youtube.com/watch?v=MVDR1QVUw9o&t=120s
[20] 2018 PWBA World Bowling Tour Championship Final
https://www.youtube.com/watch?v=P5Iq2tnBZX8
[21] 2017ABF Hong Kong Men's Semifinal 02
https://www.youtube.com/watch?v=U4Irftdo0bU

優異作品：

The crime journey of the three little pigs

School Name: Pui Ching Middle School

Name of Student: Yao Yuan Yue, Leung Tsz Wai

Supervising Teacher: Mr. Leong Wai Hong

Abstract:

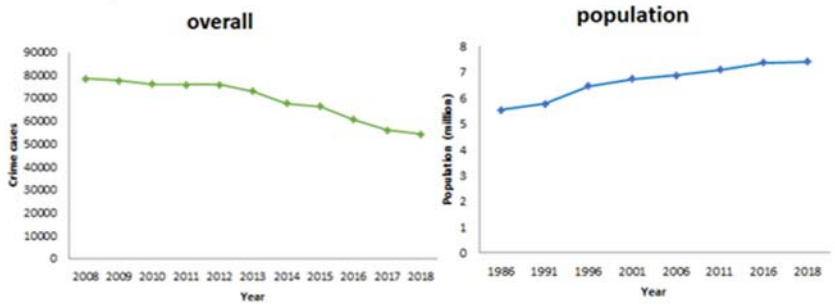
The three little pigs travelled to Hong Kong and went to a carnival. The two smaller pigs were constantly deceived with various ways. Their big brother then explained to them with statistics and they finally learnt a lesson.

Once upon a time, in a certain dimension, there lived the famous three little pigs. After they fought big bad wolf, they slowly became wealthy, and decided to travel around the world to broaden their horizons.

The three little pigs actually had names: the big brother was called Pingda, the next boy was called Pingy and the little girl was called Pinky.
Their first stop was Hong Kong.

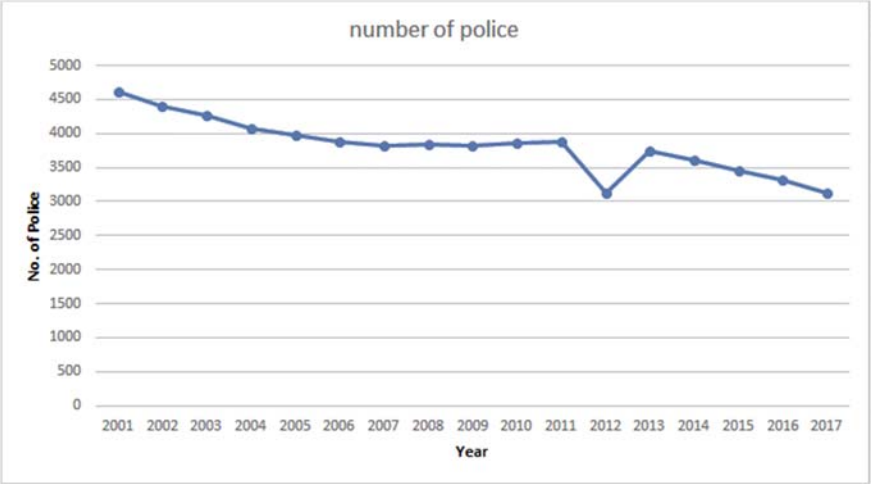
The smartest Pingda first warned his siblings, “We have to take great attention to our belongings. It is not that safe in this place, we have to be cautious all the time.”

“Really? I thought the crime rate had been dropping continuously since 2008, so it should be quite safe instead.”

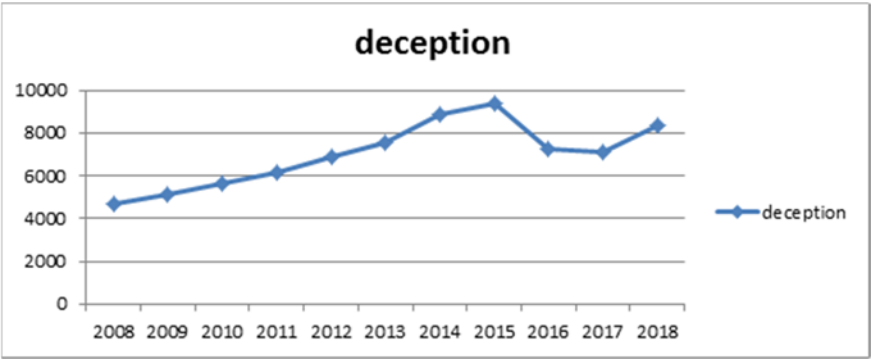


“Well you are correct. Look at the crime cases from 2008 to 2018, you could indispensably realised that the crime has kept dropping, From 78469 cases in 2008 drops to 54225 cases in 2018, which drops in a significant 30%. At the same time, the population since 1986 has been booming nonstop, there are 5524600 citizens in 1986 and it increases to 7391000 citizens, which shows a 33.8%. It is crystal clear that the relationship between population and overall crime has shown negative correlation, we only have as what my little brother just said, that means one people will only have 0.007337 chances to encounter a crime. Hong Kong is basically becoming a safe place to live.”

There was police patrolling around. So Pingda explained, “There used to have huge number of police in Hong Kong, but it has decreased over the years. But the quality of the police has improved as well to support the cut down of police.”



“But, listen carefully. The number of reported cases of all crimes had been decreasing except one- deception. In 2008, there was only 4653 cases, but after ten years it is increased to 8372 cases. It dropped a while during 2016 but the cases has risen again.”



The three little pigs decided to go to the AIA The Great European Carnival to play. It was Sunday, therefore there was a lot of people queuing to buy tickets. As the line never seemed to move, Pinky soon lost her patience, and ran around.

She was distracted by a man wearing a uniform and standing nearby. “The queue is so long. Come and buy instantly tickets here!” He yelled towards the crowd.

Wanting to go into the carnival to play as soon as possible, she approached the man. The man handed the girl pig a tablet, and asked her to fill in some information in their website to buy the ticket.

The website looked legitimate, with logos of AIA. Pinky filled in her personal information without any hesitation and the man said, “You will receive an email soon and with that email you could get into the carnival.”

Pinky ran to Pingda and shared her joy with him, but he was not happy at all.

“You were just deceived!!!!!! I told you to be cautious all the time, but you didn’t even pay a tinge bit of extra attention. Now, what information did you fill in?” Pingda scolded her and asked.

“It’s not really a lot. I filled in my ID number, email, address, name, your contact information…”

“Stop! That is obviously just a deceiver with a fraudulent website. You didn’t even fill in payment options or related stuff, how could you just simply fall into this flagrant trap!” He exclaimed.

Pinky felt sad, so she decided to scroll her phone to consume time. Soon, an email notification popped out. She clicked inside and was delighted to discover that it was the email that the man mentioned.

The whole email seemed formal, with proper registers and closings. The structure was clear and diplomatic, and at the ending of the email, it wrote, “Click into the link to pay for the tickets. To let your time in AIA Carnival be more well spend, download our app here.”

With no indecision, she clicked into the link, and a proper form of XYZ

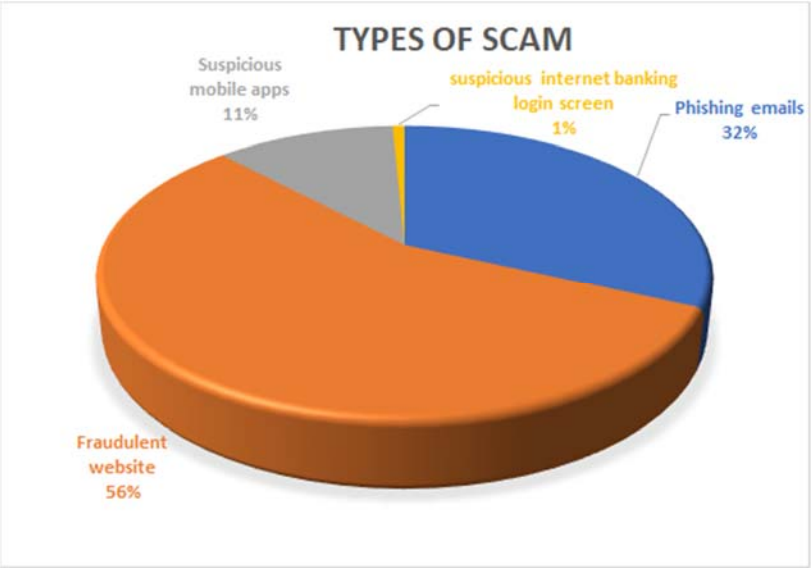
bank appeared in front of her eyes. The big brother, who was glancing at her, observed the whole process.

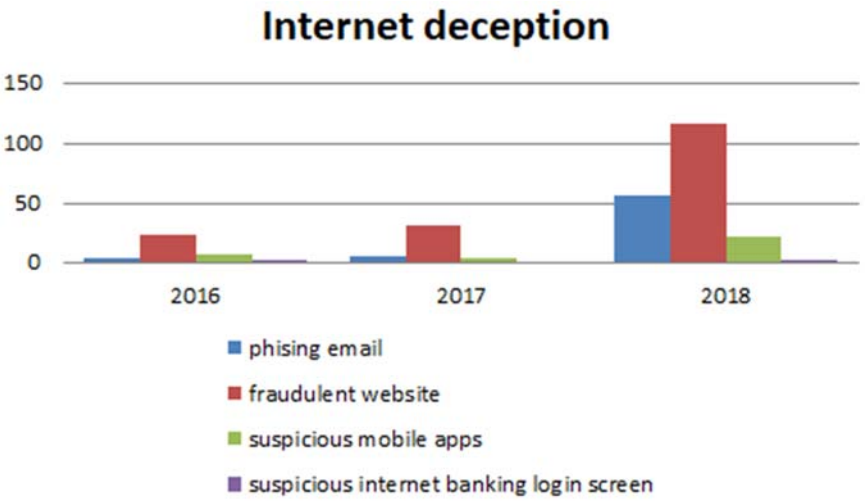
“STOP!!!!”, he shouted to her, “That’s obvious a scam as well! You just received a scam and clicked into a website with a suspicious internet banking login screen. The other link is likely a suspicious app as well. ”

Pinky was obviously not convinced. Those emails and websites looked official and trustworthy. Her other brother, Pingy, laughed at her as well. Irritated by her brother’s attitude to her, she clicked into the other link of downloading their app. Soon she was directed into an unknown website, with advertisements popping out interminably. Her phone lagged and she needed to reboot it.

Her big brother, who gave up stopping her, looked at her throughout the whole process. She was finally convinced that she was just deceived. Pingda started to explain,

“Actually deceiving is quite common nowadays, living in high-technology era. The deceivers are adroit in using the internet to create deception. That was a fraudulent website which you had a 56% chance to face if you were trapped in internet deception. So over half of the victims made the same mistake as you. In fact, if you hadn’t clicked into the website, you had a $\frac{8}{11}$ chances to get to phishing email.”





“We can find out that the internet has a sharp increase from 2016 to 2017, especially the phishing email has increased a sharp 102% and fraudulent website has increased 26.25%, it is easy to guess that the internet deception is rising in the future.”

It was finally their turn to buy tickets and enter the carnival. Suddenly, Pingy's phone rung.

"Hello, do you remember me?" A women asked.

Pingy guessed from the voice that it might be his aunt, so he replied, "Are you Auntie? "

"Yes of course I am. Actually a car accident just happened and I got injured…"

"What!!??" Pingy panicked.

"The problem is that I don't have enough money with me, could you lend me some money by transferring some to my bank account? It's so urgent! Please help me!!!" The women yelled at the phone.

Pingy was worried. He loved his aunt a lot, therefore he immediately asked Pingda for help. But Pingda proclaimed, "You just received a "Guess who I am" telephone deception! We didn't even tell her that we were travelling to Hong Kong and got a new phone card! You should calm down and analyze the facts before panicking!"

Pingda explained to him patiently, " This type of deception takes up 0.51 chance if you get into the telephone deception, the deceivers will wait for your response to deceive who he or she is, then they will fabricate a story, like being injured one ask you to pay for them. You need to be vigilant to any calls."

The three little pigs decided to play the spinning coaster. Unfortunately, the queue was long as well and they needed to wait.

“Ring…” This time it was Pinky’s phone ringing.

“I’m from the Customs and Excise Department. I’m calling to inform you that your package had some problems so it is kept here. If you want that package back, you’d better transfer a decent amount of money…”

Pinky was frightened. She sobbed, and asked Pingda for help, “My package is detained by the customs and I need to pay for it to pass…”

“...It’s fake.”, Pingda replied instantly, “Customs in Hong Kong wouldn’t ask you to pay for packages that have problem. Problems can’t be solved by money. Nowadays, many deceivers pretend to be officers, and try to deceive people’s money. The chance of getting into these crime is quite high: 0.46. It is just slightly lower than falling into “Guess Who I Am” cases.”

Pinky was feeling unwell so she went to the toilet.

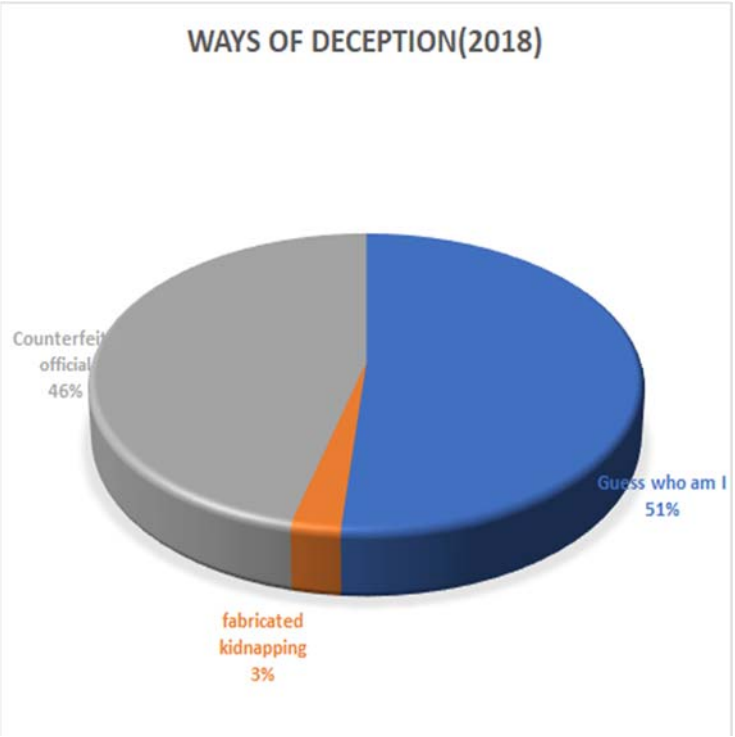
Suddenly, Pingy’s phone rung.

A vague voice came out from the phone, “Your friend is in extreme danger now. She is kidnapped by me and I would only release her if I receive $444444. Now if you dare scream….”

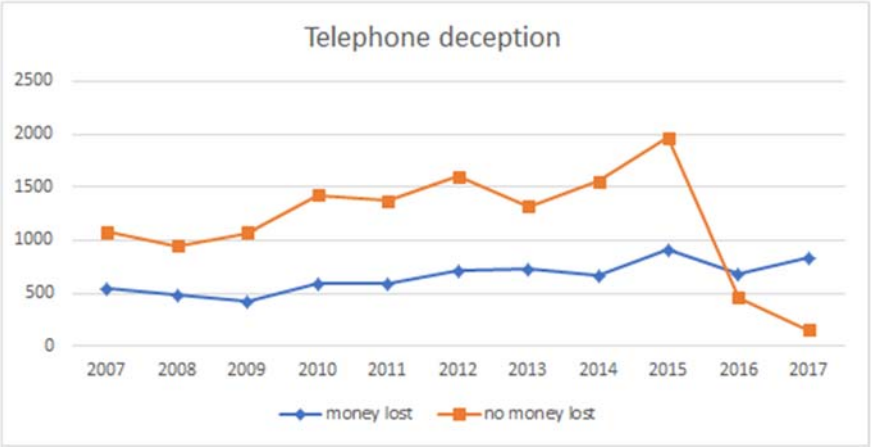
“AHHHHHHHH!!!!!!!” Pingy screamed since he was apprehensive.

Seeing Pingy acting crazy, Pingda snatched his phone and listened to the phone call. He immediately called off the phone and calmed him down, looking at Pinky who just stepped out from the toilet, “That was simply a scam. It was a fabricated kidnap deception. You need to learn how to discern which phone calls are real and which aren’t.”

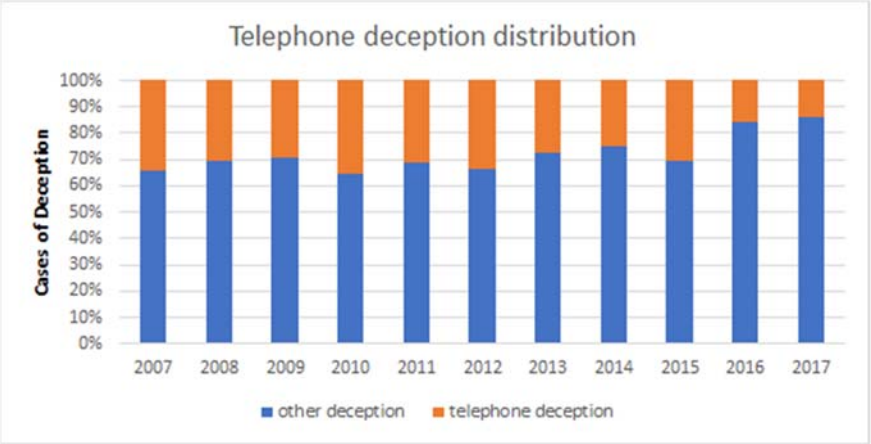
“This is the swindle of fabricated kidnapping and the swindler pretends to kidnap your relatives or friends. Only around 0.03 people will get tricked and deceived, as it is unrealistic.”



You two just received several types of telephone deception. Actually 85% of the victims would lose money in 2017, so you should be lucky that I stopped you. Normally, most people don't lose money, but starting from 2016, the numbers started to reverse."



Pingda kept on talking, "As the telephone deception is quite common, we can analyse the distribution of telephone deception cases among all the cases, from 2015 to 2017, the telephone deception is taking a less proportion with 30.8% in 2015 to 13.8% in 2017, we will encounter the telephone deception less likely suppose we get into a scam."



After a long time, it was finally their turn to play the spinning coaster. The spinning coaster made Pinky dizzy and felt sick, so they rested aside.

Suddenly, a women approached them, “Hi, little pig, you seem to be sick. I can observe that some evil spirits invading her. If you don’t clear them, bad luck would stick to her. However, with my unique blessing, the spirits will disappear into ashes…”

“Shut up!”, Pingda shouted to the women, “Stop deceiving, I know that you would then ask for money.”

The women ran away and Pingda said, “Now that kind of deceiving is called spiritual blessing gang, they will tell superstitious stuff, and once you believe you would fall into the fraud to pay money and buy the “medicine”. ”

“If you occur street deception, there is 0.6 chance you will fall into the above deception and mostly elder suffer so. So, that is why I think you are

a bit idiotic…" Pinda shaked his head helplessly.

"Hey little piggy, you screamed deep from your throat like ten minutes ago," Said a passing by man," Screaming harms your throat."

The man then took a leaflet out and pointed to it, "This medicine would fix all your sore throats. With just $1234, you could buy one. I guarantee you that your throat would feel much better after taking just a pill... ".

Pingda literally rolled his eyes, and spoke, "Please, sir, we are not idiots. You are selling fake medicine. You are deceiving innocent kids. Now please leave or else I would call the police!"

Pinky and Pingy was surprised that that man was a deceiver as well.

"Street deception includes medicine gang, and also electronic component deception, intentionally dropping money, these three cases have taken up 40% chance if one has encountered telephone deception."

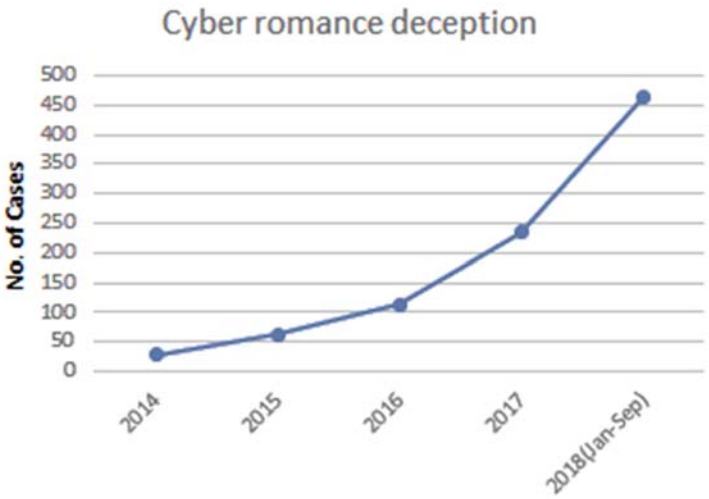
Suddenly, Pinky asked Pingda for some money, "Could you lend me $1000? My friend need some money urgently."

"Who is her? How come I didn't know the whole thing?" Pingda questioned.

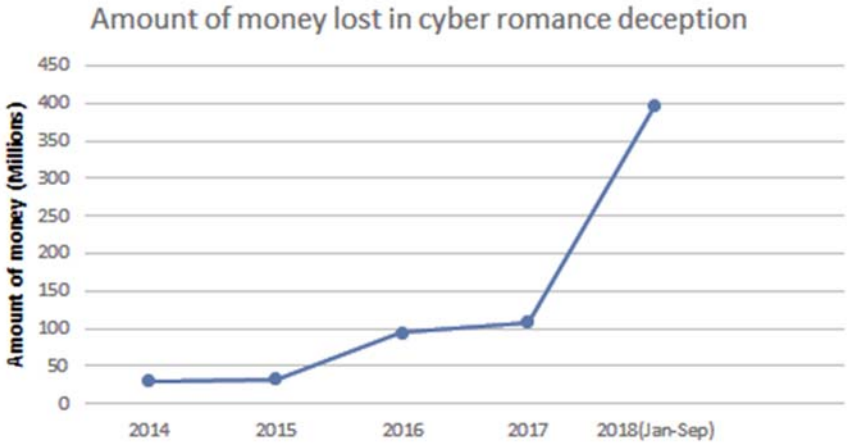
"It's a friend that I met online. He is a super talented pig and he is so handsome as well, he is so attractive! Please could I borrow some money from you?" She replied with her childish voice.

“I won’t, as he is obviously a deceiver as well. You don’t even know him in real life, how could you ensure that he is trustworthy? You can’t even know if he is really male. You just experienced another type of online scam: the online romance scam.” Pingda replied, “Online romance scam is when someone attracts another via internet and try to deceive their money, by using their conscience.”

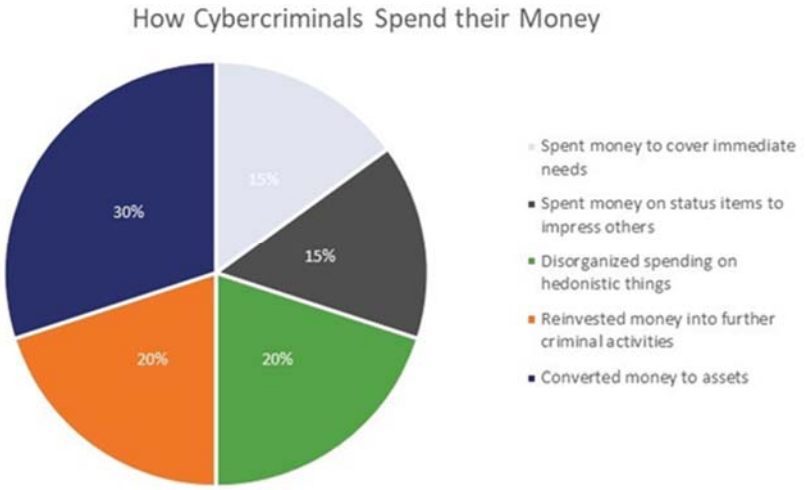
Pingda searched online, and continued, “It is also known as cyber romance deception. This scam is one of the scams that is significantly increasing these years, as social media is getting more prevalent these days.”



“Back in 2014, there were only 29 cases the whole year. But in the already incomplete data of 2018, which only consisted data from January to September, is already 463. That is a nearly 16 times difference!”



"Following the rise of cases, the amount of money lost due to them has also increased a lot. We could all obviously observe a sharp increase of money involved during 2017 to 2018. Do mind the chart actually did not have all data of 2018 yet, but the increase is already noticeable."



"It is interesting to notice the selection of what the cybercriminal will do to spend their money. 30% of them will convert money to asset, which is

quite a silly action as the police can easily check their accounts and they will be discovered of doing illegal trading. Don't fall into traps and let them spend your money!" He concluded.

After playing in the carnival for the whole day, the three pigs were exhausted. They walked out from the exit and met the man carrying a tablet that deceived Pinky the start of the day. Seeing the three pigs walking around, he ran away, and Pinky moaned.

Pinky and Pingy had surely learnt a valuable lesson that day. They were deceived for so many times, and they experienced many different kinds of deception, including clicking into suspicious website and phishing email, getting telephone deception such as "Guess who I am", "Counterfeit official", and "fabricated kidnapping". Last but not least, they also encountered cyber deception included romance deception, blessing deception and medicine deception.

They had learnt to pay attention to all the opportunities and lucky chances they we given, and to be aware of all suspicious people approaching them. Even if something seemed to be legitimate, they knew that they should think more before completely believing what others claimed to be true.

(2489 words)

References:

[1] 立法會秘書處 資料研究組-保安 https://www.legco.gov.hk/research-publications/chinese/1718issh13-crime-and-telephone-deception-20180227-c.pdf

[2] Police review https://www.police.gov.hk/ppp_tc/01_about_us/police_review.html

[3] Hong Kong Monetary Authority https://www.hkma.gov.hk/eng/smart-consumers/beware-of-fraudsters/#fraudulent-bank-websites-phishing-emails-and-similar-scams

[4]立法會 CB(2)949/17-18(03)號文件 https://www.legco.gov.hk/yr17-18/chinese/panels/se/papers/se20180306cb2-949-3-c.pdf

[5] 為金錢服務經營者舉辦的打擊「洗黑錢」研討會 https://www.fstb.gov.hk/fsb/aml/tc/edu-publicity/seminar2018/Vulnerability%20of%20MSOs%20shownin%20Telephone%20Deception%20Cases.pdf

[6] THINK HONG KON 數讀香港：網戀騙案飆升 41-50 歲最多人中招 http://www.thinkhk.com/article/2018-z10/25/30041.html

[7] Hong Kong Police Force https://www.police.gov.hk/ppp_tc/09_statistics/

[8] Three little pigs photo https://www.londonnewsonline.co.uk/an-imaginative-retelling-of-the-classic-fairy-tale-three-little-pigs-go-west/

[9] Three little pigs houses https://hometipsforwomen.com/insulation-and-the-3-little-pigs

[10] Wolf photo https://www.123rf.com/photo_87122489_stock-vector-cartoon-three-little-pigs-big-bad-wolf-blowing.html

優異作品:

NBA 勝率大謎團

主場客場逐個捉！

學校名稱：沙田官立中學

學生姓名：方貴堯、黃浩賢

指導教師：陳世雄老師

摘要:

近日 18-19 年度的 NBA 賽事開始了不久，貴堯和浩賢一向是 NBA 的狂熱粉絲, 今次也毫不例外，十分期待新賽季的賽事。今天數學堂結束後，他們便留在課室討論昨天的賽事，並打算運用所學的統計學和概率來研究主客場對勝率的影響。

引言

NBA(National Basketball Association)是不少同學喜歡的體育項目，但你們知不知道有甚麼因素會影響比賽的結果呢? 有人認為球隊在主場出賽會有優勢；又有人認為這只是純粹巧合,究竟真相是如何? 本報告將以概率和統計學角度進行分析，找出主客場是否真的是球隊獲勝的關鍵。

他們的對話如下~

堯：浩賢，你有沒有看最近11月20日巫師對快艇的比賽呀？

賢：沒有呀.....上次10月28日快艇贏得太輕鬆了。相信這次不用看就知道結果了。

堯：咦？可是這場巫師大勝呢。

賢：竟然？我還以爲快艇比起巫師實力高上一大截呢。

堯：難道你不知道主場優勢足以對賽果做成影響嗎？

讓我們以這兩場的數據做一些分析

華盛頓巫師

日期	得分	入球數量	命中率 $(\frac{\text{入球數量}}{\text{射球數量}})$	3 分球數	3 分球命中率 $(\frac{\text{3 分球數量}}{\text{射球數量}})$	失球	偷球	封阻
20-11-18 主場	125	40	43.0%	13	33.3%	7	11	8
28-10-18 客場	104	38	39.2%	5	18.5%	11	7	3

洛杉磯快艇

日期	得分	入球數量	命中率 $(\frac{\text{入球數量}}{\text{射球數量}})$	3 分球數	3 分球命中率 $(\frac{\text{3 分球數量}}{\text{射球數量}})$	失球	偷球	封阻
28-10-18 主場	136	56	54.4%	15	55.2%	15	7	9
20-11-18 客場	118	46	51.7%	19	40.7%	19	6	4

堯：你看，主場的優勢不但能在一些表面的得分中顯示出來，甚至在雙方的失誤控制亦能反映出來。

賢：看來即使到了NBA球員的級數，在一些陌生的地方也不能好好發揮自己。

堯：其實不只是因爲對球場的不熟悉，還有球迷的支持度。當所有球迷都為你的球隊歡呼時，你打起球來也會分外得心應手；相反在客場中大多球迷都會在支持對手時，你也會較爲緊張，導致更容易犯錯。有時候就是稍一分心，球就投偏了，導致命中率下滑。

賢：我明白了。可是這不能完全反映真相吧，這最多能算是個別案例吧。

堯：對。所以讓我們以整體的數據來看一下吧。讓我們分別以上中下游的球隊作例子吧。

球隊	排名	整體勝率	平均得分	主場勝率	主場 平均得分	客場勝率	客場 平均得分
多倫多速龍	東區第 1	71.4%	114.3	76.5%	116.2	66.7%	112.5
丹佛金塊	西區第 1	67.7%	110.0	81.3%	112.8	53.3%	107.1
費城 76 人	東區第 4	62.9%	114.5	84.2%	117.4	37.5%	111.1
波特蘭拓荒者	西區第 6	55.9%	111.0	72.2%	113.7	37.5%	107.9
華盛頓巫師	東區第 11	38.2%	113.2	60.0%	118.7	21.1%	108.9
達拉斯獨行俠	西區第 11	46.9%	110.7	81.3%	113.6	12.5%	107.8
克利夫蘭騎士	東區第 15	23.5%	102.8	27.8%	104.7	18.8%	100.6
鳳凰城太陽	西區第 15	23.5%	104.6	31.3%	103.0	16.7%	106.0

(數據來源截至 2018-2019 年賽季聖誕大戰前美國時間 12 月 25 日)

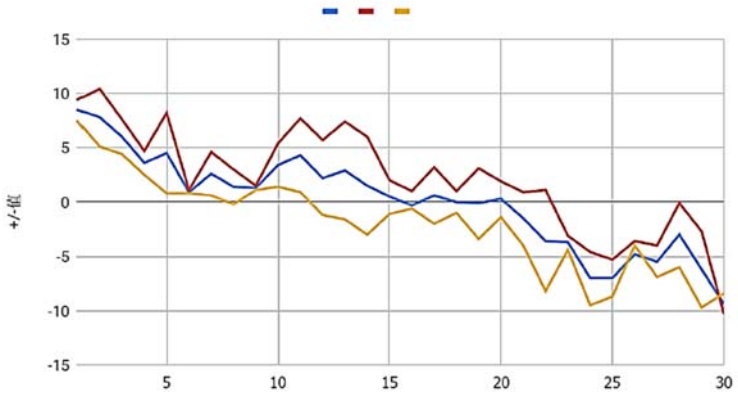
堯：從數據中我們可以得出主場有優勢並不是單一事件。我們以排名較前和排名較後的球隊作觀察，均可以發現主場的勝率明顯比客場的勝率較高，而數據中除了太陽以外，所有球隊在主場時的得分都比起客場的高，因此我們可以歸納出所有隊伍在主場都會打得更好。

賢：可是只考慮某些比賽不能反映所有賽事都是對主場有利呢[1]。

[1] 詳情請參考附錄一。

堯：讓我來統計一下吧!

2017-2018 主客場數據統計



（圖一）

備注：藍色綫條代表球隊在主場和客場的總平均正負值

紅色綫條代表球隊在主場的平均正負值

黃色綫條代表球隊在客場的平均正負值

*X*軸代表球隊由第1名到第30名的排名，詳細名次於附錄三中可以細閱

*Y*軸代表球隊的正負值

賢：這個圖表足以反映出主場的優勢，除了鳳凰城太陽外，全部球隊也在主場中表現較好。

堯:對,從圖表中可以看到前半段的球隊在主客場的表現均較爲穩定，正負值[2]差距都在正負5以內，而中段的球隊在主客場的差距就較爲明顯，而後段的球隊在主客場的表現都變得較接近。

賢：總括而言，球隊在主場往往發揮得比較好，而在客場則會因爲各種原因而表現較差。

賢：但是我們有沒有其他方法來證明主場的優勢呢？

堯：好問題！我們來研究一下吧 ！

以下是球隊分別在主場及客場的勝率和總勝率

球隊	主場勝率(%)	客場勝率(%)	總勝率(%)
侯斯頓火箭	82.9	75.6	79.3
多倫多速龍	82.9	61.0	72.0
金州勇士	70.7	70.7	70.7
波斯頓塞爾特人	65.9	68.3	67.1
費城 76 人	73.2	53.7	63.4

[2] 正負值指的是球隊與球隊之間的得分差，例如球隊 A 大勝球隊 B 10 分，那球隊 A 的正負值則是+10，而球隊 B 的正負值是−10。以正負值來表達就可以更好的展示出球隊的表現，從而更好的分析主場優勢。

堯：我們可以利用條件概率來計算一下他們在不同情況下的勝率。

(假設 H 為球隊在主場的事件[3]，R 為球隊在客場的事件，W 為球隊獲勝的事件。)

以侯斯頓火箭為例[4]

我們可以得出：

(1) $P(W|H)=82.9\%$

(2) $P(W|R)=75.6\%$

(3) $P(W)=79.3\%$

利用以上的資料，我們可以計算出當球隊勝利時，他們在主場的機率

$$P(H|W)=\frac{P(H)P(W|H)}{P(W)}$$

$$=\frac{(\frac{41}{82})(82.9\%)}{79.3\%}$$

$$=0.5227 \text{ (取小數後四個位)}$$

同樣，我們可亦用一樣的方法找到當球隊勝利時，他們在客場的機率

$$P(R|W)=\frac{P(R)P(W|R)}{P(W)}$$

$$=\frac{(\frac{41}{82})(75.6\%)}{79.3\%}$$

$$=0.4767 \text{(取小數後四個位)}$$

堯：由此可見，當球隊勝利時，事件大多發生於主場。因此我們便能推斷出球隊在主場往往能表現更好。

賢：原來如此，我明白了！

---

3 每個賽季總共有82場賽事，主場有41場，而客場有41場。

4 其他球隊勝利時分別在主客場的機率可以在附錄二中查看。

陳：咦？你們兩個在討論甚麼呢？

堯：我們正以簡單的數據分析主客場對球隊勝率的影響。

陳：這種簡單的統計方法只能發現一些較表面的資料。你們有想過用一些較深入的方法研究嗎？

賢：願聞其詳。

陳：你們有沒有想過主場影響勝率的一些原因呢？

堯：是心理質素嗎...我知道了！關鍵時刻[5]可以衡量心理質素！

賢：當球賽進入關鍵時刻時，球員會因擔心失誤導致落敗而變得格外緊張！如果球員心理質素差，就會因為太緊張而被逆轉。

堯：對！因此衡量一隊球隊在關鍵時刻的表現就能衡量他們的心理質素。

賢：以下我們以邁亞密熱火隊[6]作例子，以熱火隊分別在主客場關鍵時刻時的表現作比較，看看主客場對心理質素的影響。

邁亞密熱火隊在 2017-18 賽季關鍵時候的表現

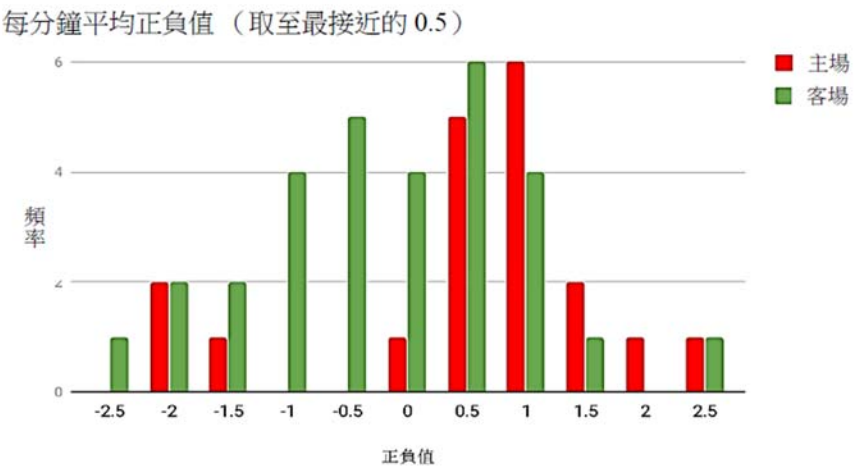
	勝率	得分	射球命中率	3 分球命中率	罰球命中率	籃板	助攻	失誤	偷球	封阻	+/-值
主場	52.94%	2.45	49.5%	39.4%	82.9%	0.95	0.45	0.26	0.11	0.10	0.72
客場	33.34%	2.18	45.9%	36.4%	77.3%	0.93	0.43	0.31	0.19	0.05	0.01

賢：不論在射球和投球時，熱火隊關鍵時刻在主場都有較好的表現，反而在客場表現非常失色。雖說大家都在相同的情況對賽，但在主場時球員往往有較好的心理質素，而且勝利的機率較高。客場時，雖是同一群球員，卻連最簡單的罰球也達不到 80%的命中率，可見心理質素對球員的影響。

5 『關鍵時刻』普遍指的是一場比賽最後的幾分鐘，而比數又只差得較少的時刻，在這裏我們會將其當作每場比賽如果到了最後5分鐘仍然相差5分以內的時刻。

6 聯盟排名第 15，在最中游

讓我們現在以邁亞密熱火在 **17－18** 年度的賽季中 **49** 次關鍵時刻中的正負值作一個圖吧[7]



從圖表中，我們可以看見到主場和客場時候正負值的分佈

	平均值	中位數	眾數	分佈域	四分位數間距
主場	0.553	1	1	4.5	1
客場	- 0.167	0	0	5	1.5

	標準差 (正負值<0)	標準差 (正負值≥0)	總標準差
主場	0.289	0.640	1.17
客場	0.663	0.676	1.13

7 由於數據過於極端，而且時間太短，因此進入關鍵時刻少於 2 分鐘的數據將不被納入計算

1） 主場時，熱火隊在關鍵時刻中的發揮比較集中在 0-2.5 的區域，有16場比賽中在關鍵時候可以表現出 0-2.5 的正負值表現，而16場當中他們更是贏了15場，可見在主場中球隊即使在比賽分數接近的情況也不會因爲心情緊張而輸掉比賽，冷靜地對賽獲得勝利。

2） 客場時，熱火隊的表現較爲不集中，其平均正負值爲正數和負數的機會是非常接近，分別為 40% 和 46%。從表格中可以看出，客場的總標準差雖然較主場的為低，然而只要我們分開少於0和大過等於0來看，主場的數字比客場為低，由此我們可以得出主場的總標準差高是因為特別有3項數據落在少於0的部分，相反客

場的數據分佈在各項，導致在少於0和大過等於0的部分標準差均比主場高，因此我們可以見到客場的表現較不穩定。

3） 而通過比較主場和客場的數據，我們可以發現熱火隊在主場時被拖入關鍵時刻的機會比起在客場時的機會遠遠為少。熱火隊在主場時只有經歷19次關鍵時刻，當中勝出了15場；反觀熱火在客場時30次被拖入關鍵時刻，卻只贏了14場。這顯示球隊在主場時能完全發揮自己的實力，代表對上比自己弱的隊伍會贏很多，而對上比自己強的隊亦會輸得比較多。反觀在客場時很靠運氣，偶爾會拖到關鍵時刻。

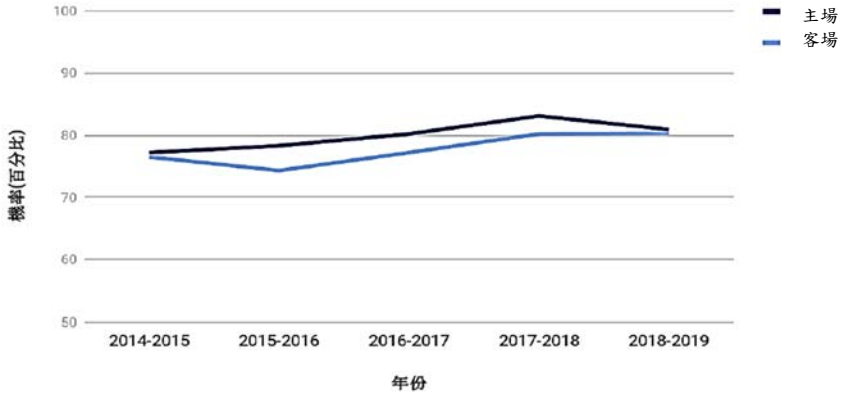
堯： 但罰球[8]也是球隊成敗的關鍵，主客場會否對罰球有影響呢？

賢：當然會啦，球隊在主場比賽往往會得到觀眾熱烈的支持，表現自然能更好啦！反而客場球隊卻會受到觀眾的唾棄聲，甚至用一些奇怪的物件分散注意力，令他們表現失準。

堯：讓我們用數據證明一下吧。

以下是金州勇士歷年在主客場罰球投入的機會:

金州勇士



（圖二）

堯：從以上的數據我們可以發現即使金州勇士擁有著5 名表現穩定的明星球員[9]，在主場時投進罰球的機會仍比客場高，雖然只是相差很少[10]，但這些小小的差別卻足以令球隊獲勝！因此可以推論出球隊在主場的罰球命中率會比在客場時高。這亦導致球隊在主場比起在客場會得到約多2-3 分，甚至可以改變形式。

陳：時間已經差不多了。總括來說，球隊在主場的表現遠比在客場時的好，因此勝率也會比較高。

(字數:2490)

8 罰球命中率往往是籃球比賽的關鍵，特別是最後時刻兩隊分差接近的時候，把握好每次罰球機會的球隊就能獲勝

9 5 名明星中包括罰球率最高最穩定的歷史性射手 Stephen Curry，歷史性 3 分射手 Klay Thompson ， 歷史性得分手 Kevin Durant ,全年度防守冠軍 Draymond Green 和現役籃底霸主之一 DeMarcus Cousins

10 在 2018-2019 賽季中罰球命中率只是相差了 0.6%

參考資料：

[1] NBA 官網 https://stats.nba.com/teams/

附錄一

1）雖然大部分球隊都在主場勝率比起客場的來得高，然而其實一定有一些例外，例如邁亞密熱火，在主場勝率只有43.8%，反而在客場有56.3%。這些數據我們亦應該在分析時歸納在內，不應該對其絕口不提。

2）以半個賽季（2018-19 年賽季聖誕節前的賽事）來做數據的分析比較不公平。首先，每隊球隊並沒有一個相同的賽程。可能有部分球隊在主場遇到的對手都是比較弱的，而在客場遇到的對手都較強，導致主客場有一個非常明顯的勝率差距。這樣會使數據變得過大，並不能真正反映問題。相反，如果以一整個賽季作爲分析，可以確保球隊之間至少交手一次，減小因爲賽程的不同而做成的不公平，而數據則可以真實地反映情況。

3）以得分多少來定論一隊球隊的好壞並不完整，即使一隊球隊在客場得分比較低，也不代表他在客場的表現比較差，可能那隊在客場時防守比較好，而進攻比較差，導致雖然得分較低，整體表現卻比較好。因此爲了可以較全面地展示球隊的表現，我們應採用正負值[11]。

11 參考頁六的註腳二

附錄二

以其他球隊為例

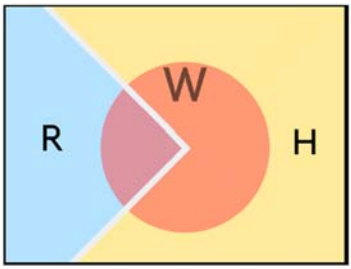
(以下均以小數點後四個位為結果)

多倫多速龍： $P(H|W)=0.5757$
$P(R|W)=0.4236$

金州勇士： $P(H|W)=0.5$
$P(R|W)=0.5$

波斯頓塞爾特人：$P(H|W)=0.4911$
$P(R|W)=0.5089$

費城 76 人： $P(H|W)=0.5773$
$P(R|W)=0.4227$



附錄三（頁六圖一）

球隊	平均正負值	主場正負值	客場正負值
侯斯頓火箭	8.5	9.4	7.5
多倫多速龍	7.8	10.4	5.1
金州勇士	6	7.6	4.4
波斯頓塞爾特人	3.6	4.7	2.5
費城 76 人	4.5	8.2	0.8
克利夫蘭騎士	0.9	1.1	0.8
波特蘭拓荒者	2.6	4.6	0.6
印第安納溜馬	1.4	3	−0.2
新奧爾良塘鵝	1.3	1.5	1.1
奧克拉荷馬城雷霆	3.4	5.4	1.4
猶他爵士	4.3	7.7	0.9
明尼蘇達木狼	2.2	5.7	−1.2
聖安東尼奧馬刺	2.9	7.4	−1.6

球隊	平均正負值	主場正負值	客場正負值
丹佛金塊	1.5	6	−3
邁阿密熱火	0.5	2	−1.1
密爾沃基公鹿	−0.3	1	−0.6
華盛頓巫師	0.6	3.2	−2
洛杉磯快艇	0	1	−1
底特律活塞	−0.1	3.1	−3.4
夏洛特黃蜂	0.3	1.9	−1.4
洛杉磯湖人	−1.5	0.9	−4
紐約人	−3.6	1.1	−8.2
布魯克林籃網	−3.7	−3.1	−4.4
芝加哥公牛	−7	−4.6	−9.5
薩克拉門托帝王	−7	−5.3	−8.7
奧蘭多魔術	−4.8	−3.6	−4
阿特蘭大老鷹	−5.5	−4	−6.9
達拉斯獨行俠	−3	−0.1	−6
孟菲斯灰熊	−6.2	−2.7	−9.7
鳳凰城太陽	−9.4	−10.3	−8.4

附錄四（頁十一圖二）

賽季	主場勝率(%)	客場勝率(%)
2018-2019	80.9	80.3
2017-2018	83.1	80.2
2016-2017	80.2	77.2
2015-2016	78.3	74.3
2014-2015	77.2	76.5

優異作品:

網絡資料審查員

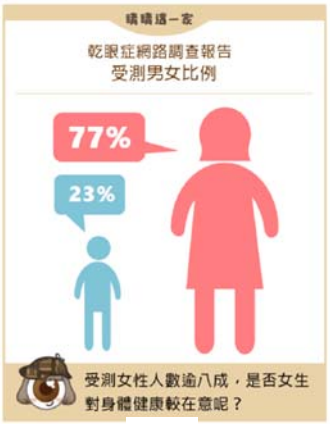
學校名稱：新亞中學
學生姓名：林琳、李銀鈴
指導教師：林新儀老師

引言

在現今網絡發達的時代，資訊隨處可尋。我們隨時可以接收不同的資訊和消息，但是不是所有的資訊都是正確的呢？這次，我們化身為審查員，在網絡上找到一份關於乾眼症的調查報告，並找出不同的問題。

這份報告的整體來來說，特別鮮艷，利用不同的顏色和圖案吸引讀者注意，也有不同圖表顯示數據，嘗試引導讀者明白其含意。雖然這份調查報告整體看起來很吸引，但經仔細閱讀及思考後，卻讓我們發現當中的內容缺少客觀的分析。

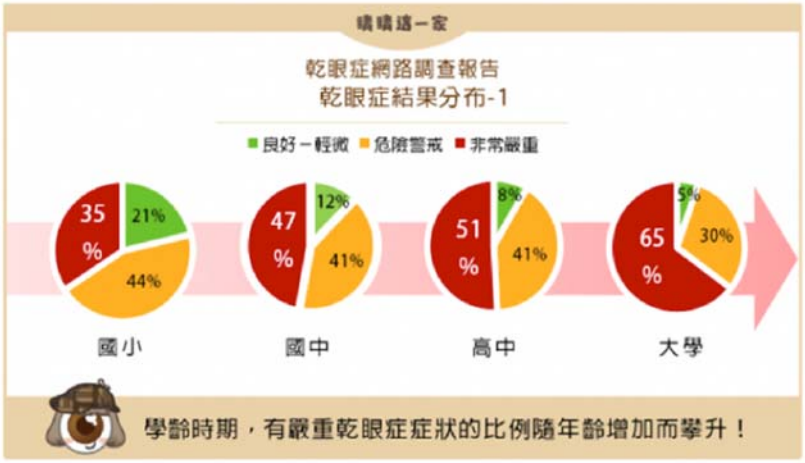
右邊的統計圖(圖一)中，男女的圖像的確美觀，但受測男女的比例與圖中男女圖形的大小比例大相逕庭，編者誇大了女生的比例，而且圖中文字寫道：「受測女性人數逾八成，是否女生對身體健康較在意呢？」但調查報告屬於隨機調查，受測的女生人數多與少並不能得出女生對身體健康是否較在意這個問題的答案，可見這個問題存在了編者的主觀意見，誤導讀者的思考。編者說，受測女性逾八成，但調查顯示受測女生的比例為77%，所以編者未能陳述事實。



圖一

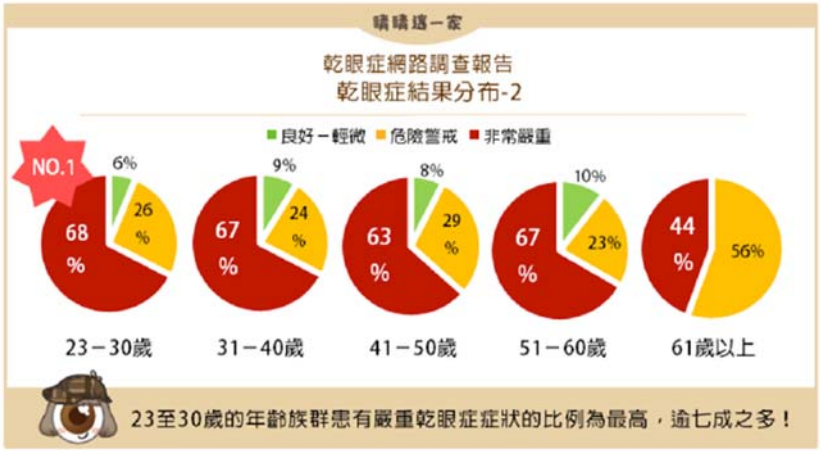
如左邊的受測年齡分佈(圖二)的棒形圖所示，受測的組別不正確，當中「大學」和「23-30」這兩個的組別極有可能包含重複的數據。到底一個二十歲的大學生，應同時在兩個組別中，還是隨意選一個呢？若把教育程度及年齡分為兩類，再為每類製作統計圖，相信能更有效突顯結果及更易加以分析。

圖二



圖三

在上述的乾眼症結果分佈-1(圖三)的各個圓形圖中，分為「良好-輕微」、「危險警戒」、「非常嚴重」四個組別，而這些組別在整份的報告中並沒有具體說明怎樣分類的。到底怎樣算是非常嚴重呢？我們不得而知。



圖四

而在上述的乾眼症結果分佈-2(圖四)中，44%及56%的「61歲以上」人士乾眼症狀分別屬非常嚴重及危險警戒，但圖二顯示61歲以上的只有9個人，以9個人代表所有這個年齡層的人的情況，結果並不公平。

總結

經過我們網絡資料審查員仔細地審核這份美觀的報告後，發現整份報告都帶有編者的主觀意見。而且，報告提及，受測總人數有2274人，而有效數據只有1554人，並未提及無效的720個數據如何無效，因而讓我們質疑整個報告的公平性。網絡世界存在的資料多不勝數，我地應小心、公平、理性地分析資料的可信性，明辨是非。

（773 字）

參考資料

[1]《你有乾眼症嗎？＜網路調查報告＞》
https://www.bigeyefamily.com/archives/955

優異作品:

離婚率

學校名稱：順德聯誼總會李兆基中學
學生姓名：彭少柏
指導教師：許俊江老師

摘要：

在看本文之前，你認爲香港人的婚姻狀況如何？如果我告訴你，香港有三分之一的人離婚，你會相信嗎？這就是誤用數據的一個例子。本文將會探討這篇參考文章中錯誤計算離婚率的情況，和政府統計處如何計算離婚率。

根據蘋果日報的一則報導，每三對夫婦就有一對離婚收場，2016年的離婚率更是高達34.38%。究竟這個百分比是如何計算的？數據的可信度有多高？本文會為你一一解答。

根據政府統計處網站，2016年的結婚數目為50,008，離婚數目則為17,196。將離婚人數除以結婚人數，轉換成百分比後約為34.4%，即報章中得出的結論。

$$\text{報章的計算方法：2016年的離婚率} = \frac{\text{2016年的離婚人數}}{\text{2016年的結婚人數}} \text{x100\%}$$

但我認為，兩者並不能直接比較，因為2016年的離婚人士不見得全部都是2016年的結婚人士，比較這兩組的數據所得出的結果沒有意義。換句話說，若果按報道所說「2016年離婚率達34.4%」，是否代表於2016年結婚的五萬多名結婚人士當中，有34.4%的人將會或已經離婚？恐怕這並不是事實，2016年的結婚人士將來會否離婚仍是未知之數，而於2016年離婚的人士是於2016年或之前的不同時間結婚，將2016 年的離婚數目除以2016年的結婚數目所得出的並不是離婚率。打個比方，若將2016年的死亡人數46900除以2016年的出生人數60900，轉換成百分比約為77%，是否能得出2016年的死亡率？這只會貽笑大方。

那麼我們應如何計算離婚率？可參考政府統計處有關粗離婚率和粗結婚率的計算。粗離婚率是指某一年內，獲頒布離婚令數目相對該年年中每千名人口的比率。

$$\text{政府統計處的計算方法：2016年的粗離婚率} = \frac{\text{2016年的離婚人數}}{\text{2016年的中期人口}}$$

相似地，粗結婚率是指在某一年內，結婚數字相對該年年中每千名人口的比率。整理政府統計處的數據後所得出的圖表如下：

從上述圖表可見，香港整體的粗結婚率由2001年的4.8逐漸上升至2012年的8.4，其後回落至2016年的6.8。至於粗離婚率，則由2001年的2.00逐漸上升至2013年的3.10，在2016年則回落至2.34。在2016年，整體粗結婚率和粗離婚率的差距為4.46（每千名人口）。對比近十年的差距，其實沒有太大差別。

總結

有些數據並不能直接比較，就像本文中的離婚率。按作者計算離婚率的方法，其實可以追蹤同一批人，找出這批人在幾年，甚至幾十年後的婚姻狀況，直到全部研究對象去世爲止，但這種方法耗時太長。所以，用粗離婚率的計算方法會較爲簡單。讀者要自行判斷內容的真確性，計算得出的數據要符合常理，不應盲目地相信報道內容。

（字數：797）

原文：

[1] 港人婚姻缺激情 性福不足現危機

蘋果日報2018年8月5日

網址連結：

https://hk.news.appledaily.com/local/daily/article/20180805/20469073

據政府統計數字，本港每3對夫婦就有1對離婚收場，日離婚率過去有上升趨勢，2016年離婚率達34.38%。有機構昨公佈「婚姻體檢」問卷調查，結果顯示本港「婚姻健康指數」平均數達73分，但在性行為及抽時間討論彼此關係的分數，則低於婚姻危機警戒線的66分。專家表示， 普遍港人因生活節奏急速而忽略維持婚姻關係，建議夫妻每天至少有15分鐘進行有質素的傾談、一起計劃活動等。

城市青年商會和香港婚姻及家庭治療協會於5月初網上發佈「婚姻體檢」問卷，截至7 月底有1,295名港人完成問卷。結果顯示，本港「婚姻健康指數」平均數達73分，而婚姻危機警戒線為66分，即低於66分的婚姻關係可能已亮紅燈。問卷結果顯示，平均分最低分項目分別是激情（62分）、他/她和我有性行為（60分）、和他/她討論我們這段關係的質素（58.5分）。

女性為家庭付出較男性高

調查又發現， 港男比港女婚姻幸福感較高，特別是在關係質素、情感交流和維持關係的行為上。商會指出，根據前線服務經驗及國外文獻，女性對婚姻關係的付出通常較男性高，對於家庭成員的需要也比較敏感。加上社會對性別角色的

期望，女性也比較會因家庭需要而犧牲個人發展與喜好，因此男性於婚姻關係中比女性的滿足感較高。

香港婚姻及家庭治療協會主席魏素華表示，普遍受訪港人忽略維持婚姻關係的行為， 與港人生活模式和節奏有關，建議夫妻每天至少有 15 分鐘進行有素質傾談、計劃一起參與活動、學習排解衝突方法等。她又指，在婚姻關係中，丈夫很多時因較少表達情感，導致雙方不了解而產生爭吵，最終離婚收場。

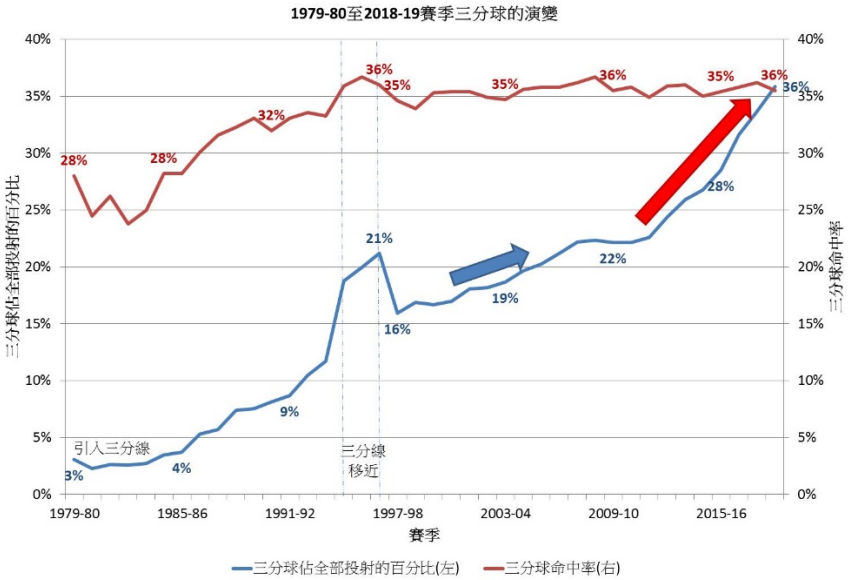
邀請作品：

淺談 NBA 統計

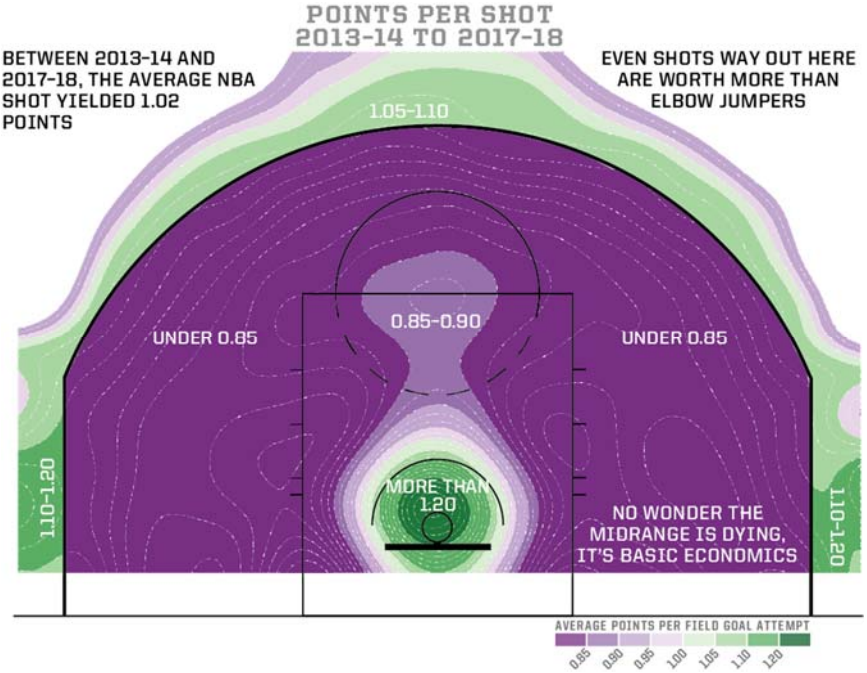
近年，不少中學生統計創意寫作比賽中的作品，是以國家籃球總會（National Basketball Association，簡稱 NBA）為題材。適逢今年比賽專題為「運動中的統計」，筆者亦希望在此作一些分享。

過往的中學生統計創意寫作比賽中的作品，都會利用 NBA 的統計數字找出一些關聯或有趣的現象。事實上，近年 NBA 亦積極運用統計分析，因而令整個聯盟出現極大的變化。最具代表性的例子，就是現今 NBA 球員的三分球投射次數愈來愈多。

圖中可見，NBA 於八十年代初引入三分線。初期三分投射並不算普遍，約二十次投射才有一球是三分球。及後三分投射慢慢增多（除 1994-95 至 1996-97 賽季因三分線移近造成短暫大幅提升）。至近四、五個賽季，三分投射更大幅提升，2018-19 球季每支球隊平均約每三次投射便有一球是三分球。



根據過往統計，在 NBA 比賽中投射一球的平均得分大約是一分。假如聯盟的三分球命中率超過 33.3%，其得分期望值（Expected value）便會高於平均值的一分。在 NBA 近二十個賽季，三分球命中率都高於此水平，因此我們不難理解三分球投射為何愈來愈多。到近年空間分析（Spatial analysis）大行其道，不少 NBA 球隊開始利用統計分析協助製訂比賽戰術及挑選球員。下圖就是利用空間分析研究球員於賽場上各個位置投射一球的得分期望值，分析利用 2013-14 至 2017-18 賽季共五年的數據編制而成。圖中可見，得分期望值高於平均值一分的位置（綠色）分佈在籃底及三分線上，而低於平均值一分的位置（紫色）則集中在籃底及三分線之間的位置（一般稱為中距離）。



於是，球隊近年所製訂的比賽戰術，都是針對性地希望球員多在三分線及籃底投射，減少中距離投射。比賽戰術的改變因而令近四、五個賽季的三分投射大幅提升。

談到以統計分析營運球隊，就不得不提起 2011 年美國有一套電影叫“Moneyball”，是以真人真事改編的電影，內容講述美國職業棒球大聯盟（Major League Baseball）球隊奧克蘭運動家如何運用數據分析在 2002 年取得驚人的成績。而在籃球界，美國有統計學者早於九十年代初開始研究進階數據（Advanced metrics）。所謂進階數據，簡單來說就是建基於傳統的比賽數據去研發更反映球員表現及效率的一些指標。要了解進階數據，首先要認識以下這些傳統數據：

傳統數據		備註
上陣時間	Minutes (**MIN**)	
投射次數	Field Goals Attempted (**FGA**)	只包括兩分及三分投射，不包括罰球投射
投射命中	Field goals made (**FGM**)	
投射命中率	Field Goal Percentage (**FG%**)	**FG%** = **FGM** / **FGA**
三分投射次數	Three Pointers Attempted (**3PA**)	
三分投射命中	Three Pointers Made (**3PM**)	
三分投射命中率	Three Point Percentage (**3P%**)	**3P%** = **3PM** / **3PA**
罰球投射次數	Free Throws Attempted (**FTA**)	
罰球命中	Free Throws Made (**FTM**)	
罰球命中率	Free Throw Percentage (**FT%**)	**FT%** = **FTM** / **FTA**
得分	Points (**PTS**)	**PTS** = 2 x **FGM** + **3PM** + **FTM**
進攻籃板	Offensive Rebounds (**OREB**)	
防守籃板	Defensive Rebounds (**DREB**)	

傳統數據		備註
籃板	Total Rebounds (**REB**)	**REB** = **OREB** + **DREB**
助攻	Assist (**AST**)	
偷球	Steals (**STL**)	
封阻	Blocked Shots (**BLK**)	
失誤	Turnover (**TOV**)	
犯規	Personal Foul (**PF**)	

基本上，傳統數據已反映了球員在比賽中的絕大部分表現。然而，當中仍有一些不足的地方。較易明的一個例子，就是數據不能直接反映球員的進攻效能。如果單純比較得分，有些球員可能因為有較多的投射次數，所以即使投射命中率低，亦可以得分高。如果比較投射命中率，則因為兩分球和三分球的價值不同，以及投射命中率不包括罰球投射，所以不太公平。

因此，統計學者修改了計算命中率的公式，其中一條公式是有效命中率（Effective field goal percentage, eFG%）。概念很簡單，就是將投射命中率中的三分投射命中乘上 1.5（因為入一球三分球比入一球兩分球多 1.5 倍分數）：

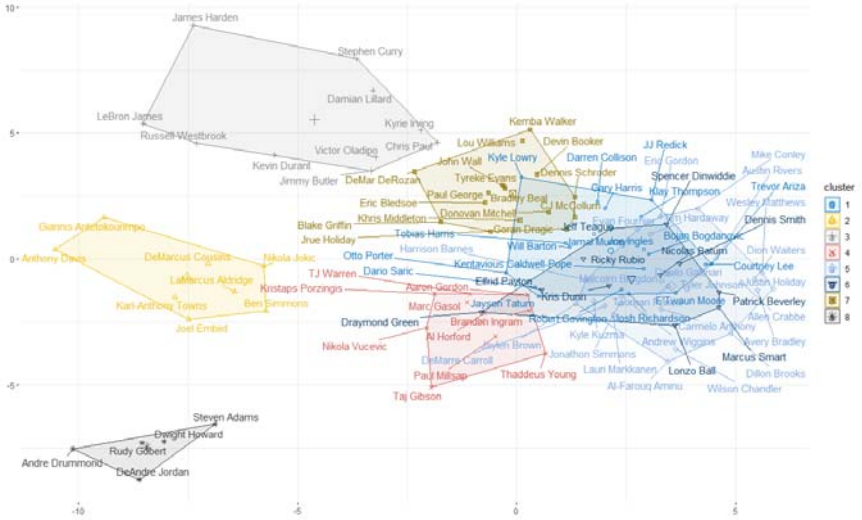
$$\text{eFG\%} = \frac{FGM + 0.5 \times 3PM}{FGA}$$

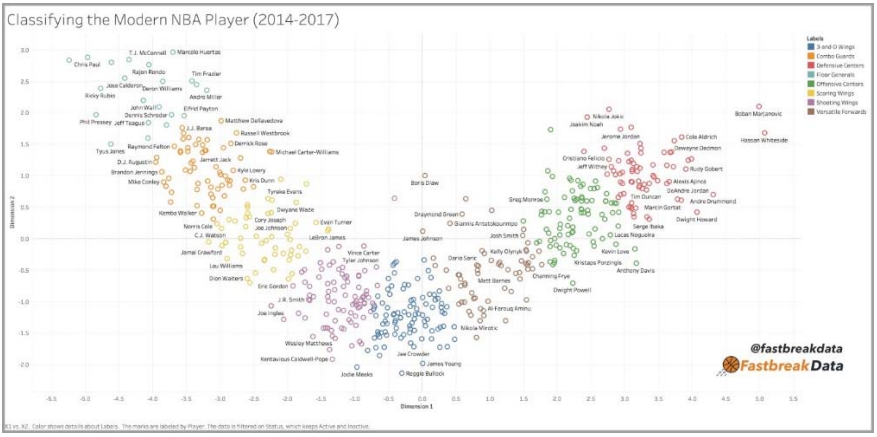
不過，有效命中率始終沒有考慮到一些進攻時搏取犯規獲得罰球的得分。因此統計學者再修改了公式，其公式為：

$$\text{TS\%} = \frac{PTS}{2 \times FGA + 0.88 \times FGA}$$

這個公式叫真實命中率（True shooting percentage, TS%）。它將罰球投射次數乘 0.88，意思是只有 88%的罰球是因進攻時搏取犯規而獲得，而 88%這數字是根據過往統計所得出的。

以上是眾多進階數據的其中一個例子，其他的進階數據例子其實在一些體育網站 ESPN 或 Basketball Reference 都可以找到。利用這一類進階數據，加上統計分析，可以協助球隊了解球員的屬性及特質。筆者隨意網上搜查，發現有不少人都進行過類似的統計分析。下面兩圖就是網上有人利用 K-means 分群（K-means clustering）及主成分分析（Principal component analysis）作分析的例子。兩幅圖都利用 K-means 分群將 NBA 球員分為八組，再用主成分分析則將眾多的進階數據簡化為兩大主要成分，並以散佈圖將結果視覺化呈現：





了解球員的屬性及特質可幫助教練團選出合適的比賽陣容，甚至方便球隊在球員交易市場羅致合適的球員。或許有一天，“Moneyball”的故事會在 NBA 發生。

參考資料:

[1] The NBA is obsessed with 3s, so let's finally fix the thing https://www.espn.com/nba/story/_/id/26633540/the-nba-obsessed-3s-let-fix-thing

[2] How Mapping Shots In The NBA Changed It Forever https://fivethirtyeight.com/features/how-mapping-shots-in-the-nba-changed-it-forever/

[3] Kirk Goldsberry (2012). CourtVision: New Visual and Spatial Analytics for the NBA https://pdfs.semanticscholar.org/46e4/a7271de62e9118625dec935c4aef1bc0ea74.pdf

[4] Assessing NBA player similarity with Machine Learning (R) https://towardsdatascience.com/which-nba-players-are-most-similar-machine-learning-provides-the-answers-r-project-b903f9b2fe1f

邀請作品：

大數據的應用與挑戰

引言

相信讀者對大數據（Big Data）一詞不會感到陌生。隨著科技不斷進步，特別要互聯網的普及，形形式式的大數據正急速發展。不論是商界、政府、學術界，甚至是普羅大眾均對大數據趨之若鶩。究竟大數據有甚麼魔力，讓世界各地的人士都紛紛去湊熱鬧，希望取得各式各樣的大數據進行分析？本文嘗試從不同角度簡單介紹大數據的應用及其面對的挑戰。

何謂大數據？

顧名思義，大數據最大的特性就是「大」，不但是數據量龐大，其增長速度及多樣化亦遠高於過往一般人所認知的數據規模。大數據普遍以 3Vs 來界定，分別是 volume（數量）、variety（種類）及 velocity（速度）。

一． 龐大的數據量 （Volume）

大數據涉及的數據量龐大，當一般市民仍以 TB 為單位去儲存數據時，現時世界各地每年所生產的大數據動輒以 ZB 為單位，根據 IDC 發表的報告，數據量未來將會高速增長，預計於 2018 年至 2025 年這 7 年間，全球所產生的數據量將由 33ZB 增加 4 倍至 175ZB。究竟 ZB 有多大?大家嘗試猜一猜，（一）1 萬個 GB；（二）1 千萬個 GB；（三）1 萬億個 GB。

答案是三，由此可以想像數據量是那麼的龐大。

二． 數據產生及處理迅速 （Velocity）

在現今科技發達的年代，數據產生的速度驚人，每分每秒數據都在不斷增長，其實大家每天亦為急速增長的數據量出一分力，簡單如利用智能電話瀏覽網頁，使用電子錢包結賬，在社交網絡談天或上載相片等，都產生大量的數據。除了增長速度快外，數據的接收、處理等的速度亦較以往更快。

三． 數據多樣化 （Variety）

多樣化不但指資料來源豐富，而且數據並非只是局限於數

字，任何形式及格式的資訊，諸如文字、圖像，甚至音訊及影像數據亦可用作分析。隨著人工智能技術越來越成熟，大數據應用層面亦日趨廣泛。

大數據的應用

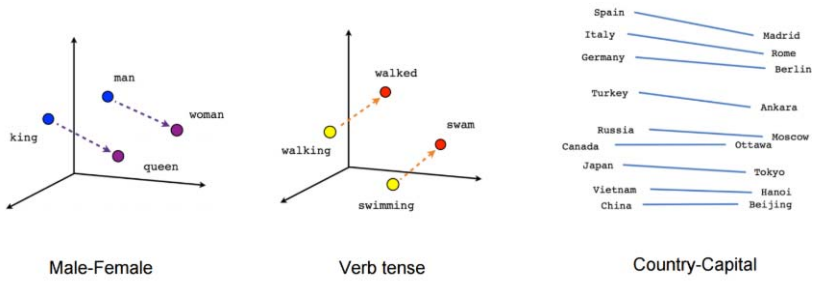
在電腦運行速度較為遜色的年代，使用數 TB 的數據進行分析已可謂天方夜譚。隨著科技的不斷進步，大數據分析不再只是局限於一些較大規模的機構的工作，普羅大眾亦可使用家用電腦，從網上下載各式各樣的大數據自行設計分析模型。現時，政府、學術界及商界均應用大數據作為輔助工具，協助機構作出大大小小的決策。

大家若是臉書的用戶，會發現臉書對你的喜好、朋友圈子、行蹤等均瞭如指掌。事實上，臉書正利用每天由你及身邊朋友收集的大量數據進行分析，如透過你曾瀏覽的網頁去了解你的喜好，利用人面辨識技術從照片找出你和你的朋友。商界往往亦利用從客人收集的數據，去制定更適切的推銷策略。政府及公營機構亦使用大數據提升工作效能及改善服務。例如香港天文台便利用大數據來分析天氣對社會帶來的影響，如 2018 年 9 月超強颱風山竹襲港時，天文台便利用塌樹或水浸等報告，並與天氣圖上的天氣資料聯繫起來作更全面的分析；醫院管理局早於十年前便利用臨床大數據，就內科長者病人再次入院機會率進行預測，從而提供更適切的服務以降低其再次入院的機會；而政府亦將利用人工智能技術，於 1823 聯繫中心和「香港政府一站通」網站正式推出聊天機械人服務以處理市民的查詢。只要留心，便會發現日常生活中有很多產品及服務和大數據息息相關。

機器學習的技術

接下來將跟大家簡單介紹現時一些流行的大數據分析技術，亦即機器學習（Machine learning）技術。如上文提及，除了數字外，文字、圖像、聲音等亦可透過機器學習，從大量的資料中發掘隱藏的資訊。

以文字為例，早期使用的文字分析方法，大多數將句子或文章中每一個字以一個不重複的數字來表示，或以 1 和 0 標示，1 代表該字在句子或文章中存在，0 代表不存在，這些方法並未包含文字與文字之間相互關聯，簡單來說，中國和北京字義上很相近，因為兩個詞語分別代表國家及其首都，但以數字代表時卻看不出其關聯性。而字詞嵌入（word embedding）的出現則能打破上述方法的限制。以一個名為 Word2Vec 的模型為例，模型透過學習大量文字數據，將每一個詞語以一個字詞的向量（字詞的向量為一組數字）代表，向量涵蓋該詞的字義，透過比較各字詞的向量，便能利用模型去猜測和目標字/句最有關聯的字/句。例如輸入中國一詞，模型能猜測到和中國相關的詞語為北京；若以圖表示，會發現相關聯的字詞位置會較接近（如下圖）。



圖片來源：https://www.tensorflow.org/images/linear-relationships.png

現時較廣泛應用的技術除了 Word2Vec 外，還有 GloVe、fastText 等。而圖像、聲音等亦可轉化為有意義的數字，利用各種機器學習的模型進行分析。由於機器學習的突破，有關技術已被廣泛應用於文章翻譯、將圖像、聲音轉化為文字，辨識面容、簽名等。

大數據時代所面對的挑戰

縱然利用大數據進行分析有很多過人之處，但亦要留意其不足的地方。首先，龐大的數據中往往夾雜很多垃圾資訊，因而影響分析的準確度；此外，數據的來源可能只涵蓋分析對象的一部分，資料未必有代表性，而依賴這些欠缺代表性的數據所得出的結論亦會有偏差。

而另一值得關注的是，不同機構往往在用戶知情甚至不知情情況下收集了大量數據。因此，將個人資料提供予其他機構前，要確保你已清楚對方收集資料的用途，以及有關機構會否向其他人士或機構披露資料，以保障個人私穩。

結論

隨著大數據應用的普及，有關使用及分析大數據的資源亦日趨豐富，只要小心處理數據，大數據無疑對人們的生活帶來更多、更新、更有用的資訊。大家亦可嘗試自行設計分析模型，去探索更多大數據的可能性。

參考資料:

[1] IBM, “What is big data? More than volume, velocity and variety…”
https://developer.ibm.com/dwblog/2017/what-is-big-data-insight/

[2] IBM, “The Four V's of Big Data”
https://www.ibmbigdatahub.com/infographic/four-vs-big-data

[3] International Data Corporation (IDC), “The Digitization of the World. From Edge to Core”
https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf

[4] Tensor Flow, Word embeddings
https://www.tensorflow.org/tutorials/representation/word2vec?hl=zh-cn

[5] 香港天文台。從大數據探索天氣對社會的影響
http://my.weather.gov.hk/m/article_uc.htm?title=ele_00515

[6] 醫院管理局。善用大數據　規劃未來服務
http://www3.ha.org.hk/ehaslink/issue97/tc/news-04-tc.html

[7] 立法會資訊科技及廣播事務委員會。電子政府服務 2019 年 6 月 10 日討論文件
https://www.ogcio.gov.hk/tc/news/legco_papers/2019/06/doc/lb_20190610.pdf

[8] 香港電台。「大數據」分析局限 乃傳統統計學問題
https://app3.rthk.hk/mediadigest/media/pdf/pdf_1490268097.pdf

邀請作品：

《標準差—何去何從？》

探討加入(或移除)的數據值不等於平均值時，標準差的變化。

朱吉樑老師

宣道會鄭榮之中學

於中學課程中，數據變化對離散度的影響是一個頗熱門的問題。一般討論方向，是加入(或移除)的數據值(或一組數據的平均值)與原本的數據平均值相同，由於加入數據後的平均值沒有變化，通過代入標準差的公式，便能得出標準差的值必定下跌的肯定結果(見附頁)。學生偶而會問：「若加入或移除的數據的平均值與原本的數據平均值不相同，標準差又會如何？」

為了方便討論，本文只探討加入數據的情況，(實際上，移除數據的處理手法也是差不多，可留給學生嘗試。) 讓我們先與學生考慮一些特殊情況作為討論的開場。

給定兩個數據 x_1 和 x_2，假設$x_1 = 1$、$x_2 = 3$，我們可得出表一的統計量：

表一：數據的平均值及標準差。

數據	平均值($\bar{x}$)	標準差(σ)
$x_1 = 1,\ x_2 = 3$	2	1

若加入的數據值(x_a)不斷增加(向右遠離平均值 2 時)[3]，平均值($\bar{x}$)和標準差(σ)會有甚麼影響？作為初探，選了$x_a = 2$(平均值)，2.5(原數據的範圍內)，3(原數據的範圍的最大值)及 3.5(原數據的範圍以外)測試一下對標準差的影響，表二記錄了當中的結果。

表二：加入數據x_a時，平均值及標準差的改變。

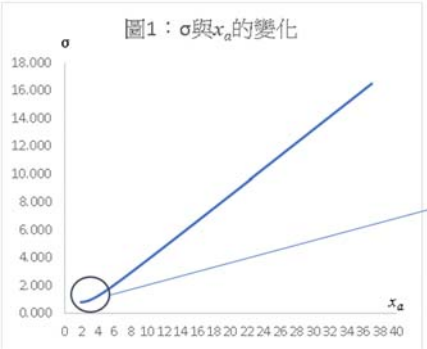
x_a	$\bar{x}$	σ
2	2	0.816
2.5	2.17	0.850
3	2.33	0.943
3.5	2.5	1.08

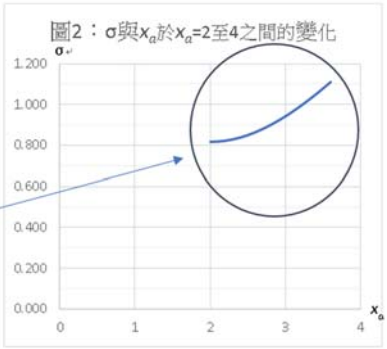
從上表可見，平均值和標準差均隨x_a的增加而上升。此外，原本數據的平均值是 2，當加入的數據值高於平均值，不難想像，新的平均值會隨x_a的增加而不斷上升；可是標準差的情況卻有點不同，原本數據的標準差是 1，但當x_a的值在原數據的範圍內時(小於或等於 3)，新的標準差的值看來會小於原本數據的標準差(當$x_a = 3$ 時，標準差也只是 $0.943 < 1$)；不難想像，當x_a的值於範圍外(大於 3)時，新的標準差的值不一定大於原本數據的標準差！

我們不能無止境的試不同的x_a值，為了了解更多，在此，運用了試算表(EXCEL)的功能，看看x_a與標準差的變化(見圖 1)，表面上看，好像是一條直線！保守一點，至少可以看到是一條遞增的線。但當仔細觀察時，聚焦於x_a的值於 2 與 4 之間，圖像清楚顯示應為曲線(見圖

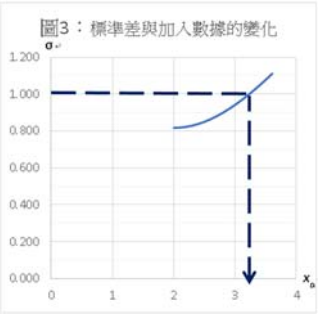
[3] 由於加入數據的值減少和增加只是反射對稱的關係，對標準差的影響相同，故本文不贅論述，只考慮增加的轉變。

2)。





另一個問題，x_a應為何值才會令標準差保持不變(即$\sigma = 1$)？從圖 3 中可見，當x_a稍稍大於最大值（$x_2 = 3$）時，標準差便開始大於本身的值，但究竟要大多少才會發生呢？而當數據大於兩個時，此現象會否不同？



對於加入數據值與標準差，現在有了基本的看法，是時候從數式找尋更深入的答案。

為了方便討論，因應標準差不受整體數據的位移影響，即當每個數據x_i加上常數 c 時，另把轉換後的數據稱為 Y (即 $Y = X + c$)時，Y的標準差和X的標準差無異(即$\sigma_Y = \sigma_X$)，圖 4 把當中的關係以圖像形式展示。

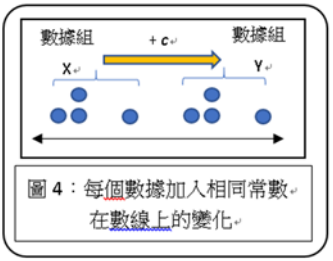


圖 4：每個數據加入相同常數在數線上的變化

現假定有 n 個數據，而當中的平均值為零，即$\bar{x}=0$。若實則數據的平均數不是零，只須通過線性轉換$Y=X-\bar{X}$，便可把數據組的平均值轉為零，而不影響本身數據組的標準差。

有了以上的簡化，再把標準差的公式[4]以另一形式展示

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

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

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

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

$$=\sqrt{\frac{x_1^2+x_2^2+\cdots+x_n^2}{n}-\bar{x}^2}$$

由於這個數據的平均值為 0，標準差為 σ，即上式可再簡化為：

$$\bar{x}=0 \text{ 及 } \sigma=\sqrt{\frac{x_1^2+x_2^2+x_3^2+\cdots+x_n^2}{n}}$$

現加入數據x_a

$$\bar{x}_{New} == \frac{n(0)+x_a}{n+1} = \frac{x_a}{n+1}$$

$$\sigma_{New}=\sqrt{\frac{x_1^2+x_2^2+\cdots+x_n^2+x_a^2}{n+1}-\bar{x}_{New}{}^2}$$

[4] 為了方便沒有修讀 M2 的同學，此處並未以連加符號"$\sum x_i$"簡化。

$$= \sqrt{\frac{n\sigma^2+x_a^2}{n+1} - \left(\frac{x_a}{n+1}\right)^2}$$

$$= \sqrt{\frac{n\sigma^2}{n+1} + \frac{x_a^2}{n+1} - \frac{x_a^2}{(n+1)^2}}$$

$$= \sqrt{\frac{n\sigma^2}{n+1} + \frac{(n+1)x_a^2 - x_a^2}{(n+1)^2}}$$

$$= \sqrt{\frac{n\sigma^2}{n+1} + \frac{nx_a^2}{(n+1)^2}}$$

$$= \sigma\sqrt{\left(\frac{n}{n+1} + \frac{nx_a^2}{(n+1)^2\sigma^2}\right)}$$

若考慮標準差不變的情況時，即

$$\frac{n}{n+1} + \frac{nx_a^2}{(n+1)^2\sigma^2} = 1$$

$$\frac{nx_a^2}{(n+1)^2\sigma^2} = 1 - \frac{n}{n+1}$$

$$\frac{nx_a^2}{(n+1)^2\sigma^2} = \frac{1}{n+1}$$

$$x_a^2 = \frac{1}{n+1} \times \frac{(n+1)^2}{n}\sigma^2$$

$$x_a^2 = \frac{n+1}{n}\sigma^2$$

$$x_a = \sqrt{\frac{n+1}{n}}\sigma$$

從以上只得兩個數據的情況為例，只有兩個數據 $x_1 = 1,\ x_2 = 3$，即 $n = 2$，$\sigma = 1$，

即當$x_a = \sqrt{\frac{2+1}{2}(1)} = \sqrt{\frac{3}{2}} = 1.2247\ldots\ldots$，標準差將維持不變。

由於例一的數據平均值是 2，只要把數值右移兩個單位，即當加入的數據為 3.2247…(與圖 3 的情況吻合)，標準差便會不變。(同學可用計算機作簡單驗證)

同學亦不難發現，當$x_a > \sqrt{\frac{n+1}{n}}\sigma$ 時，新的標準差便會大於原本的標準差，反之亦然。

再考慮新的標準差的公式：

$$\sigma_{New} = \sigma\sqrt{\left(\frac{n}{n+1} + \frac{nx_a^2}{(n+1)^2\sigma^2}\right)}$$

$$= \sqrt{\left(\frac{n}{n+1}\sigma^2 + \frac{n}{(n+1)^2}x_a^2\right)}$$

當x_a的值越大，$\frac{n}{n+1}\sigma^2$和$\frac{n}{(n+1)^2}$相對越小，以致標準差與x_a 的關係出現了圖 1 右上方，好像直線的關係。另一方面，若數據量(n)很大時，$\frac{n}{n+1} \to 1$ ，$\frac{n}{(n+1)^2} \to 0$ 則 $\sigma_{New} \to \sigma$。在此情況下，x_a的改變對於整體數據的影響將會變得微不足道。最後，若$x_a = \bar{x}$，由於$\bar{x} = 0$，數式中新的標準差又變回附頁一中的$\sigma_{New} = \sqrt{\frac{n}{n+1}}\sigma$了。若考慮移除數據又會如何？不如作為同學的功課，試試證明一下。

參考資料:

[1] Miller, I., Miller, M., Freund, J. E., & Miller, I. (2004). *John E. Freund's mathematical statistics with applications*. Upper Saddle River, NJ: Prentice Hall.

附頁一：

加入數據相等於平均數($x_a = \bar{x}$)時，標準差(σ)的變化。

考慮 n 個數據，它的平均值($\bar{x}$)及標準差(σ)為：

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

現加入數據x_a

若$x_a = \bar{x}$，新的平均值($\bar{x}_{New}$)及標準差(σ_{New})為

$$\bar{x}_{New} = \frac{x_1+x_2+x_3+\cdots+x_n+x_a}{n+1}$$

$$= \frac{x_1+x_2+x_3+\cdots+x_n+\bar{x}}{n+1}$$

$$= \frac{n\bar{x}+\bar{x}}{n+1} = \bar{x}$$

$$\sigma_{New} = \sqrt{\frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+(x_3-\bar{x})^2+\cdots+(x_n-\bar{x})^2+(x_a-\bar{x})^2}{n+1}}$$

$$= \sqrt{\frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+(x_3-\bar{x})^2+\cdots+(x_n-\bar{x})^2+(\bar{x}-\bar{x})^2}{n+1}}$$

$$= \sqrt{\frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+(x_3-\bar{x})^2+\cdots+(x_n-\bar{x})^2+0}{n+1}}$$

$$= \sqrt{\frac{n}{n+1}}\sigma < \sigma \quad (\text{因 } \frac{n}{n+1} < 1)$$

移除數據的技巧與加入數據的技巧類同，同學可自行證明當移除數據相等於平均數時，$\sigma_{New} = \sqrt{\frac{n}{n-1}}\sigma$.

邀請作品：

Matrix Completion and Its Application to Movie Recommendation

Philip L.H. Yu
Department of Statistics and Actuarial Science
The University of Hong Kong

In the era of big data, the matrix completion problem has become increasingly popular in machine learning and data mining. Matrix completion is the task of filling in the missing entries of a partially observed matrix. A typical example of this is in the Netflix movie rating challenge, launched by Netflix---a move-rental company, which aimed to improve their system for recommending movies to their users. The dataset consists of 100 million ratings of 17,770 movies (columns) on a scale from 1 to 5 given by 480,189 users (rows), resulting in an incomplete rating matrix with nearly 99% of missing entries as not all movies are rated by the same user. The table below shows a subset of the data.

	Movie 1	Movie 2	Movie 3	Movie 4	Movie 5	Movie 6	…
User 1	-	-	-	-	4	-	
User 2	-	-	3	-	-	3	
User 3	-	2	-	4	-	-	
User 4	3	-	-	-	-	-	
User 5	5	5	-	-	4	-	
…							

One of the most popular approaches is based on low-dimensional matrix factorization methods which makes use of the technique of the *singular value decomposition* (SVD). Let us first briefly introduce the conventional SVD.

Given that $\boldsymbol{X}$ is a $m \times n$ matrix with $m \geq n$ and all the entries are observed, its singular value decomposition takes the form:

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}',$$

where $\boldsymbol{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_n]$ is a $m \times n$ matrix such that $\boldsymbol{U}'\boldsymbol{U} = \boldsymbol{I}_n$, an $n \times n$ identity matrix, $\boldsymbol{V} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n]$ is a $n \times n$ matrix such that $\boldsymbol{V}'\boldsymbol{V} = \boldsymbol{I}_n$, and $\boldsymbol{\Lambda}$ is a $n \times n$ diagonal matrix, with diagonal entries $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$ known as the *singular values*.

Note that $\boldsymbol{X}\boldsymbol{X}' = \boldsymbol{U}\boldsymbol{\Lambda}^2\boldsymbol{U}'$ and $\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{V}\boldsymbol{\Lambda}^2\boldsymbol{V}'$. Hence, the $n \times n$ square matrix $\boldsymbol{X}'\boldsymbol{X}$ has eigenvalues $\lambda_1^2, \lambda_2^2, \dots, \lambda_n^2$ with corresponding eigenvectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n$ while the $m \times m$ square matrix $\boldsymbol{X}\boldsymbol{X}'$ has n possibly non-negative eigenvalues $\lambda_1^2, \lambda_2^2, \dots, \lambda_n^2$ with corresponding eigenvectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_n$.

The singular value decomposition provides a low-dimensional approximation of $\boldsymbol{X}$ which aims to find a matrix $\widehat{\boldsymbol{X}} = (\hat{X}_{ij})$ of rank d $(\leq n)$ which minimizes:

$$\sum_{i=1}^{m}\sum_{j=1}^{n}\left(X_{ij} - \hat{X}_{ij}\right)^2.$$

This has a closed form solution:

$$\widehat{\boldsymbol{X}} = \boldsymbol{U}_d \boldsymbol{\Lambda}_d \boldsymbol{V}'_d,$$

where $\boldsymbol{\Lambda}_d$ is a $d \times d$ diagonal matrix with the first d largest singular values as diagonal entries, and $\boldsymbol{U}_d$ and $\boldsymbol{V}_d$ are the first d columns of $\boldsymbol{U}$ and $\boldsymbol{V}$ respectively.

However, in the context of matrix completion, some of the entries of $\boldsymbol{X}$ are missing, and hence the above singular value decomposition cannot be used.

Given a partially observed matrix $\boldsymbol{X} = (X_{ij})$, the matrix completion aims at determining a low-rank matrix of $\boldsymbol{X}$ by finding the matrix $\widehat{\boldsymbol{X}} = (\hat{X}_{ij})$ of rank d which minimizes the sum of squared distance to the target matrix $\boldsymbol{X}$:

$$\sum_{(i,j)\in\Omega} \left(X_{ij} - \hat{X}_{ij}\right)^2,$$

where Ω is the set of the (i,j) pairs for which X_{ij} is observed.

To avoid overfitting and to encourage stability, various methods have been proposed in the literature. A common approach is to adopt matrix factorization with regularization by introducing penalty to the unknown parameters. For example, one can minimize the objective function J:

$$J = \sum_{(i,j)\in\Omega} \left(X_{ij} - \hat{X}_{ij}\right)^2 + \alpha\left(\sum_{i=1}^{m} \boldsymbol{u}_i' \boldsymbol{u}_i + \sum_{j=1}^{n} \boldsymbol{v}_j' \boldsymbol{v}_j\right),$$

where $\hat{X}_{ij} = \boldsymbol{u}_i' \boldsymbol{v}_j$ and $\alpha \geq 0$ is the penalty parameter. A larger

value of α will shrink more entries of $\boldsymbol{u}_i$ and $\boldsymbol{v}_j$ towards zero. Once all the $\boldsymbol{u}_i$'s and $\boldsymbol{v}_j$'s are determined, we can predict those missing entries X_{ij}, $(i,j) \notin \Omega$ by $\hat{X}_{ij} = \boldsymbol{u}_i' \boldsymbol{v}_j$.

Consider a toy example where there are 9 ratings given by 4 users on 4 movies:

	Movie 1	Movie 2	Movie 3	Movie 4
User 1	-	4.5	2.0	-
User 2	4.0	-	3.5	-
User 3	-	5.0	-	2.0
User 4	-	3.5	4.0	1.0

Suppose that the estimated user vectors $\boldsymbol{u}_i$'s and movie vectors $\boldsymbol{v}_j$'s for $d = 2$ are

	Dim 1	Dim 2
User 1	1.2	0.8
User 2	1.4	0.9
User 3	1.5	1.0
User 4	1.2	0.8

	Dim 1	Dim 2
Movie 1	1.5	1.7
Movie 2	1.2	0.6
Movie 3	1.0	1.1
Movie 4	0.8	0.4

The estimated rating for Movie 1 for User 1 is 1.2(1.5) + 0.8(1.7) = 3.16 and the estimated rating for Movie 4 for User 1 is 1.2(0.8) + 0.8(0.4) =1.28. So it is expected that User 1 prefers Movie 1 than Movie 4 and Movie 1 could be recommended to User 1.

In the Netflix movie rating challenge, the objective is to predict the

ratings for unrated movies, so as to better recommend movies to users. The “Cinematch” algorithm used by Netflix had a root-mean-square error (RMSE) of 0.9525 in a testing set. After launching the competition in 2006, the winner algorithm could improve this RMSE by at least 10%. Notice that the SVD played an important role among many competing algorithms in the competition including the winning algorithm. For more details about various matrix completion techniques and the Netflix movie rating challenge, see Chapter 7 of Hastie, et al. (2015).

Reference

[1] Hastie, T., Tibshirani, R. and Wainwright, M. (2015). *Statistical Learning with Sparsity*. CRC Press.

二零一八至一九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
	李思俊先生，政府統計處
	吳詠琴女士，政府統計處
	郭啟然先生，政府統計處
	張柱華先生，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

(二十三) 2016/17 中學生統計創意寫作比賽作品集

(二十四) 2017/18 中學生統計創意寫作比賽作品集

(二十五) 面積和體積