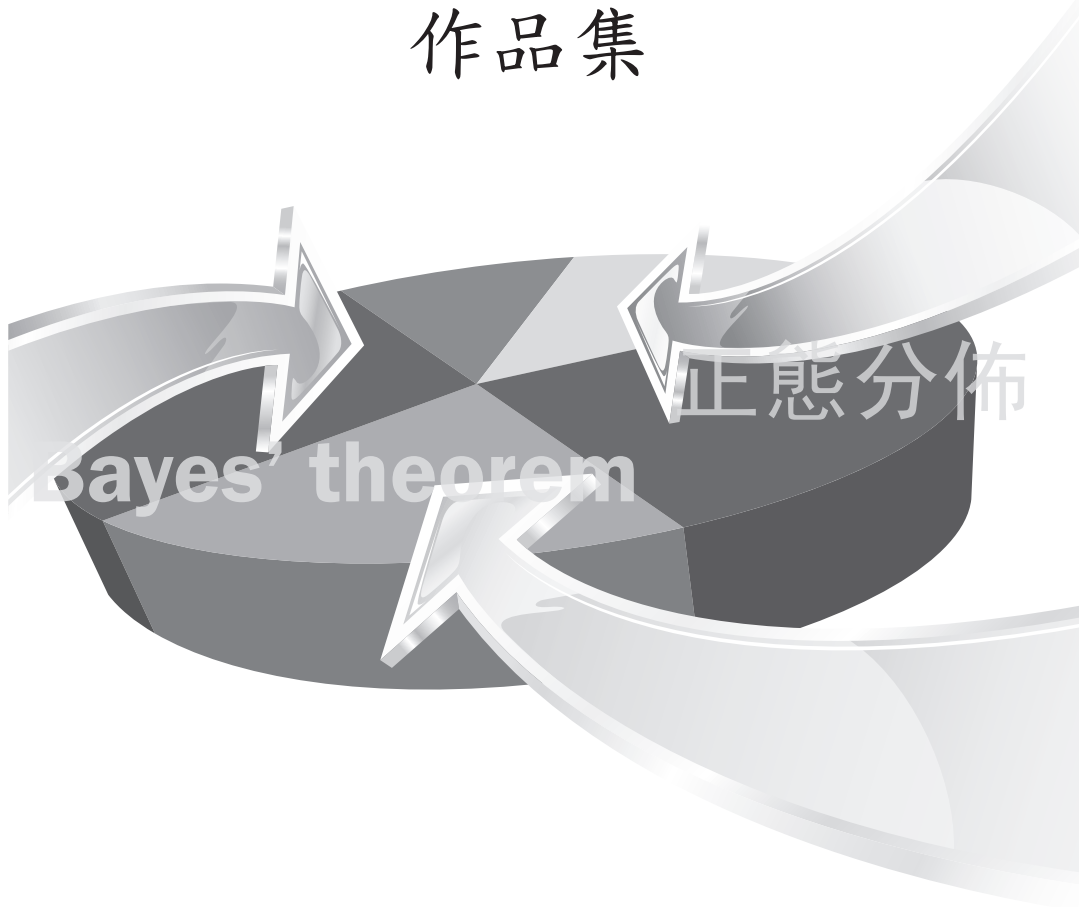


數學百子櫃系列(十)

中學生統計創意寫作比賽 2009

作品集



教育局
課程發展處數學教育組

版權

©2010 本書版權屬香港特別行政區政府教育局所有。本書任何部分之文字及圖片等，如未獲版權持有人之書面同意，不得用任何方式抄襲、節錄或翻印作商業用途，亦不得以任何方式透過互聯網發放。

ISBN 978-988-8040-67-4

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，並蒐集及整理講座資料，輯錄成《數學百子櫃系列》。本書《中學生統計創意寫作比賽 2009 得獎作品集》是這個系列的第十冊。惟本書內大部分文章為上述比賽的優勝作品，均為中學生所編寫。

隨著科學技術的快速發展，量化形式的資訊充滿了生活的每一個角落。本書正是希望透過創意寫作的故事為藍本，傳播概率統計科學知識，讓讀者理解到課堂所學的概率統計知識與我們生活息息相關。

全書共有 14 篇文章，第 1 至 9 篇為「中學生統計創意寫作比賽 2009」的冠、亞、季、殿和優異作品。其他 5 篇則為邀請作品。每篇均可單獨閱讀，望能引發讀者對概率統計的興趣。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)

前言

很多同學會認為，統計這一門學問看似簡單，但在實際應用上會令人有摸不着頭腦的感覺，因為不清楚應該使用哪一個統計方法作分析才是正確，因而減低了學習統計的興趣。有見及此，香港統計學會及教育局在 2009 年合辦了一項名為「中學生統計創意寫作比賽 2009」的比賽，旨在提高中學生對統計的興趣，並鼓勵他們運用創意的手法，以科學和客觀的精神，用文字表達出日常生活中所應用的統計概念或創作一個蘊含統計概念的故事。在此，我衷心感謝香港大學統計及精算學系對此項比賽的贊助和支持，並多謝評審委員會主席張家俊博士及各位評審委員，在近百份參賽作品中公正地選出得獎作品。除了收錄所有得獎的作品外，本書還收錄了由教育局借調老師和政府統計處的統計師創作有關統計概念的文章。我們希望藉着書中每一篇豐富有趣的文章能提高學生對統計的興趣，以及進一步掌握運用統計學的方法。

籌備委員會主席楊良河博士

2010 年 5 月 19 日

序言

香港統計學會向社會各界推廣統計教育的工作一向不遺餘力。我們每年均與教育局合辦「中學生統計習作比賽」(SPC)，目的是要增加同學對社會的認識，鼓勵他們正確運用統計數據及建立團隊精神。

既然已經有一年一度的 SPC，香港統計學會為何又與教育局創辦「中學生統計創意寫作比賽」(SCC)呢？

香港是一個中西文化交匯的地方，語文表達能力對任何一個行業都非常重要。此外，在香港這個彈丸之地，商業機構的營運成本高企，推動高增值服務行業成為香港經濟發展唯一的出路，而創新的意念正是發展高增值服務的必要條件。舉辦 SCC 的目的，正是希望提昇同學的語文表達能力及創意潛能，鼓勵他們以生動有趣的手法表達日常生活中常用但又複雜的統計概念。

書中輯錄了首屆 SCC 的得獎作品，除了對得獎同學付出的努力作出嘉許外，亦希望鼓勵師生在未來積極推廣及踴躍參與 SCC，令 SCC 可像 SPC 一樣茁壯成長。

香港統計學會主席鄧偉江

2010 年 5 月 28 日

序言（二）

是次「中學生統計創意寫作比賽」共收到 101 份參賽作品，經由教育局、政府統計處和香港大學統計及精算系內從事統計工作及教育的專業人士所組成的評審團進行評審。評審標準主要為創意和趣味性，及對統計知識的闡釋和正確應用。在此先鳴謝一眾評審團成員，感激他們公正嚴謹地為我們選出優勝作品。以下是評審團對參賽及得獎作品的意見。

綜觀眾多參賽作品，同學普遍能於取材及表達手法上發揮創意，亦能正確地運用各種統計圖表及統計和概率的知識去分析問題，值得嘉許。特別是勝出的作品，文章題材富創意，緊貼時事，趣味盎然，活用不同的統計方法，如冠軍作品選用了考試出錯題目後的調分方法作分析，題材不但與師生息息相關，當中的分析亦很有系統；或像亞軍的作品，以虛構又有趣的故事帶出問題，引發讀者思考後再道出問題的癥結；另外季軍的作品以電視劇作為藍本，改編成生活化的例子來指出數據運用上的謬誤；其他得獎作品題材亦相當廣泛，如校園驗毒、候車時間、補習社及纖體中心的宣傳技倆等，可見同學們緊貼時事外，還有獨立及批判性的思考。

另一方面，評審團亦指出了一些可改善的地方。首先，同學在闡釋統計概念時，要謹記指出分析背後之假設和合理性，亦應增加

就統計模型的適用性及局限等的討論，以提醒讀者在應用模型時要注意的地方；在文章的組織方面，中心思想可更清晰明確，應避免過多不必要的分析，使文章的結構可以更緊密；在文字表達方面，要減低錯別字或口語，從而提昇文章的可讀性。希望來屆參加的同學能多下工夫，寫出更佳的作品。

本書所收錄的優勝作品經過簡單修潤，每篇均有值得欣賞的地方，亦有瑕疵疏漏，恕未一一點出。務請同學仔細閱讀，老師從旁指導。

評審委員會首席評審員 張家俊博士

2010 年 5 月 28 日

目錄

編者的話	i
前言	iii
序言	v
序言（二）	vii
冠軍作品：不公平的考試	1
亞軍作品：一寸光陰一寸金	9
季軍作品：絕代商驕之誰是主角	15
殿軍作品：「校本驗毒計劃」中的概率應用	21
優異作品：等「你」等到我心痛	29
優異作品：哪一間補習社較好呢？	37

優異作品：你今日驗咗未呀？－校園驗毒計劃成效預測 . . .	43
優異作品：數字不會說謊！?	51
優異作品：哪所補習社比較好?	59
邀請作品：主場之利	67
邀請作品：哪一科老師較易在最受學生歡迎的老師選舉中獲 勝?	75
邀請作品：Game of Chance.	81
邀請作品：抽樣調查的陷阱	93
邀請作品：奧運與概率	97

冠軍作品

不公平的考試

學校名稱：景嶺書院

學生姓名：劉景耀

級別：中五

顧問老師：卓永康

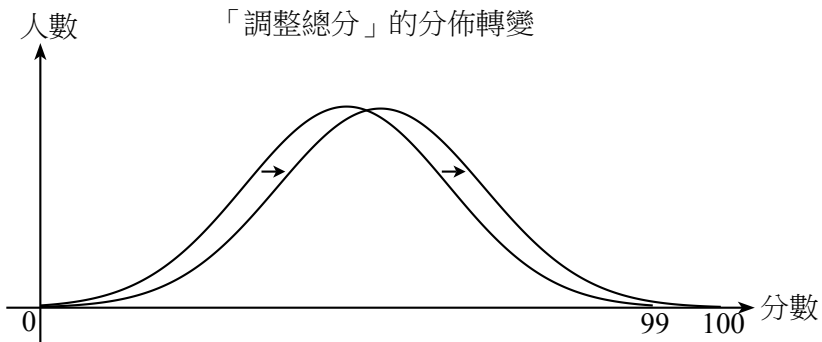
引言

很多時在學校的考試，甚至公開試都有機會出錯題目。老師發現後，普遍會使用兩種常用方法解決，「調整總分」和「全體加分」。這兩種方法看似公平，但實際上又是否完全公平呢？讓我們探討一番。

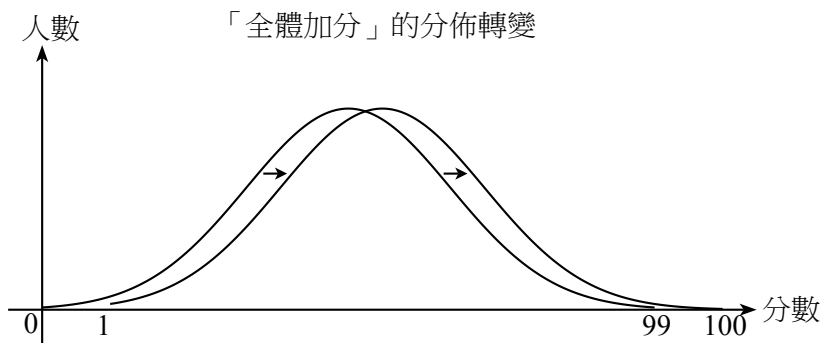
為方便探討，現在先設一例：在某一次數學科公開考試中，有 100 條多項擇題，每題一分。以 a 表示學生的分數。假如其中一條題目出錯，且無法計算答案，在此情況下，學生有機會獲得的分數範圍是 100 分中的 0 至 99 之間的整數。

調分方法

「調整總分」是把原本總分減去 1 分（出錯題目所佔的分數），每一位學生的分數仍然是 a ，但總分則會是 99 而不是 100，以分數表示答對題數與總分比例的話，應是 $\frac{a}{99}$ 。但為著不同行政上的便利，我們往往把學生分數 a 乘上固定的因子（在此情景下，因子是 $\frac{100}{99}$ ）令總分改回 100。而經調整後， a 的可能值則會介乎於 0 和 100 之間。下圖顯示調分方法對分佈的影響：



至於「全體加分」則代表把原本學生分數 a 加上 1 分（出錯題目的分數），在這個情況下，經調整後學生分數範圍則在 1 和 100 之間。參看下圖：



不同調分方法的分析

根據以上圖表所示，「調整總分」的調分方法是把原來的分佈曲線拉闊，而「全體加分」則只把原來的曲線向右平移一分。哪一種方法比較公平？在評論之前，我們有必要了解考試的功能及不同統計工具的意義。考試主要目的有三，包括：

1. 反映學生的實力；
2. 反映學生之間的差異；
3. 「排龍」，即把學生成績排序，以方便相關的行政工作。

由於「調整總分」和「全體加分」都會經過數字上的運算，因此兩種方式都會提升學生的得分及平均值（同學獲得 0 分並用上「調整總分」的調分方法除外，他只會繼續取得 0 分），若調整分數厲害的話，會錯誤反映學生實力。雖然兩種調分方法都會令學生分

數提升，但提升程度卻不一。哪一個方法較佳？我們試考慮一個特殊例子：假如同學在這次考試沒答對一題，取得零分。如果因為有一題題目出錯，在「全體加分」的情況下，令同學無故得分，但同學所獲得這一分並不是反映他的實力，而是分數上的調整，這是否有點荒謬？相反，「調整總分」則對這些學生無影響，調整後分數仍然是零。因此，若要反映學生能力，「調整總分」這個方法看似比較合適。若討論學生差異，我們會使用「標準差」來分析。「調整總分」會使分佈的差距擴大（因原本的分佈曲線被拉闊），但若用上「全體加分」，分佈曲線只是向右平移一分，沒有影響標準差。所以，若要了解學生的差異，「全體加分」好像比較合適。至於從「排龍」角度來看，兩種調分方法均不會影響同學在分數上的名次。我們把上述內容總結下表：

討論事項	使用的統計工具	「調整總分」方法	「全體加分」方法
反映學生實力	學生分數	提升 (按能力提升)	提升 (獲同一的提升)
	平均值	影響相對小	影響較大
反映學生差異	標準差	學生差異擴大	無影響
「排龍」	學生分數的名次	兩者皆無影響	

根據以上的圖表，我們得知，兩種解決「出錯題目」的方法都不能完全公平地處理學生實力和學生差異的情況（「排龍」除外）。調整後的數字只能盡量把不公平的程度減至最低。

其它因素

可是，以上的討論只是從宏觀角度考慮。其實有很多微觀的因素例如：考試時間、出錯題目的題號以及學生能力並沒有在上述討論考慮到，但它們卻會有著舉足輕重的影響。

在上例中，當同學遇上「出錯」題目時，學生一般會花上額外時間思考「出錯」題目。但考慮到學生能力不同，每位學生面對「出錯」題目所花的時間不一，繼而影響其他題目作答的時間也不一。由於之前的宏觀例子並不能深入探討時間這個因素，我們試從以下個別個案中探究。

處境：同樣地，假設數學科公開考試，有 100 條多項選擇題目，每條一分，限時 100 分鐘。我們假設以下是三位考生 X、Y 及 Z 的資料：

	同學 X	同學 Y	同學 Z
數學能力	強	中	弱
每道題目所花的平均時間（分鐘）	0.8	0.95	1.4
答對題目的百分比	90%	70%	50%

如果這次考試沒有任何題目出錯的話，學生所得的分數及完成時間應如下：

	同學 X	同學 Y	同學 Z
總分期望值（分）	90	70	35.5 ¹
所用時間的期望值（分鐘）	80	95	140

如果這次考試其中一道題目出錯，總分便會稍有改變。我們考慮以下兩種情況：

情況 1： 假設「出錯」題目在第一題，三位學生都花時間作答「出錯」題目，影響的是考生所作答的時間，尤其是同學 Z，他會因為能力稍遜而花上更多的思考時間在「出錯」題目上，從而失去更多的作答時間，使 100 分鐘內不能完成 71 道題目。設同學 Z 現在只能完成 66 道題目（不計算出錯題目）。而同學 X 及同學 Y 因有多餘的時間花在「出錯」題目上，所以作答題數不受影響。

情況 2： 假設出錯的題目在試卷較後位置，同學 Z 沒時間碰上出錯題目，影響的只是時間充足考生作答時間，如同學 X 和 Y。根據以上兩種情況，我們會得出以下數據。以下圖表內的分數顯示學生得分與總分之比。

¹由於考試只有 100 分鐘，限時內只做到 71 道題目

	同學 X	同學 Y	同學 Z
原本所得分數	$\frac{90}{100}$	$\frac{70}{100}$	$\frac{35.5}{100}$
情況一（所有學生都嘗試做「出錯」的題目）			
「調整總分」後得分	$\frac{90}{99}$	$\frac{70}{99}$	$\frac{33}{99}$
「全體加分」後得分	$\frac{91}{100}$	$\frac{71}{100}$	$\frac{34}{100}$
情況二（同學 Z 沒有做「調整總分」）			
「調整總分」後得分	$\frac{90}{99}$	$\frac{70}{99}$	$\frac{35.5}{99}$
「全體加分」後得分	$\frac{91}{100}$	$\frac{71}{100}$	$\frac{36.5}{100}$

分析上表，在情況一下，相對於同學 X 和 Y，同學 Z 在分數上有著較大的影響；在情況二下，同學 Z 的影響則比較小。

總結

從上例看出，「出錯題目」的彌補方法不能單從平均值或標準差來考慮，因為涉及其他的變數太多。在例子中，同學有沒有作答「出錯」的題目、同學本身的能力以及花在思考「出錯」題目上的時間，均會對調分後的成績有著不同的影響。但這些因素卻很難以有系統的方法研究。所以，我們總結：若一份出錯題目的試卷流入試場，怎樣彌補也是不公平。因此，再三檢查題目，確保沒有任何問題，才是同學獲得公平考核的不二法門。

亞軍作品

一寸光陰一寸金

學校名稱：粉嶺禮賢會中學

學生姓名：黃熙舜

級別：中五

顧問老師：朱吉樑

引言

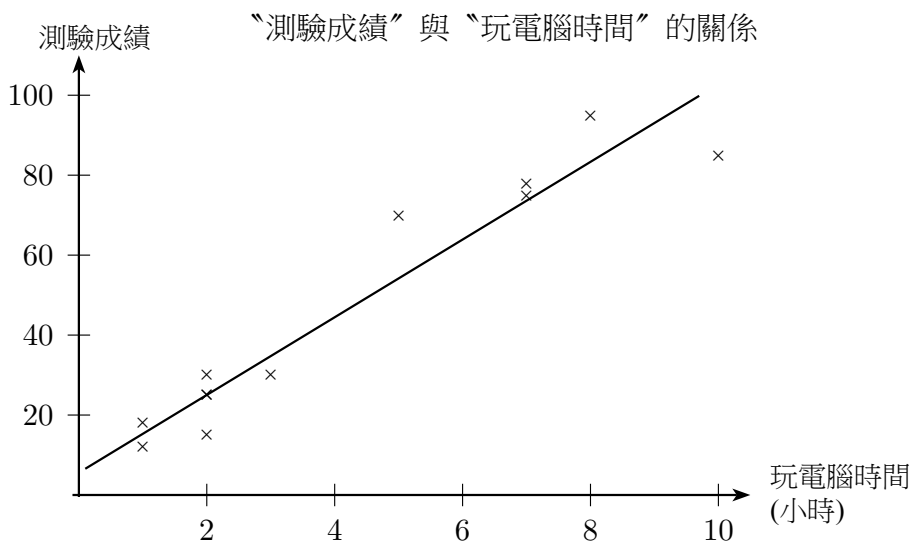
散點圖為中一數學課程中探討兩個變項之間的線性關係的方法。
熙舜以兩母子之間發生的故事，帶出誤用時所帶出的奇怪現象。

小舜是一個喜歡玩電腦遊戲的小朋友，每天都要使用電腦，否則便會心緒不寧。有一次，小舜的測驗成績非常差，當他回到家中，把測驗卷交給媽媽看時，媽媽對小舜說：「知不知道經常玩電腦的小朋友，成績會不斷退步？」小舜立即反駁媽媽，說：「我並不是因為玩電腦才成績差，只是題目太艱深。況且，玩電腦可以訓練手眼協調，保持頭腦清醒，成績反而提升呢？」

小舜媽媽沒有理會小舜的說話，並且立即沒收了小舜的電腦，弄得小舜又氣又急，小舜心想：「今天老師教的散點圖，正好可考慮兩個變項的關係。看看兩者是否真的有關係。」

於是，小舜將每天把活動記錄下來的時間表拿來細閱一會。他把數據以散點圖表示，發現了驚人的結果……小舜心想：「哼！當你看完這個結果，應該把電腦還給我。」小舜拿著圖表給媽媽看。

小舜：「媽，散點圖是看兩個變項的關係，你看看我把「玩電腦的時間」和「測驗成績」作比較，發現它們是正相關的，即是玩電腦的時間越多，測驗成績越好……」（見下圖，圖中的直線為趨勢線。）



小舜媽媽說：「你是不是說謊，這些數據是不是假的，快從實招來。」

小舜堅定地說：「我沒有說謊，你要求我每天記錄活動的時間，我把所需繪畫散點圖的資料組織了，現在給你看。」(見下表)

日期	9月10日	9月11日	9月12日	9月13日	9月14日	9月17日
玩電腦時間	1	7	2	8	2	10
成績	12	75	30	95	25	85

日期	9月18日	9月19日	9月20日	9月21日	9月24日	9月25日
玩電腦時間	2	5	1	7	3	2
成績	20	70	18	78	30	15

小舜說：「我沒有欺騙你。」

小舜媽媽心想：「從數據上看，圖像真的沒有問題，而且，小舜亦不敢做假，但是，感覺怪怪的，而結果亦不甚合理。」

小舜嚷著說：「快把電腦還給我，你也不想我成績退步吧！」

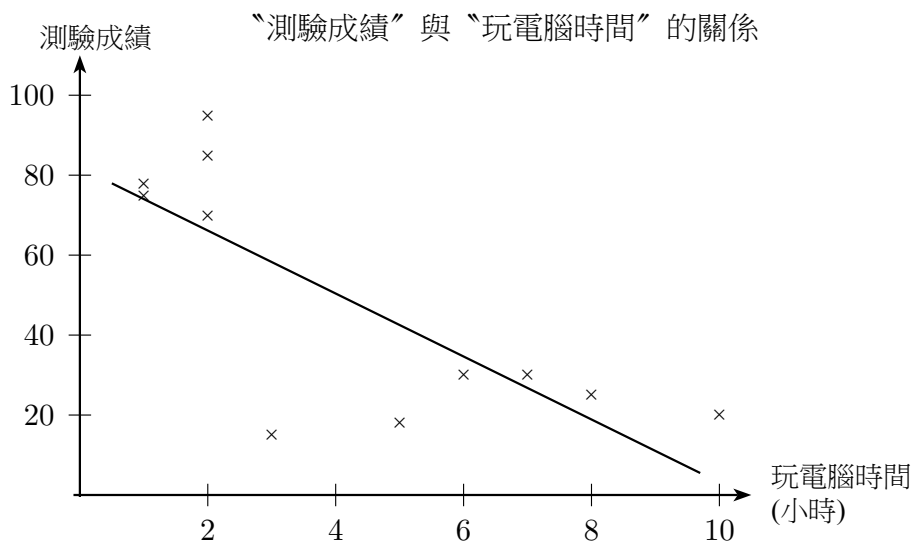
小舜媽媽心念一轉：「測驗成績應與溫習時間有一定關係，你把溫習時間也一併寫出來，若果你的發現正確，我便把電腦還給你。」

日期	9月10日	9月11日	9月12日	9月13日	9月14日	9月16日*	9月17日
溫習時間	5	1	6	1	6	7	0
玩電腦時間	1	7	2	8	2	2	10
成績	12	75	30	95	25	---	85
日期	9月18日	9月19日	9月20日	9月21日	9月23日*	9月24日	9月25日
溫習時間	6	2	5.5	2	2	1	5
玩電腦時間	2	5	1	7	6	3	2
成績	20	70	18	78	---	30	15

(*9月16及23日為星期日，所以沒有測驗。)

小舜媽媽亦是一個聰明人，當完整的時間表呈現出來，一眼便看出哪裡有漏洞。

小舜媽媽說：「你先看看真實的關係圖(見下圖)。」

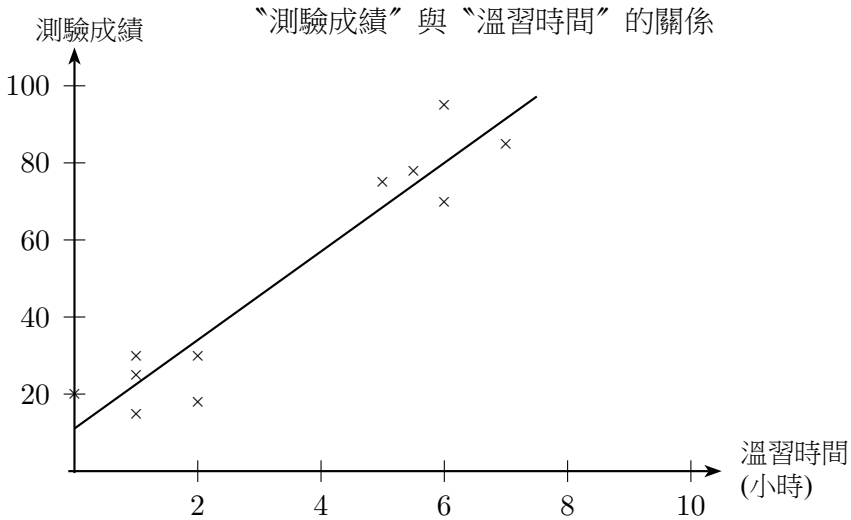


「你錯誤地運用了同一天的" 測驗成績" 和" 玩電腦時間"，要知道你當天的測驗是與你之前的活動有關，而不是當天，所以當你考慮相關性時，應把當天成績與前一天的活動掛鉤。(見下表)」，如此，你便可見到兩者是負相關。(見上圖)

日期	9月10日	9月11日	9月12日	9月13日	9月14日	9月16日*	9月17日
溫習時間	5	1	6	1	6	7	0
電腦時間	1	7	2	8	2	2	10
成績	12	75	30	95	25	---	85

日期	9月18日	9月19日	9月20日	9月21日	9月23日*	9月24日	9月25日
溫習時間	6	2	5.5	2	2	1	5
電腦時間	2	5	1	7	6	3	2
成績	20	70	18	78	---	30	15

小舜媽媽說：「更重要的，是你獲得的好成績是由於你用了較多溫習時間的成果(見下圖)。」



小舜媽媽說：「小舜，一寸光陰一寸金的道理，你還不知道麼！你若果上課時多留心，就不會犯以上的錯誤。看來，你的電腦還要給我管理好一段日子了。」

小舜：「媽媽，我知道了!」

季軍作品

絕代商驕之誰是主角

學校名稱：粉嶺禮賢會中學

學生姓名：凌家樑

級別：中五

顧問老師：朱吉樑

引言

在統計分析前，其中一個重要的考慮是統計量是否對準命題，就好比是否用合適的量度工具作量度。家樑被一句廣告口號「總有一間在身邊」吸引，在網上搜尋資料，發現網站中的解說，錯用了不合理的變量作比較，他以改編的故事把當中的問題帶出來討論，看看誰才是調查的主角。

相信大家都曾看過由黃子華與佘詩曼合演的「絕代商驕」，原來由黃子華扮演的麥提爽不單熟悉經濟學，還精通統計學。以下是一段發生在他們身上的小故事。

自從在麥提爽身上學懂不少關於經濟學的原理，林淼淼便開始自己的事業。她開了一間名為「1-17」的便利店。幾年下來，淼淼借「總有一間在身邊」的口號，在香港開了超過 500 間分店。不過有人嫉妒淼淼的事業，指出「總有一間在身邊」這口號是並不真實。也有人指出「1-17」的分店在部分地區太密集，分店雖多，但並不是大眾市民所能受惠。其後，淼淼便請教麥提爽。

淼： 爽哥！有事請教。

爽： 我已將平生所學有關做生意的知識教授給你，還有甚麼好說？

淼： 事情是這樣的……

一些人提出「1-17」在香港的密度不夠高，平均每 2.2 平方公里只有一間「1-17」。亦有人指「1-17」的分店數目在部分地區太密集，例如灣仔區，不能照顧普羅市民。我該怎樣做呢？

爽： 如何得出哪些數據？

淼： 在 2009 年 10 月，香港陸地總面積為 1,104.38 平方公里，我有 500 間「1-17」，平均來說，每 2.20876 平方公里即有一間「1-17」。灣仔區的面積有 9.94 平方公里，灣仔分店的數目有

40 間，即平均來說每 0.24850 平方公里即有一間「1-17」，似乎真是密集了些。

爽：大錯特錯。怎樣可以直接運用香港陸地總面積數據來計算。荒島可以很大，但缺少市民居住，「1-17」只是建於荒島又有何用？我們設立「1-17」的目的是為了方便消費者，並不是為土地服務。用地區面積計算「1-17」的密度是不合理的，量度對象不對。我們可以改為試用人口作分析。每區人口數目不同，「1-17」的數目亦不同，顧及每區人口情況來考慮店舖分佈相對較為合理。

淼：怎樣做才對？

爽：先看看用土地計算的問題，你看看香港十八區的地區面積、人口和「1-17」的數目吧！

(見下表)

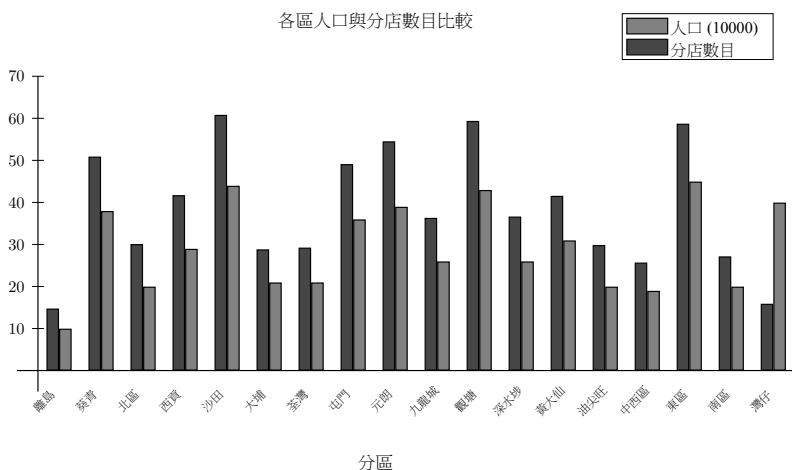
	離 島	葵 青	北 區	西 貢	沙 田	大 埔
人口數目	147 900	509 500	301 300	417 700	608 800	288 700
1-17 數目	10	38	20	29	44	21
人口除以 1-17 數目	14 790	13 408	15 065	14 403	13 836	13 748
	荃 灣	屯 門	元 朗	九龍城	觀 塘	深水埗
人口數目	292 900	491 600	545 800	363 700	594 400	366 800
1-17 數目	21	36	39	26	43	26
人口除以 1-17 數目	13 948	13 656	13 995	13 988	13 823	14 108
	黃大仙	油尖旺	中西區	東 區	南 區	灣 仔
人口數目	416 100	299 100	257 300	587 800	272 000	159 300
1-17 數目	31	20	19	45	20	40
人口除以 1-17 數目	13 423	14 955	13 542	13 062	13 600	3 983

註：表中人口為 2009 年年中數字，數字不包括公共機構/社團院舍的住院人士及水上居民。

	離 島	葵 青	北 區	西 貢	沙 田	大 埔
區議會地區面積 (km^2)	176.42	23.34	136.53	136.32	69.27	148.18
1-17 數目	10	38	20	29	44	21
地 區 面 積 除 以 1-17 數 目 (km^2)	17.642	0.614	6.827	4.701	1.574	7.056
	荃 灣	屯 門	元 朗	九龍城	觀 塘	深水埗
區議會地區面積 (km^2)	62.62	84.64	138.56	10.02	11.27	9.36
1-17 數目	21	36	39	26	43	26
地 區 面 積 除 以 1-17 數 目 (km^2)	2.982	2.351	3.553	0.385	0.262	0.360
	黃大仙	油尖旺	中西區	東 區	南 區	灣 仔
區議會地區面積 (km^2)	9.30	6.99	12.54	18.71	39.40	9.94
1-17 數目	31	20	19	45	20	40
地 區 面 積 除 以 1-17 數 目 (km^2)	0.300	0.350	0.660	0.416	1.970	0.249

註：表中區議會地區面積為 2009 年 10 月數字，數字不包括落馬洲河套區整幅面積 0.97 平方公里之土地。

爽：根據上表，香港某些地區的面積與人口比例的差異十分之大，例如離島區。雖然離島區的土地是全港最大的，但人口卻是全港最少的。如果我們以地區面積分析，平均每 17.642 平方公里才會有一間「1-17」，好像非常不合理。但我們不應只單獨考慮地區的面積，還可以考慮該區的人口數目。請看看下圖有關香港區議會地區的人口和「1-17」的分布：



爽：其實人口與「1-17」分店數目比例相若，平均約 13 000 人口就有一間「1-17」。以人口來看「1-17」的分布尚算合適。另一方面，對於「1-17」的分店數目在灣仔區太密集的批評，我們可這樣看。雖然灣仔區的人口數目只有 159 300，較離島區多出約 1 萬左右，但區內的「1-17」數目超過離島區的 3 倍。為甚麼呢？雖然灣仔區的人口數目較少，但其日間人流量卻十分多。中環廣場、銅鑼灣商業區、跑馬地馬場等地方都屬於灣仔區，名副其實是一個商業區域。平日上班，購物的人數均很多。人流多，自然生意也多。與香港其他地區不同，其他地區有不少住宅區，而人流量對灣仔區相對重要。人流量越多，生意額按理會越大，由於較容易得到灣仔區和南區「1-17」的生意額數據，因此，為簡單起見，我們以兩區於某一星期的生意總額代替人流量，作間接的比較（見下表）。

南區 (20 間)	Mon	Tue	Wed	Thu	Fri	Sat	Sun
每天該區生意額 (\$)	47,600	46,000	48,200	46,800	47,000	47,820	46,672
平均一間「1-17」 的每天生意額 (\$)	2,380	2,300	2,410	2,340	2,350	2,391	2,334
灣仔區 (40 間)	Mon	Tue	Wed	Thu	Fri	Sat	Sun
每天該區生意額 (\$)	148,000	147,420	147,330	148,200	193,208	187,560	196,000
平均一間「1-17」 的每天生意額 (\$)	3,700	3,686	3,683	3,705	4,830	4,689	4,900

根據上表的數據顯示，雖然灣仔區面積較小，但人流量多，生意額大，所以開 40 間分店是合理的。這證明了「1-17」的分店數目在部分地區太密集的說法是錯誤的。

運用統計學最重要的地方就是調查對象要正確，如果對象不正確，無論多簡單的問題都會錯，而且會發生不少的誤會呢！

淼：知道了，讓我回去反擊他們的論說！

參考資料：

1. 維基網站：<http://zh.wikipedia.org/wiki/%E9%A6%99%E6%B8%AF>
2. 社會福利署網頁：http://www.swd.gov.hk/tc/index/site_district/page_cwsi/sub_districtpr2/
3. 地政總署測繪處, 香港地理資料 (包括區議會地區面積) 網頁：<http://www.landsd.gov.hk/mapping/tc/publications/publications.htm>
4. 政府統計處, 統計地圖 -最新人口及住戶統計數字 (2009) 網頁：http://www.censtatd.gov.hk/hong_kong_statistics/statistics_on_map/index_tc.jsp

殿軍作品

「校本驗毒計劃」中的概率應用

學校名稱：香港神託會培基書院

學生姓名：黎子婧、張顯文

級別：中四

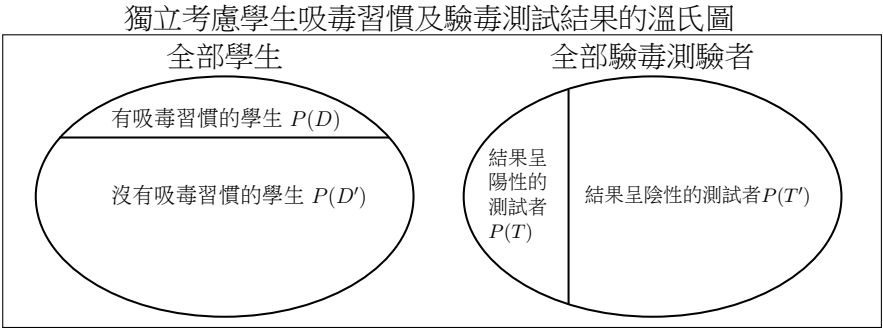
顧問老師：黃智君

引言

青少年吸食毒品的問題日益嚴重，相關的負面新聞不斷曝光，而吸毒地點更包括校園及公共場所。為遏止毒風在校園蔓延，政府高姿態地提出了「校本驗毒計劃」，並選擇以大埔區作為計劃的試點。社會各界對此反應不一，有學者認為計劃能收預防之用，可有效地提升青少年的抗毒意識，亦有學生團體關注計劃的假陽性、私隱及配套等問題。透過統計學的分析及概率的應用，本文正要為這個炙手可熱的問題把脈。



若我們先排除私隱、法律、輔導等等相關的顧慮，單從統計學的角度去分析「校本驗毒計劃」，這個概率的問題並不複雜。透過以下的溫氏圖，我們可以作更深入的探究。無論我們獨立地或是聯合地考慮學生吸毒習慣及驗毒測試結果，每幅溫氏圖內的概率總和均等於一，即代表圖中包含所有學生或驗毒測試者的樣本。



聯合考慮學生吸毒習慣及驗毒測試的結果

		驗毒測試	
		陽性 (T)	陰性 (T')
吸 毒 習 慣	有 (D)	學生有吸毒習慣，並於驗毒測試中呈陽性反應 $P(D \cap T)$	學生有吸毒習慣，並於驗毒測試中呈陰性反應 $P(D \cap T')$
	沒有 (D')	學生沒有吸毒習慣，並於驗毒測試中呈陽性反應 $P(D' \cap T)$	學生沒有吸毒習慣，並於驗毒測試中呈陰性反應 $P(D' \cap T')$

註: 一、 $D(\text{Drug})$ 代表學生有吸毒習慣，而 D' 則代表學生沒有吸毒習慣。
 二、 $T(\text{Test})$ 代表驗毒測試呈陽性反應，而 T' 則代表驗毒測試呈陰性反應。

為使問題更易理解，本文先考慮驗毒計劃只作單次性樣本測試的情況，並作以下的假設：首先假設學生有吸毒習慣的概率

$P(D) = 0.01$ ，由於所有樣本的概率總和等於一，所以學生沒有吸毒習慣的概率 $P(D') = 1 - 0.01 = 0.99$ 。因為驗毒測試對少數較罕見的毒品種類無效，並只能驗出學生在數天內有否吸毒，所以本文假設若學生有吸毒習慣，而其驗毒測試結果呈陽性反應的概率 $P(T | D) = 0.95$ 。雖然其準確率很高，但在某些特殊的情況下，驗毒測試亦會出現偏差，例如一些曾進食罌粟籽麵包或火麻仁油的人，卻會無辜地被驗出曾吸食毒品，現假設類似情況的概率 $P(T | D') = 0.02$ 。

若將學生吸毒習慣及驗毒測試結果作同時考慮，我們只要將樹形圖（見附圖）上兩者的概率相乘，便可以分別計算出以上四個情況的聯合概率。現以學生有吸毒習慣，並於驗毒測試中呈陽性反應的概率為例：

$$\begin{aligned}P(D \cap T) &= P(D) \times P(T | D) \\&= 0.01 \times 0.95 \\&= 0.0095\end{aligned}$$

從另一個角度看，若果某學生剛完成了驗毒測試，而其結果呈陽性反應，這學生真的有吸毒習慣的機會有多大呢？解答這個問題需要應用到條件概率及貝氏定理。現先解釋何為條件概率，若在已知 B 事件發生了的前題下，另一事件 A 又同時發生的概率，即「在 B 條件下發生 A 的條件概率」，可標記成 $P(A | B)$ ，其計算方法為 $P(A | B) = P(A \cap B) \div P(B)$ 。另外，貝氏定理則是結合事前概率及條件概率，以計算出事後概率的過程，其算式

為 $P(A | B) = \frac{P(B|A) \times P(A)}{P(B|A) \times P(A) + P(B|A') \times P(A')}$ 。在「校本驗毒計劃」的例子中，若在已知的驗毒結果下，我們便可找出學生吸毒習慣的概率，並可歸納成真陽性、假陰性、假陽性及真陰性四個情況，現以 $P(T)$ 及 $P(D' | T)$ 的計算作說明。

驗毒測試結果呈陽性的概率

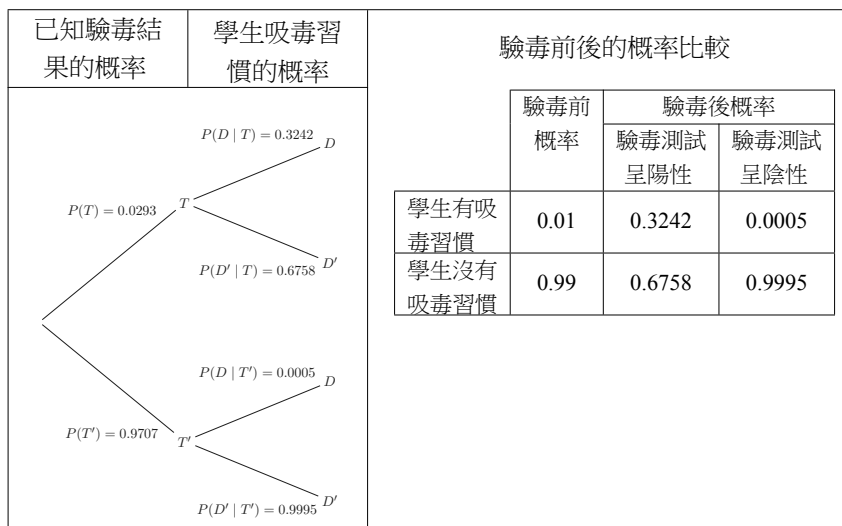
$$\begin{aligned}
 &= P(T) \\
 &= P(D \cap T) + P(D' \cap T) \\
 &= P(D) \times P(T | D) + P(D') \times P(T | D') \\
 &= 0.0095 + 0.0198 \\
 &= 0.0293
 \end{aligned}$$

假陽性的概率

$$\begin{aligned}
 &= \text{在已知陽性驗毒結果下，學生沒有吸毒習慣的概率} \\
 &= P(D' | T) \\
 &= P(D' \cap T) \div P(T) \\
 &= P(D') \times P(T | D') \div P(T) \\
 &= 0.0198 \div 0.0293 \\
 &= 0.6758(\text{取小數點後四個位})
 \end{aligned}$$

事前概率 (在進行驗毒測試前，學生吸毒習慣的概率)	新的訊息 (驗毒測試結果的概率)	聯合概率 (同時考慮吸毒習慣及驗毒結果的概率)	事後概率 (在已知的驗毒結果下，學生吸毒習慣的概率) (取小數點後四個位)
		學生有吸毒習慣，並於驗毒測試中呈陽性反應的概率 $= P(D \cap T)$ $= 0.0095$	真陽性：在已知陽性驗毒結果下，學生有吸毒習慣的概率 $= P(D T)$ $= 0.3242$
		學生有吸毒習慣，並於驗毒測試中呈陰性反應的概率 $= P(D \cap T')$ $= 0.0005$	假陰性：在已知陰性驗毒結果下，學生有吸毒習慣的概率 $= P(D T')$ $= 0.0005$
		學生沒有吸毒習慣，並於驗毒測試中呈陽性反應的概率 $= P(D' \cap T)$ $= 0.0198$	假陽性：在已知陽性驗毒結果下，學生沒有吸毒習慣的概率 $= P(D' T)$ $= 0.6758$
		學生沒有吸毒習慣，並於驗毒測試中呈陰性反應的概率 $= P(D' \cap T')$ $= 0.9702$	真陰性：在已知陰性驗毒結果下，學生沒有吸毒習慣的概率 $= P(D' T')$ $= 0.9995$

經計算後所得的概率可用下列的圖表及樹形圖作表達，以便比較驗毒前後的概率；在新的樹形圖中，因為驗毒結果屬於已知的事件，所以驗毒結果 (T) 及吸毒習慣 (D) 的位置將會倒轉，驗毒結果的概率會先置於圖的左方。



作為中學生的筆者最關注的莫過於計劃中私隱及假陽性的問題。如果自己的測試結果真的出現了假陽性，必會擔心自己被父母及同學誤會。究竟，校本驗毒測試的準確度有多高呢？在這個例子中，在驗毒前學生有吸毒習慣的概率是 0.01，而經過檢查判斷為"陽性"的概率則為 0.3242。在測試結果呈陽性反應的學生當中，真正有吸毒習慣的只有 0.3242，而其餘的學生 (0.6758) 則會被冤枉成隱君子，這現象稱為「假陽性詭論」。為何驗毒測試顯示學生曾服食毒品，但他們實際上並沒有吸毒習慣呢？

為剖析假陽性的現象，我們必須追索到計算假陽性概率時的分子 $P(D')$ ，即驗毒測試結果呈陽性的概率。由於本文假設 $P(D) = 0.01$ 及 $P(T | D') = 0.02$ ，以致計算出來 $P(T)$ 的數值極小；在分母極小的情況下，只要分子在數值上有輕微的改變，亦會對假陽性的概率 $P(D' | T)$ 起舉足輕重的作用。如果有吸毒習慣的

學生比本文假設的 0.01 更少，出現假陽性的機會將會更大。在假陽性的情況下，部分沒有吸毒的學生可能會被冤枉，但這樣的打擊又豈是人人都能承受得到呢？為免驗毒結果呈陽性的學生被標籤，政府必須正視學生私隱的保障及驗毒後的輔導工作等相關配套措施；另一方面，為減少出現假陽性的機會，政府宜採用不同牌子的驗毒工具，為首次測試呈陽性的學生，即場進行另一次的快速複驗。

本文以「校本驗毒計劃」為例，顯示了統計學及概率在日常生活中的廣泛應用，而相關的分析能有助優化計劃的細節及相關的配套措施。其實，政府推行「校本驗毒計劃」的目的並不是要拘捕有吸毒習慣的學生，而是透過驗毒計劃起預防作用，以遏止毒風在校園蔓延。另外，政府可通過驗毒計劃中所得的數據，更準確地掌握香港學生的吸毒情況，並能探討青少年沉淪毒海的原因，繼而進一步採取具針對性的禁毒措施。

參考資料：

1. 牛津/勤達數學系列—新世紀數學 (高中 M1B 冊)，第 23 至 25 頁。
2. 您真的被檢驗出疾病了嗎？條件機率與貝氏定理。李幸娟 (應數博 96)。

優異作品

等「你」等到我心痛

學校名稱：粉嶺禮賢會中學

學生姓名：朱婉儀、蕭智汶

級別：中五

顧問老師：朱吉樑

引言

婉儀和智汶透過一個簡單故事，把加權平均數和標準差的誤用與應用，以漫畫表現出來，提醒我們處理統計數字時要格外留神。

人物介紹

阿純



人如其名，單純的好人，做事細心認真，容易緊張，愛未雨綢繆。

阿才



自稱「數學天才」，生性懶惰，愛以數學知識捉弄人。

安琪



天使一名，熱心助人，充滿智慧。

阿純和阿才要做經濟科專題報告，
所以相約 9:30 於粉嶺圖書館門口







不是你的錯，是阿才捉弄你。讓我們來看看以下數據：

21 號巴士

時段	繁忙	正常	非繁忙
平均候車時間(分鐘)	5	10	30
百分比	20%	60%	20%

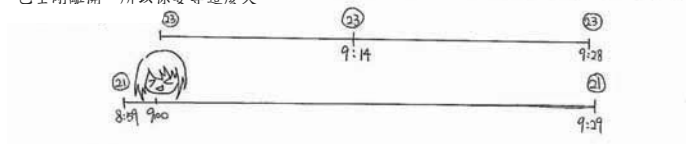
因為 21 號巴士是分繁忙、正常和非繁忙時段行駛的。繁忙時段平均候車時間為 5 分鐘，約佔整個服務時段的 20%，而正常時段平均候車時間為 10 分鐘，約佔整個服務時段的 60%，最後，非繁忙時段平均候車時間為 30 分鐘，約佔整個服務時段的 20%。

所以，21 號平均候車時間為

$$= 5 \times 20\% + 10 \times 60\% + 30 \times 20\%$$

$$= 13 \text{ 分鐘}$$

另外，23 號巴士是沒有劃分繁忙、正常或非繁忙時段行車的。平均候車時間是 14 分鐘。而非繁忙時段由早上 9:00 開始，所以你要等 30 分鐘才有一班車。你抵達車站時，23 號巴士剛離開，所以你要等這麼久。





2:30

完成

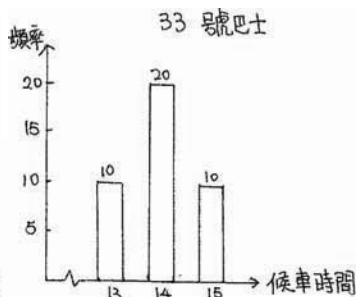
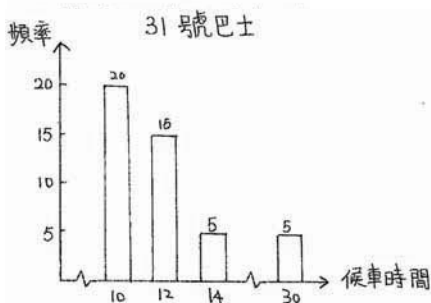


註：為方便考慮，此部分故事假設 31 及 33 號巴士沒有劃分「繁忙與非繁忙時段」。





阿才沒有騙你，他真的是為你好，想你快點到達，但是你今天就是運氣不佳。事實上，巴士的候車時間會受到不同路線影響的。比如行車路線上紅綠燈的數量、路線上車輛的數量。另外，由於 33 號巴士所行駛的是高速公路，所以較其他巴士準時。約 14 分鐘便有一班車，看看下圖的模擬真實情況：



假設 31 號巴士候車時間為 10、12、14、30 分鐘，該時段就相對有 20、15、5 和 5 輛車，而 33 號巴士候車時間為 13、14 和 15 分鐘，該時段就相對有 10、20 和 10 輛車。然後計算它們候車時間的平均值，

$$31 \text{ 號巴士的平均候車時間} = \frac{10 \times 20 + 12 \times 15 + 14 \times 5 + 30 \times 5}{45} = 13.33 \text{ 分鐘}$$

$$33 \text{ 號巴士的平均候車時間} = \frac{10 \times 13 + 20 \times 14 + 10 \times 15}{40} = 14 \text{ 分鐘}$$

它們之間的平均值相差僅 1 分鐘，計算它們的標準差：

31 號巴士的標準差

$$= \sqrt{\frac{20 \times (10 - 13.33)^2 + 15 \times (12 - 13.33)^2 + 5 \times (14 - 13.33)^2 + 5 \times (30 - 13.33)^2}{45}}$$

$$\approx 6.04 \text{ (三位有效數字)}$$

33 號巴士的標準差

$$= \sqrt{\frac{10 \times (13 - 14)^2 + 20 \times (14 - 14)^2 + 10 \times (15 - 14)^2}{40}}$$

$$\approx 0.707 \text{ (三位有效數字)}$$

由此可見，31 號巴士的候車時間的標準差較 33 號巴士的候車時間的標準差大了很多。今天你真倒霉，碰巧遇到需等 30 分鐘才有一班車的時段。



優異作品

哪一間補習社較好呢？

學校名稱：東華三院李潤田紀念中學

學生姓名：黃澄珊、韓靜文、吳明慧

級別：中四

顧問老師：陳浩文

引言

現今大部分香港中學生均會於課餘進行補習。補習社對於他們來說變成生活的一部分。究竟他們是如何揀選補習社呢？是否只單憑廣告的描述已足夠呢？

優異作品：哪一間補習社較好呢？

現今，林林總總的補習社無處不在，其廣告可謂隨處可見。補習社主要以星級名師、精美筆記、奪 A 數字、獎學金及各種優惠等等來吸引學生報讀。學生們為了能在公開試奪得佳績，加上抱著人補我補的羊群心理，因此紛紛去報讀補習班，使補習的人數不斷上升，成為社會上的熱潮，等同於全民皆股的現象。可是，每間補習社都聲稱自己的公開試成績最佳，那麼怎樣分辨真假，當中又是否有任何誤導的成份呢？

以下是 ABC 補習社於今天在報紙所刊登的資料（資料一）：



3000 名學生考獲 A、B、C

	考獲 A、B、C 人數	進步人數	會考未能考獲 A、B、C 人數*
2009 年	3000 人	300 人	只有 3 人(全部均獲 D 級成績)
2008 年	2000 人	200 人	只有 3 人(全部均獲 D 級成績)
2007 年	1000 人	100 人	只有 5 人(全部均獲 D 級成績)

*符合成績保證計劃學生：學生修畢 Ben Sir 整個會考課程並於學期末之唯一一次多項選擇題測驗中取得 80 分(100 分滿分)或以上。(所有有關題目均曾於課堂上講解，絕不加入新題目。)

以下是 XYZ 補習社於今天在報紙所刊登的資料（資料二）：



蟬聯全港最多 A 09 高考及會考勇奪

3000A

考獲 A 級成績的學生更可獲得\$500 獎學金(以單科計)

學生的家庭月入必須低於\$10,000 才符合領取獎學金的資格

ABC 補習社標榜自己的學生於 2009 年會考考獲 A、B、C 的人數有 3000 人，為全港之冠。XYZ 補習社則聲稱自己於 2009 年度蟬

聯全港最多 A（包括高考及會考），數目達 3000 個。究竟哪一間補習社較好呢？

根據資料一，我們可以看見考獲 A、B、C 的人數從 2007 年的 1000 人升至 2009 年的 3000 人；進步人數則從 2007 年的 100 人升至 2009 年的 300 人；會考未能考獲 A、B、C 的人數就從 2007 年的 5 人下降至 2009 年的 3 人，而且當中全部均考獲 D 級成績。從整體的數據來看，該補習社的學生在公開試的表現的確不斷上升。可是，當中可能涉及很多的誤導成份。

例如：

1. 2009 年有 3000 人考獲 A、B 及 C，但我們不知道當中的比重是多少，可以是 A：2000 人，B：700 人，C：300 人；也可以是 A：300 人，B：200 人，C：2500 人。兩者分別可謂是差天共地。
2. 該補習社又用字體的大小及顏色來誤導讀者。2009 年只有 3 人未能考獲 A、B、C，但仍有 D 級成績。可是，在表格的底下印有一段文字，字體細小而且十分模糊，一般人都不會察覺，因此往往被人誤導。那段文字的意思是指在符合成為保證計劃的學生當中，未能考獲 A、B、C 的人數只有 3 人，而且全部均獲得 D 級成績。事實上，我們只要細心一想，就知道符合該計劃的學生的成績必定不俗，考獲 D 級不是什麼奇怪的事。如果沒有看見那段文字的人，便會以為該補習社未

能考獲 A、B、C 的學生人數只有 3 人。此外，該補習社又故意將學生在 2009 年會考的表現放大和變成紅色，以吸引讀者的注意力。

3. 學生的進步人數在 3 年內不斷上升，這可能是因為補習社會要求那些成績低劣的學生退出，目的是使各方面的數據更為理想，可以吸引更多學生報讀。這裡又帶出一個問題，補習社如何界定學生是否進步呢？進步 1 分就算進步？補習社並沒有在這方面提供資料。
4. 補習社並沒有提供原來報讀人數資料。假如在 2007 年報讀的人數有 1200 人，而考獲 A、B、C 人數有 1000 人，則在當年考獲 A、B、C 的百分率是 83.3%；而在 2009 年報讀的人數可能有 10000 人，而考獲 A、B、C 人數有 3000 人，則在當年考獲 A、B、C 的百分率是 30%。那麼，2009 年考獲 A、B、C 的百分率還要較 2007 年的低。因此原來的報讀人數亦是非常重要的。

資料二對學生來說是十分吸引，可是同樣地也包含了不少誤導成份。

例如：

1. 同樣地，XYZ 利用顏色和字體來誤導讀者，與上者相若。

2. 該補習社聲稱自己是會考及高考的全港奪 A 之冠，數目有 3000A。可是，這樣亦未必能完全反映真實的情況。根據一些雜誌的報導，報讀該補習社的學生每人可能只報讀了一科，而其他科都是靠自己溫習的，但該補習社的奪 A 數字可能亦會包括一些未有報讀而又考得 A 的科目在內。假若該補習社有 210 名學生考獲 10A（會考），150 名學生考獲 6A（高考），在這 360 名學生之中，他們奪 A 數字已達 3000，但他們每人可能只報讀了一科，那麼，該補習社的奪 A 數字事實上只有 360 個。

因此，僅靠那些補習社在報紙上所提供的統計資料，是不足以去比較它們的好壞。此外，當中資料可能包含了極多的誤導成份，所以容易造成誤會。因此，我們不應盡信廣告，應加以分析。我們又可以在報讀前詢問一下老師的意見和有補習的同學有關的資料，這樣便是萬全之策。

參考資料：

1. 《頭條日報》

優異作品：哪一間補習社較好呢？

優異作品

你今日驗咗未呀？—校園驗毒計劃成效預測

學校名稱：聖公會李福慶中學

學生姓名：朱韋謙、李志雄、伍錦煊

級別：中七

顧問老師：楊定邦

引言

學生吸毒情況於過去數年越來越嚴重。有鑑於此，官方打算於校園推行自願驗毒計畫。究竟成效如何呢？

校園吸毒的情況日益嚴重，香港特區政府有意遏止青少年吸毒的歪風盛行，故於本年度推行校園驗毒計劃。香港教育局及保安局禁毒處於二零零九年八月份公佈，大埔區的廿三所中學最快於十二月起，率先推行校園驗毒試行計劃，為期半年。由醫藥人員、社工及數據管理員組成的專業隊伍每月平均到校兩次，抽查 32 至 40 名學生（以 800 名參與學生為例）。檢驗前須先徵得學生及家長同意，校方將提供已簽署同意書的學生名單讓專業隊伍隨機抽查，學生被分為中一至中五，以及中六至中七兩組，抽查人數為 2：1。曾作毒品測試的同學不會從名單中剔除，除非已被驗出有吸食毒品。若結果呈陽性，樣本會送往化驗室再次測驗以確認結果，而檢驗結果將保密，警方並不會提出檢控。

現在，讓我們來探討大埔區中學願意接受驗毒的中學生佔區內中學學生總數的百分比、驗出吸毒學生的人數佔總抽查人數的百分比、驗毒計劃能驗出有吸毒學生的機會率，以及半年內能驗出吸毒學生人數的期望值。

為了方便計算，我們作了以下假設：

假設一： 每間中學學生總人數為 1000 人

假設二： 大埔區每間中學吸毒的中學生為校內中學生總數的 5%

假設三： 每間中學每月隨機抽樣調查 36 人（資料提供的數據為 32 – 40 人（以 800 名參與學生計算））

假設四： 學生只分為一組，所有級別的學生被抽查的機會都是均等

假設五： 每所中學內，假設有吸毒的學生 $100y\%$ 願意參與該計劃， $100(1-y)\%$ 則不願意；至於不吸毒的學生假設有 $100(1-x)\%$ 願意參與， $100x\%$ 不願意

假設六： 被抽中的學生必定會參與驗毒而不會臨時反悔

假設七： 驗毒結果是 100% 準確

首先，根據禁毒常務委員會早前所做的調查發現，有八成大埔區的學生願意參與計劃，因此作出假設八：區內每間中學均有八成學生願意參與計劃。

但是我們並沒有數據顯示當中吸毒學生所佔的比例，因此我們設定三個有可能的個案，計算吸毒及不吸毒的學生願意及不願意接受測試的比例：

個案一：當 $x = y$

設願意參與驗毒的吸毒學生比例為 $100y\%$ ，在校內 1000 位學生中，隨機抽出一位學生

$$P(\text{願意參與計劃}) = P(\text{吸毒} \cap \text{願意}) + P(\text{不吸毒} \cap \text{願意})$$

$$0.8 = 0.05y + (1 - 0.05) \times (1 - y)$$

$$0.8 = 0.05y + 0.95 \times (1 - y)$$

$$0.8 = 0.05y + 0.95 - 0.95y$$

$$-0.15 = -0.9y$$

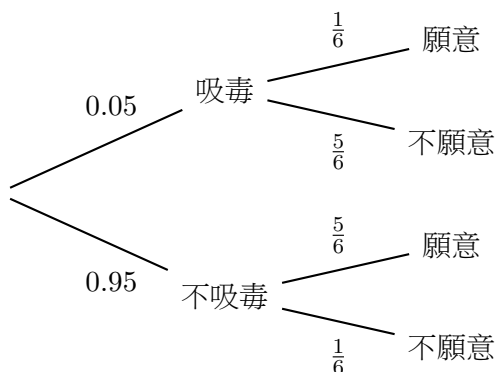
$$y = \frac{1}{6}$$

$\therefore \frac{1}{6}$ 吸毒學生願意接受驗毒，而不吸毒的學生中， $\frac{5}{6}$ 願意接受驗

毒。

如果第一次驗毒結果顯示有一名學生吸毒，在第二次驗毒中抽中吸毒學生的機會率會相應減少，如此類推。為方便計算，我們作出假設九：在每一次驗毒中抽中吸毒學生的機會率都相同。

接著，我們建議用樹形圖（如下圖）去處理以上較繁複的數據：



$$P(\text{抽查的學生發現有吸毒}) = 0.05 \times \frac{1}{6} \div 0.8 = 0.01042$$

從上述結果顯示，抽查的學生發現有吸毒的機會率為 0.01042。換言之，在願意接受驗毒的學生中抽取 36 名學生而被驗出有吸毒的人數為「二項分佈」。以一間有 1000 名學生的學校來計算：

每個月能驗出吸毒學生人數的期望值 = $36 \times 0.01042 = 0.375$ 人

假設十：區內每間中學半年內每個月均有 36 名學生被抽查驗毒

再計算每間學校於計劃所推行的半年內能驗出吸毒學生的期望值 = $0.375 \times 6 = 2.25$ 人

所以，整個大埔區內 23 間學校半年內能驗出吸毒學生人數的期望值 = $2.25 \times 23 = 51.75$ 人

最後，計算在大埔區中學驗出吸毒的學生佔總抽查人數的百分比
 $= 51.75 \div (36 \times 6 \times 23) \times 100\% = 1.04\%$

個案二：當 $x = 0.8y$

設願意參與驗毒的吸毒學生比例為 y ，在校內 1000 位學生中，隨機抽出一位學生

$$P(\text{願意參與計劃}) = P(\text{吸毒} \cap \text{願意}) + P(\text{不吸毒} \cap \text{願意})$$

$$0.8 = 0.05y + (1-0.05) \times (1-x)$$

$$0.8 = 0.05y + 0.95 \times (1-0.8y)$$

$$y = \frac{15}{71}$$

$$\text{以及 } x = 0.8 \times \frac{15}{71}$$

$$x = \frac{12}{71}$$

$\therefore \frac{15}{71}$ 吸毒學生願意接受驗毒，而不吸毒的學生中， $\frac{59}{71}$ 願意接受驗毒。

$$P(\text{抽查的學生發現有吸毒}) = 0.05 \times \frac{15}{71} \div 0.8 = 0.0132$$

每個月能驗出吸毒學生人數的期望值

$$= 36 \times 0.0132$$

$$= 0.4752 \text{ 人}$$

每間中學於計劃所推行的半年內能驗出吸毒學生人數的期望值

$$= 0.4752 \times 6$$

$$= 2.8512 \text{ 人}$$

整個大埔區內 23 間中學半年內能驗出吸毒學生人數的期望值

$$= 2.8512 \times 23$$

$$= 65.5776 \text{ 人}$$

∴ 最後，計算在大埔區中學驗出吸毒的學生佔總抽查人數的百分比

$$= 65.5776 \div (36 \times 6 \times 23) \times 100\% = 1.32\%$$

個案三：當 $x = 1.2y$

設願意參與驗毒的吸毒學生比例為 y ，在校內 1000 位學生中，隨機抽出一位學生

$$P(\text{願意參與計劃}) = P(\text{吸毒} \cap \text{願意}) + P(\text{不吸毒} \cap \text{願意})$$

$$0.8 = 0.05y + (1-0.05) \times (1-x)$$

$$0.8 = 0.05y + 0.95 \times (1-1.2y)$$

$$y = \frac{15}{109}$$

$$\text{以及 } x = 1.2 \times \frac{15}{109}$$

$$x = \frac{18}{109}$$

∴ $\frac{15}{109}$ 吸毒學生願意接受驗毒，而不吸毒的學生中， $\frac{91}{109}$ 願意接受驗毒。

$$P(\text{抽查的學生發現有吸毒}) = 0.05 \times \frac{15}{109} \div 0.8 = 0.0086$$

每個月能驗出吸毒學生人數的期望值

$$= 36 \times 0.0086$$

$$= 0.3096 \text{ 人}$$

每間中學於計劃所推行的半年內能驗出吸毒學生人數的期望值

$$= 0.3096 \times 6$$

$$= 1.8576 \text{ 人}$$

整個大埔區內 23 間中學半年內能驗出吸毒學生人數的期望值

$$= 1.8576 \times 23$$

$$= 42.7248 \text{ 人}$$

∴ 最後，計算在大埔區中學驗出吸毒的學生佔總抽查人數的百分比 $= 42.7248 \div (36 \times 6 \times 23) \times 100\% = 0.8600\%$

以上不同的個案中顯示當 x 相對 y 越少，即相對越多吸毒學生願意接受驗毒的話，成效就越大。但學生的意願還是優先被考慮，官方絕不強逼學生參與。雖然這次大埔區校園驗毒計劃表面上成效很低，但其實這次驗毒計劃的真正目的不在於驗出所有吸毒的學生，而是為那些吸毒的學生帶來阻嚇性，及對有意戒毒的學生給予幫助，讓他們能儘快重返正途。

參考資料：

1. 商業電台（新聞） <http://www.881903.com/Page/ZH-TW/news-detail.aspx?ItemId=155527>
2. 保安局禁毒處 http://www.nd.gov.hk/tc/school_drug_test_tp.htm

優異作品：你今日驗咗未呀？－校園驗毒計劃成效預測

優異作品

數字不會說謊！？

學校名稱：保良局八三年總理中學

學生姓名：陳愛珠、方雅儀、鍾小曼

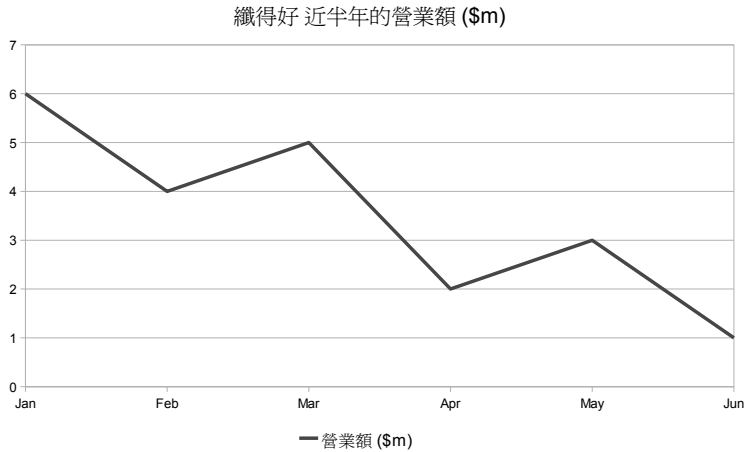
級別：中四

顧問老師：林惠邦

引言

現今的香港人較以往更加注重儀容，很多人尤其是女士會想盡辦法纖體。她們會透過纖體廣告揀選纖體公司，這種做法是否可取？對她們是否有利呢？

「特務甲乙丙顧問公司」是一間專門利用圖表來拯救經營出現問題公司的顧問公司。



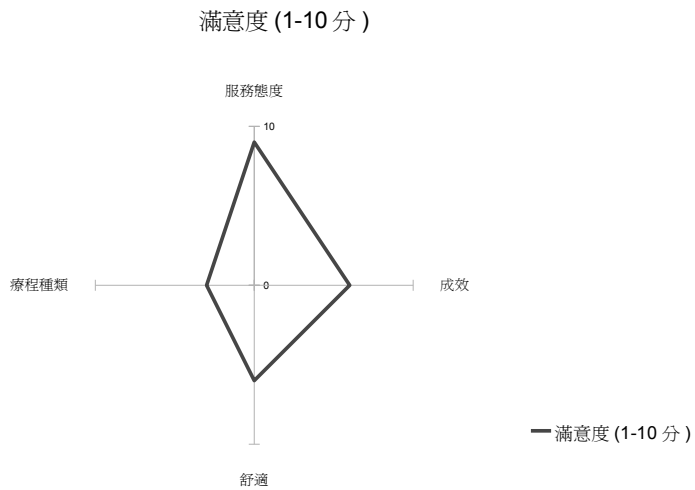
經過特務甲地毯式搜索資料後，發現「纖得好」營業額逆市下降，正面臨嚴重的財政危機，岌岌可危。因此，顧問公司派出特務乙，作詳細的調查和分析，發現原因只有一個——宣傳方式錯誤。以下是它對外宣傳的數據和圖表：

資料 1：最近 1 個月的 10 個顧客

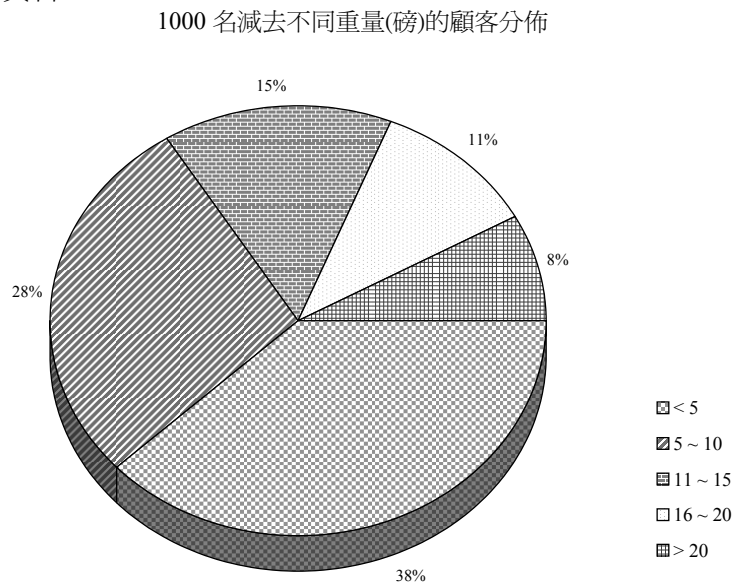
顧客	1	2	3	4	5	6	7	8	9	10
減去重量 (磅)	3	3	3	6	10	15	7	12	2	20

眾數 = 3 磅

資料 2



資料 3



經過特務乙仔細分析後，發現「纖得好」的行政人員太誠實，發放對公司不利的訊息。

首先，此公司根據資料 1，對外聲稱大部分顧客都可在一個月內減去 3 磅，但數值毫不吸引，與其他競爭者相比，成效不但不大，反而突顯此公司成績甚為差劣，商譽下降，營業額自然下跌。

其次，此公司利用了雷達圖表示出顧客的滿意程度，當中的對稱性不大，令此公司暴露了缺點，給顧客的印象也不太好。

此外，此公司將一千名減去不同重量的顧客分佈作一個圓形圖，由於數據太多，而且數據偏向成效差的一方，使圖像失去宣傳的意義，不能突顯此公司的成效，顧客自然不會考慮。

特務丙詳細考慮特務乙呈交的分析報告，作出以下的改善：

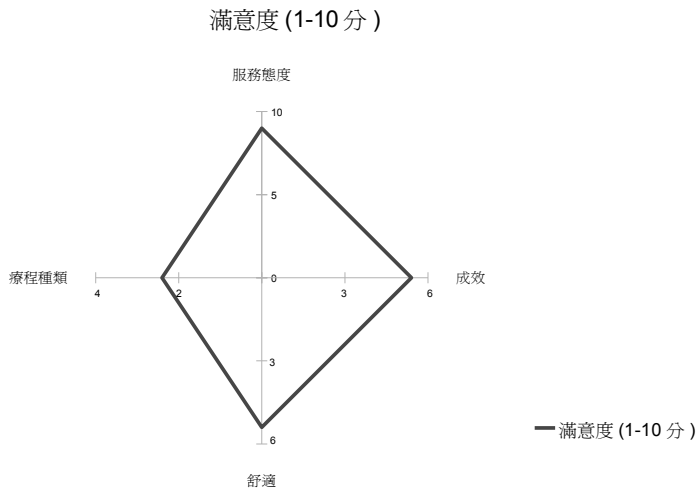
首先，資料 1 顯示顧客能減去的重量偏低，令此公司不能突圍而出，是因為它選用了眾數（出現最多次的數字）作為參考指標。事實上，它可以選擇選用平均數（計算時將所有數目加起來再除以數值的數量）或中位數（把所有觀察值高低排序後找出正中間的一個數值）作為表達方式。根據我們的計算：

$$\text{平均數} = \frac{3+3+3+6+10+15+7+12+2+20}{10} = 8.1 \text{ 磅}$$

$$\text{中位數} = 6.5 \text{ 磅}$$

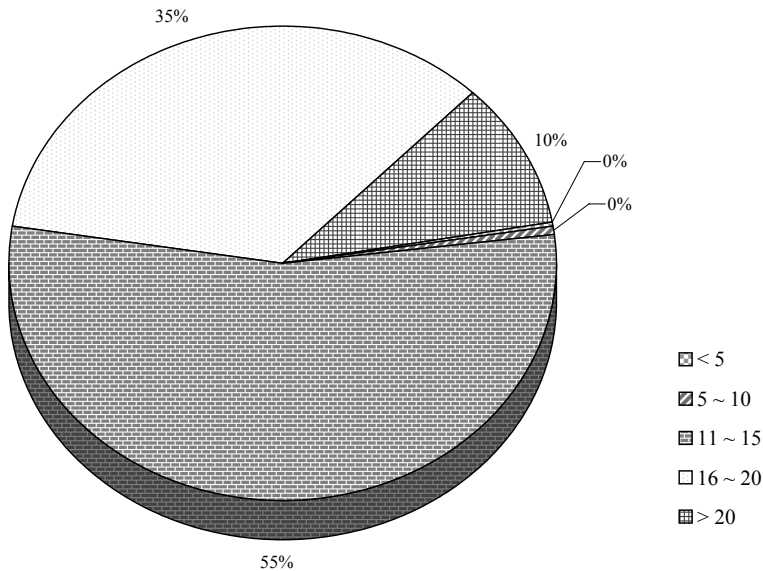
由於這組數據中有一些比較極端的數值，因此選用平均數較能誤導消費者，達到宣傳的作用。

其次，資料 2 選用的雷達圖（透過軸線與格線，讓人能依據許多準則來描述某件事的圖像）圖形的對稱性不夠大。由於雷達圖的主要精髓是追求對稱，它的構圖形式主要就是希望能達成平衡與調和，愈為規則的數據組，代表它在各不同準則之間是最平衡的。所以當資料 2 顯示的圖像愈為對稱，就表示此公司在各方面都表現良好。只要我們稍為調整雷達圖的刻度，就能達致宣傳的效果，效果如下：



最後，資料 3 的圓形圖（一個劃分為幾個扇區的圓形圖表，用於描述量、頻率或百分比之間的相對關係）顯示太多數據，而且數據偏向表現差的一方，所以我們應該從選擇數據入手，減少數據的數量和選取有利的數據，例如選擇減肥較成功的個案（10 磅或以上），及將數據數量從 1000 減至 100，這樣圓形圖就能突顯對此公司有利的數據，從而增加營業額。效果如下：

100 名減去不同重量(磅)的顧客分佈



經過以上的改變，相信「纖得好」的業績定能穩步上揚，節節上升。

從商業角度來看，這改變是統計的應用，而從消費者的角度卻是誤用。但有些商人會為求達到目的，便撇開個人的道德良心不談，以謊話來欺騙群眾。從以上的例子可見，數字只能影響別人的判斷，卻不會說謊的。因此，只要我們細心留意不同的圖表，在商人的角度上出發，思考圖表的意義，謹記「逢商必奸」和數字不會說謊的道理，我們便能防止被誤導。

參考資料：

書籍

1. 赫夫、蓋斯繪（2005）。別讓統計數字騙了你（鄭惟厚譯）。天下遠見。
2. 鄭惟厚（2007）。你不能不懂的統計常識。天下文化。
3. 瓊斯（2005）。別讓統計圖表唬弄你（葉偉文譯）。天下遠見。

優異作品：數字不會說謊！？

優異作品

哪所補習社比較好？

學校名稱：寶血會上智英文院

學生姓名：鄭昭彤

級別：中六

顧問老師：曾黎思韻

引言

公開考試將至，學生甲為了在考試獲取最好的成績，決定報讀補習班。可是坊間補習社林林總總，究竟應該選擇哪一所呢？

翻開報紙，A 補習社的廣告佔了半個版面。

智強教育

讓你穩奪高分！

有學生報讀 XX 導師的課程後，
公開考試數學科由 D 變 A

無論甚麼學生，
補 XX 一定有進步！

學生甲眼前一亮，不得了，如今補習班導師竟已到如此境界，可化腐朽為神奇。學生甲雖然成績平平，看到廣告後也燃起非奪 A 不可的決心。此時學生甲的姐姐走過，學生甲不禁興奮地對她說：「只要報讀這導師的補習班，我的成績由 C 躍升到 A 級也不是難事！」

姐姐疑惑地瞥見學生甲手上的報紙，大笑出聲。「要是世界上有一隻豬懂說人話，難道你就相信全世界的豬其實都懂說人話嗎？」看見學生甲一臉不解的樣子，姐姐止住了笑聲。「我是說，要確定這補習社或者 XX 導師的水平，至少也要從一些統計數字著手吧！如果廣告中說的由 D 變 A 的學生在全部學生裡只有一兩個，那麼你

認為你會是其中之一嗎？」

「那……」雖然姐姐說這種進步極大的學生只有一兩個的說法有點不妥，學生甲卻也無法否認，這些「奇蹟」可能只是個別的例子。突然他發覺原來廣告旁邊還有其他宣傳語說不定就是姐姐所說的「統計數字」。

全城第一，可喜可賀！
今年兩大公開考試中，我們的學生勇奪超過
1000 個 A

「你看！奪 A 的人才不只一兩個，我還是有機會的！」學生甲指著宣傳語說道。

姐姐托托眼鏡，笑說：「你這未免太武斷了。第一我們始終不知道究竟有多少那些進步神速的學生。第二，雖說這補習社的奪 A 數目很多，但這種大型補習社的學生更多。假若有一萬個學生，每人報讀兩至三科，考獲 A 級的百分比也只有……」她隨手拿起紙張，寫下算式：

考獲 A 的百分比

$$= \frac{1000}{10000 \times 2} \times 100\%$$

$$= 5\%$$

「一般來說考獲 A 級的考生大概佔全體考生的 2.5% 至 3%，就算報讀了補習班，奪 A 的人也只多了兩個百分點。雖然看上去還是

比全體考生的數字多了，但我們不可以排除報讀補習班的學生，他們原本的成績分佈比例和全體考生不相同的可能性。

事實上，不少成績好的學生為了精益求精而選擇補習，無心向學、成績較差的學生卻不會花金錢和時間在補習上。那麼在補習的學生當中成績較優的學生佔的比例較多，奪 A 率較高不足為奇。在統計學來說，由於補習學生原本的成績分佈和全體考生不同，不算是隨機抽樣，得出來的數據就沒有意義可言，亦不可以和全體考生的數據相比。」

「我怎麼覺得腦袋一團糟……」

「你該不會連百分率也忘了吧？」姐姐開玩笑道。

學生甲吐吐舌頭：「課堂上是有教過，但實際上沒多少機會應用過啦！」

「很多時候只有單一數字是不能顯示整體水平的，通常要和整體計算百分率，才能知道實際上某部分和整體的比例呢！『1000』這數字雖然給人感覺很大，但和補習學生的人數相比卻依然只是少數。而且即使知道了百分率，也得清楚統計數據的來源，資料是否可靠，否則很容易就會被誤導了呢！」

學生甲似乎沒有在聽姐姐說話，很快又在報章上翻到另一所補習社的廣告。

賀

優秀教育的學生在公開考試共獲 200 A，
連續三年全城第一*！

* 以百分率為比較標準

「這麼奇怪？為甚麼兩則廣告都宣稱自己是全城第一呢？」學生甲困惑不已。明明「全城第一」應該只有一個啊！難道其中有一則是騙人的？

姐姐也好奇地湊過頭去看看這則廣告。她眯著眼，逐行逐行讀清楚，終於發現關鍵的地方。「你試試看清楚這句子……」她指著廣告底部一行小小的字，對學生甲說，原來廣告的最後有一句小小的字：

* 以百分率為比較標準

「哦，我明白了！前一則廣告是用奪 A 的數字作比較，而這廣告則是用百分率，結果當然不同囉！既然如此，應該是奪 A 百分率較高的這所補習社比較好。」

姐姐卻皺起眉頭。「你解釋得沒錯，只是也不能這樣斷言這所補習社比較好。廣告上沒有列明奪 A 的百分率，說不定只比剛才那所高零點二、三巴仙呢！」

學生甲有點不耐煩了。「怎麼這些廣告這麼惱人啦？說是騙人卻

不是，但甚麼都好像只說一半，也不讓人看到詳細點的資料——就像用數字來唬人！」

「這就是廣告，多數都只會寫出好的方面來吸引消費者，也不會列舉太詳細的數據，總不能把自己最差的一面顯露出來吧！所以我們只能更細心地閱讀、判斷廣告上的內容，總是抱著半信半疑的態度……

而且要顯示一所補習社的教學質素，並不只限於廣告上時常強調的奪 A 的數字或百分率，要是能使原本成績不佳的學生進步良多、找到學習動機，也可說是成功了……我想與其依靠廣告作比較，還不如親自去問問較為直截了當。你把看上去可以考慮的補習社的電話號碼抄下來，我替你打電話問問詳細點的資料吧！」

過了一會兒，姐姐像個洩了氣的皮球般放下電話。「這是甚麼服務態度！」

學生甲好奇地問：「怎麼了？」

「我只是問了他們學生總體成績的平均數、中位數、分佈域、標準差，報考各科的人數、每科奪 A 的百分率、學生來自的學校的成績組別、學生與老師的比例還有其他基本的資料而已……他們不但支吾以對，而且最後竟然掛線了！」

學生甲好不容易才忍住笑意。「姐，我想我不用出外補習了，你給我補習就行了！」

「你想得美！以為你姐不用讀書麼？」

參考資料：

1. 《機會的數學》，陳希孺著，牛頓出版公司

優異作品：哪所補習社比較好？

邀請作品

主場之利

學校名稱：皇仁舊生會中學

2008 至 2009 年度借調老師：陳德貞

引言

相信有留意英格蘭超級足球聯賽的同工均會發現一般球隊於主場作賽表現會較為出色，得勝機會會較高。這種情況有沒有數據支持？如何解釋？

在日常生活中，我們遇到的事情有三個情況：必然發生、有可能發生、沒可能發生，在數學裏會用一個數值的大小來表示事件發生的可能性，數值愈大表示事件愈有可能發生。在基礎的課程裏，概率的計算都是以數算的方式為主，即把符合事件的數量除以事件的總數，例如：一個箱子裏有 1 個白球、2 個黑球、3 個紅球，那抽出白球的概率為 $\frac{1}{6}$ 。

但許多時候在現實生活裏遇到的問題都不能以這麼簡單的方法去計算，例如：測驗只有合格與不合格兩個情況，那全級第一名與第二百名的同學的合格概率均等於 $\frac{1}{2}$ ？這顯然不是，即使我們不懂太多的概率理論，但也會懂得需按這兩人過去的測驗表現而作出恰當的推測。在這種有條件性下的情況所計算出的概率，稱為條件概率。

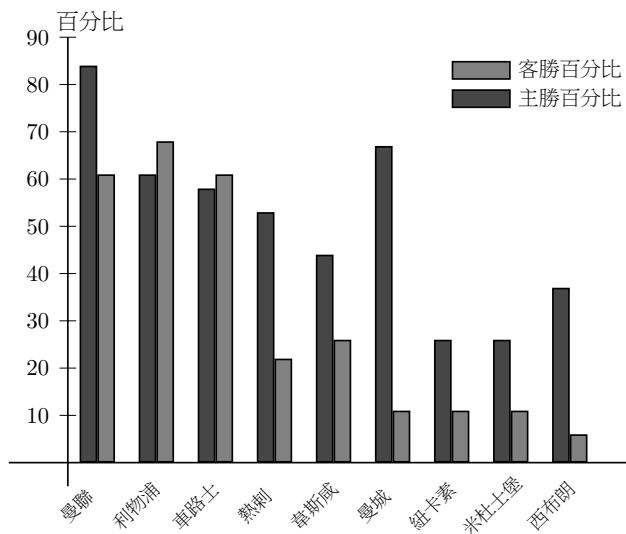
以下是兩位同學在網上的對話：





讓我們來探討「主場之利」這問題：以 2008-09 賽季英格蘭超級聯賽的成績作參考，在總分排名榜二十名中的榜首、榜中、榜末各選取三隊球隊，比較各隊分別在主場和客場作賽的成績。

總分排名	球隊	主場作賽			客場作賽			主勝%	客勝%	相差 百分點
		積分	已賽	主勝	積分	已賽	客勝			
1	曼聯	50	19	16	37	18	11	84	61	23
2	利物浦	40	18	11	43	19	13	61	68	-7
3	車路士	39	19	11	41	18	11	58	61	-3
8	熱刺	35	19	10	16	18	4	53	22	31
9	韋斯咸	26	18	8	22	19	5	44	26	18
10	曼城	36	18	12	11	19	2	67	11	56
18	紐卡素	22	19	5	12	18	2	26	11	15
19	米杜士堡	24	19	5	8	18	2	26	11	15
20	西布朗	24	19	7	7	18	1	37	6	31

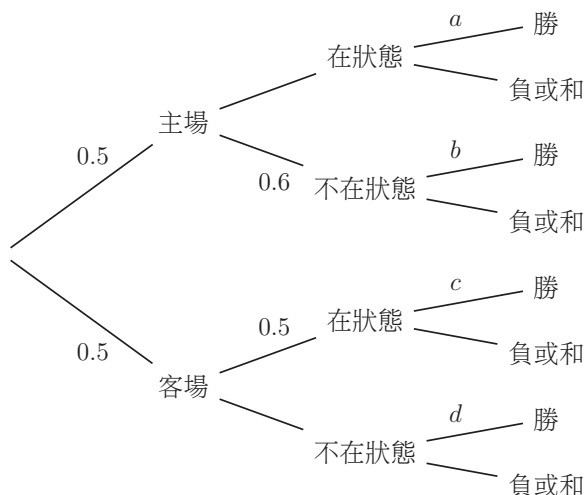


由上表可見，榜中和榜末的球隊在主場和客場作賽時成績有顯著的分別，「主場之利」對實力稍遜的球隊是有幫助，當中以曼城的差距最大；反之，榜首的球隊在主場和客場作賽時成績有沒有顯著的分別，即實力較強的球隊不論在那兒作賽也能發揮水準、保持佳績。

在現實的情況裏，球隊能否勝出要視乎許多因素，例如天氣、球員的狀態、教練的策略等，當中以球員的狀態和教練的策略尤其重要。假設韋斯咸在主場作賽，60% 球員因滿心歡喜地認為自己佔主場之利能有較大的機會勝出，作賽時掉以輕心而不在狀態；假設韋斯咸在客場作賽，50% 球員因而激發鬥心，拚命作賽。根據以往數據，當韋斯咸在主場作賽時，有 44% 的機會勝出，反之在客場作賽時，有 26% 的機會勝出，假設其他情況不變，那麼該隊能否反敗為勝？

- 設
- a 為韋斯咸在主場作賽及球員在狀態時勝出的概率；
 - b 為韋斯咸在主場作賽及球員不在狀態時勝出的概率；
 - c 為韋斯咸在客場作賽及球員在狀態時勝出的概率；
 - d 為韋斯咸在客場作賽及球員在不狀態時勝出的概率。

以樹形圖（下圖）去表達現在的情況：



根據以往數據，當韋斯咸在主場作賽時，有 44% 的機會勝出；即：

$$P(\text{勝} \mid \text{主場}) = 0.44$$

$$\frac{P(\text{勝} \cap \text{主場})}{P(\text{主場})} = 0.44$$

$$P(\text{勝} \cap \text{主場}) = 0.44 \times 0.5$$

$$0.5(0.4)a + 0.5(0.6)b = 0.44 \times 0.5$$

$$4a + 6b = 4.4$$

$$a = \frac{4.4 - 6b}{4} \quad \dots\dots(i)$$

按理，當球隊在主場作賽和球員不在狀態時勝出的概率 (b) 是較小，例如 $b = 0.2$ ，則 $a = 0.8 > 0.5$ ；由此可見，當球隊在自己的主場作

賽，勝出的概率會因球員在狀態下超過一半。另一方面，根據以往數據，當韋斯咸在客場作賽時，有 26% 的機會勝出；

$$P(\text{勝} \mid \text{客場}) = 0.26$$

$$\frac{P(\text{勝} \cap \text{客場})}{P(\text{客場})} = 0.26$$

$$P(\text{勝} \cap \text{客場}) = 0.26 \times 0.5$$

$$0.5(0.5)c + 0.5(0.5)d = 0.26 \times 0.5$$

$$5c + 5d = 2.6$$

$$c = \frac{2.6 - 5d}{5} \quad \dots\dots(ii)$$

按理，當球隊在客場作賽和球員不在狀態時勝出的概率 (d) 是很小，例如若 $d = 0.02$ ，則 $c = 0.50$ ；由此可見，雖然球隊的整體勝出的概率只是兩成多，但當球隊在自己的客場作賽時是在狀態的話，勝出的概率會提升至一半。這樣我們看見無論韋斯咸是在主場或客場作賽，勝出的概率均是少於一半：

$$P(\text{勝} \mid \text{主場}) = 0.44 < 0.5$$

$$P(\text{勝} \mid \text{客場}) = 0.26 < 0.5$$

因此我們不能說「主場之利」是必勝之因，那麼這是否代表韋斯咸勝出的機會永遠是少於「負或和」的機會呢？從以上的計算，我們看見當韋斯咸球員是在狀態下作賽，無論韋斯咸是在主場或客場作

賽，勝出的概率均是大過或等於一半：

$$P(\text{勝} \mid \text{主場}, \text{在狀態}) = 0.8 \geq 0.5$$

$$P(\text{勝} \mid \text{客場}, \text{在狀態}) = 0.5 \geq 0.5$$

而

$$P(\text{勝} \mid \text{主場}, \text{不在狀態}) = 0.2 < 0.5$$

$$P(\text{勝} \mid \text{客場}, \text{不在狀態}) = 0.02 < 0.5$$

這種在分組比較和在合併考慮中得出不同或甚至是相反的結論的現象可稱為辛普森悖論。以這例子來說，考慮了球隊是否在狀態下作賽的因素，球隊在主場作賽仍比在客場作賽有較佳勝算，但是我們不能說「主場之利」是必勝之因，反之，在狀態下作賽也可能反敗為勝。

因此，不論是在瀏覽網路新聞或報章雜誌中的統計數據資料時，我們應注意此現象，特別是當統計數據是以某種分類的方式作比較時，一定要仔細計算合併後的數據，以免遭到誤導。

參考資料：

1. 2008-09 賽季英格蘭超級聯賽主勝積分排名

<http://app.goool.com/competition.do?lid=>

邀請作品：主場之利

邀請作品

哪一科老師較易在最受學生歡迎的老師選舉中獲勝？

學校名稱：保良局馬錦明夫人章馥仙中學

2008 至 2009 年度借調老師：何綺紅

引言

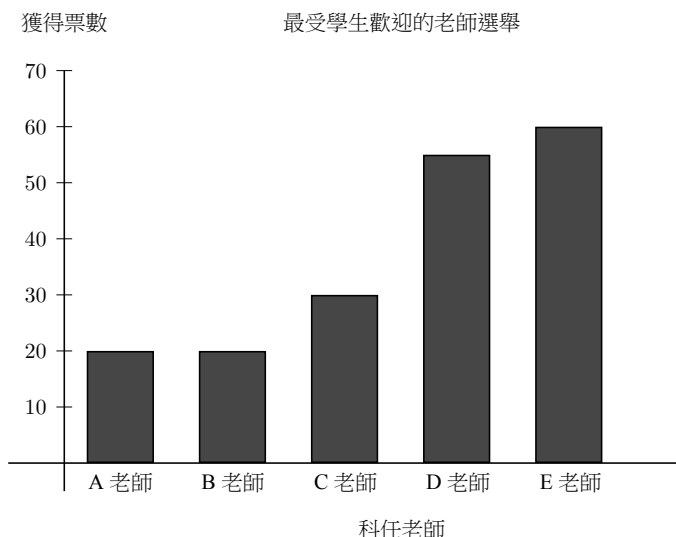
有些學校會於每年舉辦最受學生歡迎的老師選舉，增強師生之間的感情。任教哪些科目的老師會較易獲獎呢？原因何在？

邀請作品：哪一科老師較易在最受學生歡迎的老師選舉中獲勝？

學生會每年都會進行最受學生歡迎的老師選舉，讓學生以一人一票的方式選出最受學生歡迎的老師，自然地獲得最高票數的老師會獲得最受學生歡迎的美譽。

去年的最受學生歡迎的老師選舉的部分投票情況如下：

老師	A 老師	B 老師	C 老師	D 老師	E 老師
教授科目	中文科	英文科	數學科	音樂科	體育科
獲得票數	20	20	30	55	60



從以上的圖表中，可見 E 老師獲得最多票數，所以 E 老師獲得最受學生歡迎老師的美譽。

根據過往數年的選舉經驗，我發現獲得最受學生歡迎的老師多為術科的科任老師，如體育科、音樂科的科任老師。

為何術科的科任老師較一般科目的科任老師會較多被學生選為最受歡迎老師？是否因為一般科目的科任老師由於要面對課程內的家課及考試要求，所以會對學生較嚴格，經常催逼學生交功課甚至要罰留堂溫書等，導致不受學生歡迎；另一方面，術科的課程不用面對太大的考試壓力，上課的氣氛也較一般科目輕鬆，學生便能愉快學習；自然地，較少學生會將神聖的一票投給一般科目的科任老師，而大部分學生會投票給術科的科任老師。

除了上述的原因外，還有沒有其他的科學方法能夠解釋為何最受學生歡迎的老師選舉中較多術科的科任老師獲獎？

初中數學科課程的數據處理範疇中有認識期望值的意義，我們可從期望值的意義解釋為何最受學生歡迎的老師選舉中較多術科的科任老師獲獎。

根據維基百科，期望值的定義為：「在機率論和統計學中，一個離散性隨機變量的期望值（或數學期望、或均值，亦簡稱期望）是試驗中每次可能結果的機率乘以其結果的總和。換句話說，期望值是隨機試驗在同樣的機會下重複多次的結果計算出的等同「期望」的平均值。需要注意的是，期望值並不一定等同於常識中的「期望」——「期望值」也許與每一個結果都不相等。」

舉例說明，假設學生只認識教授他們的科任老師，自然地學生

邀請作品：哪一科老師較易在最受學生歡迎的老師選舉中獲勝？

只會投票給教授他們的科任老師，並且每位老師獲學生投票為最受歡迎老師的概率均等，如初中的學生每學年會由約 12 位老師教授不同的科目，即某位老師獲初中學生投票的概率為 $\frac{1}{12}$ ；如高中的學生每年會由約 8 位老師教授不同的科目，即某位老師獲高中學生投票的概率為 $\frac{1}{8}$ 。

假設某校由初中至新高中，即中一至中六級，每級有五班，每班 40 人。根據敝校的教學安排，由一位音樂科老師教授中一至中三級的音樂科，兩位體育科老師教授全校的體育科；一位任教中文科或英文科的老師通常教三班，一位任教數學科的老師通常教四班。各科的科任老師獲得最受學生歡迎老師票數的期望值如下：

科任老師	教授班別	最受學生歡迎老師票數的期望值
音樂科	初中共 15 班	$\frac{1}{12} \times 15 \times 40 = 50$
體育科	初中及高中各 15 班，每班 20 人	$\frac{1}{12} \times 15 \times 20 + \frac{1}{8} \times 15 \times 20 = 62\frac{1}{2}$
中文科／英文科	2 班高中及 1 班初中	$\frac{1}{12} \times 1 \times 40 + \frac{1}{8} \times 2 \times 40 = 13\frac{1}{3}$
數學科	2 班高中及 2 班初中	$\frac{1}{12} \times 2 \times 40 + \frac{1}{8} \times 2 \times 40 = 16\frac{2}{3}$

從上述的計算結果，可以發現體育科與音樂科的老師必定是最受學生歡迎老師選舉中的大熱，因為體育科老師獲得票數的期望值是中文科/英文科老師獲得票數的期望值的四至五倍，由此可見，一

般科目的科任老師除非有化腐朽為神奇的魅力，否則都難以獲得最受學生歡迎老師的美譽。

下表計算獲得票數多於期望值的百分比：

科任老師	期望值	獲得票數	獲得票數多於期望值的百分比
音樂科	50	55	$\frac{55-50}{50} \times 100\% = 10\%$
體育科	62.5	60	$\frac{60-62.5}{62.5} \times 100\% = -4\%$
中文科／ 英文科	$13\frac{1}{3}$	20	$\frac{20-13\frac{1}{3}}{13\frac{1}{3}} \times 100\% = 50\%$
數學科	$16\frac{2}{3}$	30	$\frac{30-16\frac{2}{3}}{16\frac{2}{3}} \times 100\% = 80\%$

若選舉制度是根據票數最高者而獲勝，這樣體育科老師獲得 60 票便會獲勝；而數學科老師獲得的票數不是最多，當然不會獲得最受學生歡迎老師的美譽。另一方面，若計算獲得票數多於期望值的百分比時，體育科老師是 -4% ，而數學科老師卻有 80% ，這百分比卻是最大的，從這數值可以發現體育科老師獲得的票數少於期望值，相反數學科老師獲得的票數比期望值高出 80% 。

雖然最受學生歡迎老師選舉是由學生以一人一票的方式選出，但這選舉制度是否對於每一位老師都公平呢？我們可以怎樣修改選舉制度，令最受學生歡迎的老師選舉得以延續下去，讓辛勤的老師能獲得一些美譽嘉許呢？

我建議改善選舉制度的方法：當老師獲得的票數多於期望值的票數時便可獲得受學生歡迎的老師的美譽，讓更多老師可以獲得嘉

邀請作品：哪一科老師較易在最受學生歡迎的老師選舉中獲勝？

許，當然，這個選舉制度可能令很多老師同時獲獎；另一個方法是根據獲得票數多於期望值的百分比來判斷，百分比最高便是最受學生歡迎的老師。當然，這選舉制度亦會有一些不公平的情況，例如，副校長通常只教授一班高年級的班別，這樣，副校長獲得最受學生歡迎老師票數的期望值 $= \frac{1}{8} \times 40 = 5$ ，當計算獲得票數多於期望值的百分比時，由於期望值較小，獲得票數不用太多，已經令百分比很大，例如，獲得票數是 9 票，這樣，獲得票數多於期望值的百分比 $\frac{9-5}{5} \times 100\% = 80\%$ ，足以成為最受學生歡迎老師的得主。

若你是學生會的幹事，你會怎樣處理最受學生歡迎的老師選舉呢？

參考：

1. 維基百科網站

<http://zh.wikipedia.org/w/index.php?title=%E6%9C%9F%E6%9C%9B%E5%80%BC&variant=zh-hk>

邀請作品

Game of Chance

學校名稱：佛教善德英文中學

2008 至 2009 年度借調老師：程國基

Introduction

Have you ever played a game called “Da Feng Chui” (the big wind blows 大風吹) during school Christmas party? This game is very exciting and everybody will enjoy it. Is there any strategy to win the game?

How to play the game called “Da Feng Chui” (the big wind blows 大風吹)? The rule is very simple. At the beginning, all participants sit in a circle except one person comes out and stands at the middle of the circle. Exciting moment comes when that person says something that someone have. Those participants need to leave their own seats immediately and find another seat as quickly as possible. At the same time, that person can find a seat to sit. Clearly, there will be somebody who cannot find his/her own seat; he/she will then receive punishments. Since the game requires participants with quick responses and fast movements, there will be many interesting things happening and they may feel a lot of fun. Apart from the amusement of the game, some mathematics can be found from this activity.

Let's think about a simple example: There are only Mary(M), Susan(S), John(J), Dick(D) and Peter(P) joining the game. John is selected to come out first. The following figure (a) is used to show the situation:

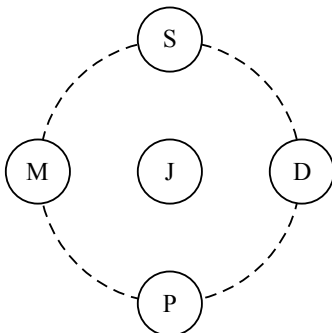


Figure (a)

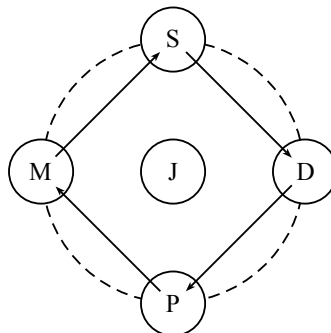


Figure (b)

If John says "Those wear shoes, they must leave", all participants will

leave their own seats. In order to make John find no seats, it may be: Mary $\Rightarrow$ Susan; Susan $\Rightarrow$ Dick; Dick $\Rightarrow$ Peter; Peter $\Rightarrow$ Mary, where $\Rightarrow$ means moves to the seat of somebody as shown in figure (b).

Do you know totally how many ways are there for Mary, Susan, Dick and Peter to find the new seats in order to prevent John to get a seat?

The following is one of the ways presented in tabular form:

	Mary	Susan	Dick	Peter
Mary	✗	✓		
Susan		✗	✓	
Dick			✗	✓
Peter	✓			✗

- ✗ – means that each player cannot take his/her original seat again.
- ✓ – means that the player on left hand side takes the seat of another player.

For example, refer to the table, Mary moves to the seat of Susan.

With the aid of the above table, you only need to consider the following cases.

Case 1: If Mary has chosen Susan's seat, Susan will have 3 choices to move.

If Mary and Susan have chosen their new seats, Dick will have only 1 choice left. If Mary, Susan and Dick have chosen their new seats, it is clear that only 1 way is provided for Peter to move.

Case 2: If Mary has chosen Dick's seat or Peter's seat and Susan has chosen Mary's seat, Dick and Peter both will have only 1 choice to select.

Case 3: If Mary has chosen Dick's seat and Susan has chosen Peter's seat, Dick will have 2 choices to move. If Dick has chosen his new seat, Peter will have only 1 choice left.

Case 4: If Mary has chosen Peter's seat and Susan has chosen Dick's seat, Dick will have 2 choices to move. If Dick has chosen his new seat, Peter will have only 1 choice left.

$$\begin{aligned} & \text{Hence, the total number of ways for these players to select other seats} \\ & = 1 \times 3 \times 1 \times 1 + 2 \times 1 \times 1 \times 1 + 1 \times 1 \times 2 \times 1 + 1 \times 1 \times 2 \times 1 \\ & = 9 \end{aligned}$$

Although the above calculation is not so complicated, it will be more difficult for us to count and we need to consider more cases if there are more players joining this game. That is why we are going to use “probability” in order to solve this problem in a more systematic way. You may have learnt that for $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ any events A, B . How about $P(A \cup B \cup C \cup D)$ for any events A, B, C, D ? Do we have a similar result? Yes and it can be obtained by the “inclusion – exclusion principle” for probability:

$$\begin{aligned}
& P(A \cup B \cup C \cup D) \\
&= P(A) + P(B) + P(C) + P(D) - P(A \cap B) - P(A \cap C) \\
&\quad - P(A \cap D) - P(B \cap C) - P(B \cap D) - P(C \cap D) \\
&\quad + P(A \cap B \cap C) + P(A \cap B \cap D) + P(A \cap C \cap D) \\
&\quad + P(B \cap C \cap D) - P(A \cap B \cap C \cap D)
\end{aligned}$$

Let A = the event that Mary takes her original seat.

B = the event that Susan takes her original seat.

C = the event that Dick takes his original seat.

D = the event that Peter takes his original seat.

From the above formula and with the fact that $P(A) = P(B) = P(C) = P(D)$, $P(A \cap B) = P(A \cap C) = P(A \cap D) = P(B \cap C) = P(B \cap D) = P(C \cap D)$ and $P(A \cap B \cap C) = P(A \cap B \cap D) = P(A \cap C \cap D) = P(B \cap C \cap D)$, we have,

$$\begin{aligned}
& P(A \cup B \cup C \cup D) \\
&= \frac{1 \times 3!}{4!} \times 4 - \frac{1 \times 1 \times 2!}{4!} \times 6 + \frac{1 \times 1 \times 1 \times 1 \times 1}{4!} \times 4 \\
&\quad - \frac{1 \times 1 \times 1 \times 1 \times 1}{4!} \\
&= 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!}
\end{aligned}$$

Hence, P(none of 4 players Mary, Susan, Dick and Peter take their original seats)

$$\begin{aligned}
&= 1 - P(A \cup B \cup C \cup D) \\
&= 1 - \left(1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!}\right) \\
&= \frac{3}{8}
\end{aligned}$$

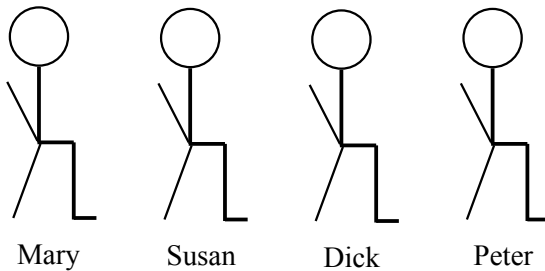
Since the number of ways that these participants select their seats without restrictions = $4!$.

Hence the number of ways that none of the participants take their original seats = $4! \times \frac{3}{8} = 9$.

Up to now, the problem seems to be solved and you may feel everything OK. However, I can tell you that Mathematics does not stop here. You may learn more if you investigate further...

Let's look some interesting results below.

From the above, for four players



$$P(\text{none of 4 players take their original seats}) = 1 - \left(1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!}\right).$$

With some calculations for only 2 players or 3 players, we have the following expressions:

$$P(\text{none of 2 players take their original seats}) = 1 - \left(1 - \frac{1}{2!}\right)$$

$$P(\text{none of 3 players take their original seats}) = 1 - \left(1 - \frac{1}{2!} + \frac{1}{3!}\right)$$

Observe all these expressions. Do you agree that they all seem to follow a pattern?

Yes, that is $P(\text{none of } n \text{ players take their original seats})$

$$= 1 - \left(1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \dots + \frac{(-1)^{n-1}}{n!} \right)$$

$$= 1 - \left(\frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \dots + \frac{(-1)^n}{n!} \right) \dots\dots\dots (*)$$

where n is any positive integer.

Although the proof of (*) is omitted here, the proof will be similar as in the case when $n = 4$. Try by yourself if you like!

We can now generalize a formula for the number of ways that none of n players move to their original seats.

By using original definition of probability,

$$\begin{aligned} & P(\text{none of } n \text{ players take their original seats}) \\ = & \frac{\text{Number of ways that none of } n \text{ players move to their original seats}}{\text{Number of ways that } n \text{ players find their seats without any restrictions}} \\ & \text{Combined it with (*), } 1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \dots + \frac{(-1)^n}{n!} \\ = & \frac{\text{Number of ways that none of } n \text{ players move to their original seats}}{n} \end{aligned}$$

Hence, the number of ways that none of n players move to their original seats

$$= n! \left(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \dots + \frac{(-1)^n}{n!} \right) \dots\dots\dots (**)$$

This value of (**) has a name called **subfactorial** and denoted by $!n$ but you don't confuse with n factorial $n!$.

From (**), when n starts from 0, we have $!0 = 1$, $!1 = 0$, $!2 = 1$, $!3 = 2$, $!4 = 9, \dots$ You may find the value increasing rapidly.

For example, $!10 = 1334961$.

The above "Seat" problem has its history. It is originated from a classical old problem first proposed by the French mathematician **Pierre Remond de Montmort** in 1708. This old problem was recorded in his treatise on the analysis of games of chance. In Montomort's version of the problem, it was about a game of playing with a deck of 13 cards. So the game was called **Treize**, French of thirteen. Another name of the game is called **Rencontres**(means coincidences) or **Montmort's Matching Problem**:

Suppose you have a deck of 13 cards numbered 1, 2, 3,..., 13 respectively. Cards are drawn one by one, what is the probability that no card is drawn in the order indicated by its label? This means that first card is drawn with the label not "1"; second card is drawn with the label not "2",...,etc.

This problem was not so easily to be solved at that time but if you consider "card" just like "seat" and apply (*), you can obtain:

$P(\text{no card is drawn with its order equal to its label})$

$$= 1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \dots + \frac{(-1)^{13}}{13!}$$

$$= 0.3678794 \text{ (correct to 7 decimal places)}$$

The following table shows the value of (*) and (**) for $n = 0$ to 13:

n	$\left(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \dots + \frac{(-1)^n}{n!}\right)$	$n! \left(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \dots + \frac{(-1)^n}{n!}\right)$
0	1	1
1	0	0
2	0.5	1
3	0.3333333	2
4	0.375	9
5	0.3666667	44
6	0.3680556	265
7	0.3678571	1854
8	0.3678819	14833
9	0.3678792	133496
10	0.3678795	1334961
11	0.3678794	14684570
12	0.3678794	176214841
13	0.3678794	2290792932

(The above values are exact or correct to 7 decimal places.)

Would you find out some values in the table that seems to be equal if n becomes larger?

Although the overall story is about probability, the result we obtain can be connected with a special number called Euler number, “ e ” .

When $n \rightarrow \infty$,

$$\left(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \dots + \frac{(-1)^n}{n!}\right) \rightarrow \left(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \dots\right)$$

$$\text{Since } \left(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \dots\right) = 1 + \frac{(-1)^1}{1!} + \frac{(-1)^2}{2!} + \frac{(-1)^3}{3!} + \frac{(-1)^4}{4!} +$$

$$\dots = e^{-1} \text{ for } e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

Hence $\left(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \dots + \frac{(-1)^n}{n!}\right) \rightarrow e^{-1} \approx 0.36787944\dots$
as $n \rightarrow \infty$.

That is why the values of (*) become stable and equal to 0.3678794 correct to 7 decimal places.

You may use this result to win money: Prepare a deck of n cards numbered 1, 2, 3... n . Shuffle them thoroughly. Turn up the cards one by one and you shout 1, 2, 3, ..., n , 1, 2, 3, If the number you shout matches the number on the card, the game ends and you will get \$1 from your friend, otherwise the game continues until the number of the last card is shown. If this number still does not match what you shout, then you need to give \$1 to your friend. If you repeat playing this game for enough number of times, your friend will lose. About this game, is it like a popular children game called “襟綿胎” ?

Do you agree that some games are not just for fun but you may learn lot of mathematics? From this article, the games introduced above are called **Games of Chance**. You may think that the mathematics used for games of chance is probability or statistics. However, skills of algebra and calculus are also widely used. Hence, when you play games with your friends next time, remind yourself to try some mathematics and this may increase your chance of success. Good Luck!

Remark:

For method of inclusion – exclusion principle, you may refer the following websites

1. http://en.wikipedia.org/wiki/Inclusion-exclusion_principle
2. <http://en.wikipedia.org/wiki/Derangement>

邀請作品：GAME OF CHANCE

邀請作品

抽樣調查的陷阱

看看圖一及圖二兩位軍官的對話，你同意他們的分析嗎？為什麼？



假設在受傷的 100 名軍人中，有 80 名是手部受傷，其餘 20 名則是頭部受傷，即手部受傷的比例為 80%。如此看來，加強手部保護是必須的。但這分析並沒有考慮那些陣亡軍人的致命原因。



若有 200 名軍人陣亡，當中有 90% 是頭部受傷的。那麼，頭部受傷軍人佔整體為 67%，這遠較佔生還軍人的 20% 為高。因此，頭部裝備更加需要加強。

	手部受傷	頭部受傷	總數
生還軍人	80	20	100
陣亡軍人	20	180	200
總數	100	200	300

由此可見，如果沒被樣本涵蓋的範圍 (non-coverage) 佔大部分，而它們的特質與被抽樣涵蓋的部分截然不同，所編製統計數字便沒

有代表性。

統計數字的編製，往往依靠抽取樣本，假設樣本能夠確實反映沒被抽中的部分，然後再依一定比例放大到全體，所以抽樣涵蓋範圍的代表性對於所編製數據的質量尤其重要。一旦選錯抽樣涵蓋範圍，遺留了重要的部分，而沒涵蓋部分的特質異於抽樣涵蓋的範圍的話，所抽取的資料便會有偏誤，不能代表整體，依比例放大計算出來的數字亦不能代表到全體。這樣的話，即使再多的資料、再精確的計算都不會改善數字品質，只是白費功夫。

正如上述指出，抽樣涵蓋範圍的代表性固然重要，但抽樣隨機性 (randomness) 在統計學上也佔有一個重要的角色，隨機的抽樣亦避免我們搜集了嚴重偏誤的資料來進行分析。

我們延續以上例子。假設軍官們選取了正確的抽樣涵蓋範圍，包括了生還和陣亡的軍人，但我們怎樣確保抽樣方法的隨機性呢？理論上，每一位出戰的軍人都有一個已知並且不等於零的被抽中的機會，才算是一個正確的概率抽樣 (probability sampling)。但概率抽樣要在現實中執行並不容易。如果我們在夜間到營舍進行抽樣調查，需要在晚上外出站崗的軍人的資料便未必有機會被抽取。如果我們在醫院進行抽樣調查的話，接受治療的軍人和在醫院的陣亡軍人都會被涵蓋在內，但一些自行治理的軍人的傷患個案便沒有被調查的機會。要確保抽樣涵蓋範圍內的樣本都有機會被抽中，其中一個辦法是利用分層抽樣 (stratified sampling)，將抽樣涵蓋範圍分為日

間或晚上駐守、接受治療和自行治理等，在不同的分層上利用上述可能的隨機抽樣方法搜集數據，並根據特定的機率綜合每一個分層的資料再製成數字，才能確保數字的隨機性。

由此可見，要製作一個好的統計調查要考慮很多因素，選取正確的抽樣範圍以及抽樣的隨機性是不可或缺的環節。因此我們在進行抽樣調查和理解調查結果時，都需要十分小心。

參考資料：

1. 劉順仁（2008）。決勝—在看不見的地方。時報文化。

邀請作品

奧運與概率

在日常生活中，我們常常會遇到很多「有可能」會發生的事情，例如明天會下雨或是會放晴？今晚英超聯賽中曼聯獲勝還是阿仙奴獲勝？以數學的角度來看，用來表示一個事件發生的可能性大小的數字叫該事件的概率。在相同的條件下，可能發生也可能不發生的事件叫隨機事件。概率的研究目的就是研究隨機現象所服從的規律，從而令人們能夠預測事件的可能性。

北京二零零八年奧運會比賽共有二十八項比賽項目。其中，羽毛球、排球及乒乓球項目都是大家熟悉的比賽項目。根據第 29 屆奧林匹克運動會網站的資料顯示，羽毛球項目採用 3 局 2 勝賽制，排球項目採用 5 局 3 勝賽制，而乒乓球項目則採用 7 局 4 勝賽制。究竟為什麼會有多局賽制的出現？哪一種多局賽制最能夠準確分辨出技術較優的一方呢？

其實，在各項形形色色的比賽中，我們都有所謂的「大熱門」。「大熱門」的意思只是表示某參賽者在某一場賽事中有較大的機會獲勝。假如我們都認為甲是「大熱門」，甲比乙有較大的機會

獲勝，甲勝出的概率有 80%。這數字表示在 100 場比賽中，甲勝出場數的期望值是 80 場，而乙則只有 20 場。那麼，在一局定勝負的賽制中，甲和乙都有機會勝出。然而，雖然我們都認為甲比乙強，但是他敗給乙的概率仍然有 20%，甲勝出並不是一個「必然的」事件。

在北京奧運會比賽場上，參賽選手的技術良莠不齊。假設奧運會比賽場上只有兩位選手，而甲比乙強，甲勝出的概率有 80%。在一場以 3 局 2 勝賽制的賽事中，甲和乙勝出的局數組合分別羅列如下：

甲勝出的組合	概率	乙勝出的組合	概率
甲甲	0.8×0.8	乙乙	0.2×0.2
甲乙甲	$0.8 \times 0.2 \times 0.8$	甲乙乙	$0.8 \times 0.2 \times 0.2$
乙甲甲	$0.2 \times 0.8 \times 0.8$	乙甲乙	$0.2 \times 0.8 \times 0.2$

從上述的結果計算，甲「大熱勝出」的概率為 0.896，而乙「爆冷門」的概率則是 0.104。因此，3 局 2 勝賽制比一局定勝負能夠準確分辨出技術較優的一方。

根據以上的運算方式，在 5 局 3 勝賽制和 7 局 4 勝賽制中，甲「大熱勝出」的概率分別為 0.942 和 0.967。而乙「爆冷門」的概率則分別是 0.058 和 0.033。由此可見，在參賽者的實力不同的情況下，賽制的局數越多越能夠準確分辨出技術較優的一方。因此，為了能夠準確分辨出技術較優的一方，大部分比賽都採用多局賽制。

那麼，在參賽者的實力旗鼓相當的時候，是否一樣局數越多越好呢？假設奧運會比賽場上仍然只有兩位選手，而甲和乙的實力不相伯仲，甲和乙勝出的概率一樣是 50%。計算後得出以下的結果：

項目	賽制	甲勝出的概率	
		旗鼓相當	強弱懸殊
---	一局定勝負	0.5	0.8
羽毛球	3 局 2 勝	0.5	0.896
排球	5 局 3 勝	0.5	0.942
乒乓球	7 局 4 勝	0.5	0.967

在一局定勝負的賽制中，甲和乙勝出的概率是 0.5。而在 3 局 2 勝、5 局 3 勝和 7 局 4 勝賽制中，甲和乙勝出的概率仍然是 0.5。由此推論，當參賽者實力旗鼓相當的時候，在任何局數的賽制下，任何一方勝出的概率都是一樣。

雖然概率在賽制的決定中確實扮演著一個重要的角色，但是除了概率之外，有沒有其他的因素影響賽制的決定呢？答案是「有可能」。影響賽制決定的因素有很多，例如比賽時間、體能的限制等。舉例來說，北京奧運會的羽毛球項目競賽規則採用了羽毛球世界聯合會於 2006 年 5 月正式決定實行的三局二十一分新賽制。當中提及的提高運動員的積極性、減少運動員受傷以及電視轉播等也是決定這個新賽制的因素。另一方面，乒乓球競賽的賽制同樣是經過多次

的會談後才決定的。由此可見，影響賽制決定的因素確實有很多。所以，賽制的設立往往是經過多次的討論及研究後的結果。而當中有沒有涉及其他的統計學理論呢？這就要留待大家去探討了。

參考資料：

1. 蕭文強、林建（2007）。概率萬花筒。香港統計學會。
2. 第 29 屆奧運會官網。 <http://www.beijing2008.cn/index.shtml>
3. 國際乒聯官網。 <http://www.ittf.com.cn>