13.1 解讀新聞標題

工作紙

題目

STEP I 1992，第 14 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

step.maths.org/questions/92-s1-q14

提示

請先自行嘗試解題。如果遇到困難，請逐一閱讀提示，每讀完一個便回到題目再試。標有停一停，試一試的提示是關鍵步驟。每個提示之後均留有空位供你演算。

> **想一想**　假如死亡事件以穩定的比率獨立而隨機地發生，每年的總數應有多大的變動？若平均每年為 1080，每年與平均值的典型偏差是多少？

階段 1　「坐飛機比步行安全六倍」

涵蓋：第 (i) 部分

> **想一想**　比值 $1080/180 = 6$ 比較的是總死亡人數。這與比較風險是同一回事嗎？一個公平的風險量度需要考慮哪些因素？

階段 2 「我們的天空更安全：1990 年僅 125 人喪生」

涵蓋：第 (ii) 部分

想一想 空難死亡是獨立（independent）的罕見事件，還是會成批出現？你預期每年的總數會有多大的波動？單憑一年的數字能看出趨勢嗎？

階段 3 「道路慘劇有增無減：星期二 7 人喪生」

涵蓋：第 (iii) 部分

停一停，試一試 平均每天有多少名行人喪生？若每天的死亡人數服從以此為平均值（mean）的泊松分佈（Poisson distribution），你預期一年之中有多少天會有 7 人或以上喪生？單憑一天的數字，又能否說明某事正在增加？

階段 4　**「升幅驚人：1990 年 1350 名行人喪生」**

涵蓋：第 (iv) 部分

> ? **想一想**　在平均值為 1080 的泊松模型下，1350 比平均值高出多少個標準差（standard deviation）？若這並非偶然，還有甚麼其他原因可以解釋？想一想 16 年的平均數可能掩蓋了甚麼。

背後的數學

訊號與雜訊

每一個統計問題，歸根究底都是要把*訊號*（signal，即真實的效應或趨勢）與*雜訊*（noise，即隨機波動）分開。泊松模型為獨立事件計數中的雜訊提供了一把尺：標準差為 $\sqrt{\mu}$。罕見事件的*相對波動*很大：計數為 9 時，波動約為 ± 3，即其大小的三分之一；而計數為 10 000 時，波動只有約 1%。正因如此，小數目（某一天的死亡人數、某一年的墜機次數）才會屢屢被人過度解讀，而小鎮上的癌症「羣組」最終也多半只是偶然。

均值回歸

還有另一個陷阱。道路安全攝影機往往安裝在剛剛經歷了異常惡劣一年的地點。翌年，意外減少，功勞就歸於攝影機。但異常惡劣的一年過後，本來就很可能迎來較為平常的一年：這就是*均值回歸*（regression to the mean），最早由高爾頓（Francis Galton）在 1880 年代提出。要把攝影機的效果與均值回歸分開，就必須與沒有安裝攝影機的類似地點作比較，這正是對照實驗（controlled experiment）背後的理念。

南丁格爾的統計學

南丁格爾（Florence Nightingale）以克里米亞戰爭中的護士身分聞名於世，同時也是統計學的先驅。她在 1858 年繪製的「玫瑰圖」顯示，死於可預防疾病的士兵遠多於死於

創傷的士兵，並說服了政府改革軍隊的衛生狀況。她堅持使用比率（每年每千名士兵的死亡人數）而非原始總數，並堅持同類才可相比，這些正是本題所考核的要點。她是首位獲選為皇家統計學會（Royal Statistical Society）會士的女性。

延伸練習

1 **泊松分佈的尾部。**設 X 服從平均值為 2.96 的泊松分佈，驗證 $\mathrm{P}(X \geq 7) \approx 0.031$。你預期每年有多少天會有 9 名或以上的行人死亡？

2 **成羣出現的死亡。**假設墜機事故按泊松過程（Poisson process）發生，比率為每年 1.2 宗，而每宗奪去 150 人的性命。求每年空難死亡人數的平均值和標準差。一年之內有 125 人或以下死亡，是否不尋常？有沒有可能某一年恰好有 125 人死亡？

3 **顯著的升幅？**假如 1974–1989 年的數字一直平穩地維持在 1080 ± 33，試（利用正態近似（normal approximation））求單一年份出現 1350 或以上的概率。

4 **擬一個更好的標題。**為四個標題逐一改寫，各擬一個數據真正支持的標題。

延伸閱讀

- **David Spiegelhalter, *The Art of Statistics***（Pelican）。講述如何從數據中學習，並附有大量取自新聞的例子。
- **Tim Harford, *How to Make the World Add Up***（Bridge Street）。思考傳媒中的數字的十條法則。
- **Darrell Huff, *How to Lie with Statistics***（Penguin）。1954 年出版的短篇經典，至今仍不過時。

完整的引導式解答見第 13.1 節：
https://beyondstep.keithmcnulty.org/zh-hant/ch13/headlines.html

每星期一道 STEP 題目，完整講解：The STEP Up，以電郵免費發送，訂閱網址：keithmcnulty.substack.com