詳細解答

STEP I 1992，第 14 題

概率與統計 • 統計論斷的詮釋

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

本題在 *Beyond STEP* 第 13.1 節解讀新聞標題中有完整的引導式講解：https://beyondstep.keithmcnulty.org/zh-hant/ch13/headlines.html

關鍵思路

每則標題都是拿着一個數字，就一躍跳到結論。對每一則標題，都要問兩個問題。第一，它是否在拿同類的東西作比較：同一樣事物，用同一種方法量度，並按承受風險的程度（exposure）作出調整？第二，這個差異會不會純屬偶然？對於第二個問題，一把很自然的尺是泊松模型（Poisson model），它描述罕有而互相獨立的事件的發生次數，其標準差（standard deviation）等於平均值的平方根。懂得這個模型何時適用、何時不適用，正是 (ii) 至 (iv) 的核心。

解答

以下所說的「平均」，一律指 1974–1989 年的年平均數：每年 1080 名行人死亡，180 人死於空難。

第 (i) 部分

「六倍」之說來自比率 $1080/180 = 6$，但這個比率比較的是原始總數，而不是風險。風險是指每單位風險承擔量的死亡人數：按每程、每小時或每公里計算。在 Poldavia，幾乎人人每天都在馬路附近步行，而乘飛機的人相對很少，而且只是偶一為之。要比較兩者

的安全程度，我們需要的是類似

$$\frac{\text{行人死亡人數}}{\text{步行的總時數（或總公里數）}} \quad \text{對比} \quad \frac{\text{空難死亡人數}}{\text{飛行的總時數（或總公里數）}},$$

這樣的比較，而標題卻沒有給出任何一個分母。視乎所用的量度方法，答案甚至可能倒轉過來：按每公里計，飛行看來非常安全；但按每小時或每程計，就未必那麼安全了。

此外還有其他問題。這裏所謂的「步行」，其實是指徒步時被車輛撞倒，這個數字完全沒有反映行人在僻靜地方的情況。空難死亡人數除了乘客，還包括機組人員。而且一段時期的平均數根本無法「證明」1991 年的安全情況；它們充其量只是關於 1974–1989 年的證據。

第 (ii) 部分

乍看之下，125 遠低於平均數 180。這個跌幅是否超出了偶然的範圍？

假如死亡事件是互相獨立、逐宗發生的，那麼採用平均值為 180 的泊松模型便算合理。它的標準差是 $\sqrt{180} \approx 13.4$，因此 125 大約比平均值低 4.1 個標準差，純粹出於偶然的可能性極低（$P(X \leq 125) \approx 10^{-5}$）。

但空難死亡並不是逐宗發生的，而是成羣出現：一次墜機就可以奪去數十以至數百人的性命。較自然的模型是：致命墜機的次數服從泊松分佈，而每次墜機的死亡人數是隨機的。這樣一來，變異程度便大大增加。舉例說，假設平均每年有 6 宗致命墜機，每宗有 30 人死亡。那麼全年死亡人數是 $30N$，其中 $N \sim \text{Po}(6)$，其標準差為 $30\sqrt{6} \approx 73$，而

$$P(\text{死亡人數} \leq 125) = P(N \leq 4) \approx 0.29.$$

按這些（僅作說明用的）數字，一年有 125 人死亡實在平常不過：只需要比平常少一兩宗墜機便會出現。如果連每宗墜機的死亡人數也容許變動，離散程度還會更大。

所以單憑一年的數字，我們所知甚少。我們還想知道 1990 年的飛行量有多少（如果航空交通減少了，死亡人數本來就應該較少），以及好幾年的數字，才能看出是否真的有趨勢。標題的說法並無根據。

第 (iii) 部分

這裏有兩個不同的毛病。

單憑一天不能顯示上升。「上升」是關於一段時間內趨勢的說法，而單獨一天的數字根本無從比較。我們需要一段長時間內每天（或每年）的數字。

七人根本不算異常。我們手上唯一的道路數字是行人的：每年 1080 人，即

$$\lambda = \frac{1080}{365} \approx 2.96 \text{（每天）}.$$

用泊松分佈來模擬行人死亡相當合適：這類事件罕有，大多是每宗意外一人死亡，而且大致互相獨立。設 $X \sim \mathrm{Po}(2.96)$ 為某一天死亡的行人數目。那麼

$$P(X \geq 7) = 1 - \sum_{k=0}^{6} \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!} \approx 0.031,$$

所以我們預期一年之中大約有 $365 \times 0.031 \approx 11$ 天，會有最少 7 名行人死亡。況且標題說的是「死亡人數」，那還包括司機、乘客和騎單車的人。每天道路死亡總人數的平均率必定高於 2.96，於是 7 就更加不足為奇了。

想一想 某一天的死亡人數是平均數的兩倍有多，難道不令人震驚嗎？

數目小的時候並不然。當平均值約為 3 時，泊松標準差是 $\sqrt{3} \approx 1.7$，與平均值相比已經很大，所以每天之間按比例計出現大幅波動是正常的。專挑最差的一天，也是一種選擇效應（selection effect）：一年有 365 天，報章總能找到差的一天。那天是星期幾也有關係，因為交通流量和外出步行的人數會隨星期幾和季節而變化。速率恆定的泊松模型只是一個粗略的初步近似。

第 (iv) 部分

這一則標題確實有值得認真回應之處。先看統計，再談詮釋。

*1350 是否超出了偶然的範圍？*基於 (iii) 所述的理由，泊松模型在這裏是個不錯的起點。平均值為 1080 時，標準差是 $\sqrt{1080} \approx 32.9$，而

$$\frac{1350 - 1080}{32.9} \approx 8.2.$$

在這個模型下，一個比平均值高出 8 個標準差的數值，其概率的數量級為 10^{-15}。即使實際上每年之間的離散程度是泊松值的兩三倍（比方說因為嚴冬），相對於 1080 這個穩定的速率，1350 仍然非常異常。所以極有可能有些東西改變了。問題是：改變的是甚麼？

*與甚麼比較？*標題是拿 1990 年與一個 16 年的平均數比較，而不是與 1989 年比較。假如行人死亡人數在這段時期內一直穩步上升，那麼平均數 1080 描述的就是這段時期的中段（約 1981–82 年），而 1990 年理應遠高於這個數。例如，由 1974 年約 855 人起，每年穩定增加 30 人，平均數便是 1080，而 1990 年則約為 1335 人。在這種情況下，1990 年根本算不上甚麼「急升」，只是長期趨勢的延續。這個趨勢或許值得上頭條，但應該是另一則標題。我們需要逐年的數字，才能分辨這兩種說法。

*風險承擔量有沒有改變？*人口增加、車輛增加或步行增加，都會令死亡人數上升，卻不代表按每程計步行變得更危險。公平的比較應該用比率，例如每 100 000 人口的死亡人數，或每步行一公里的死亡人數。

*量度方法有沒有改變？*死亡個案的記錄或分類方法一旦改變，數字便會出現表面上的跳

升。例子包括：意外發生後多久之內死亡仍算作道路死亡的準則有變、誰算是行人的定義有變，或者呈報情況有所改善。任何這類改變都必須先予排除。

用詞是否恰當？「傷亡」通常是指死亡或受傷的人。所引述的數字只計死亡，因此標題誇大了這個數字所能說明的事。受傷的人數多得多，而且可能呈現截然不同的規律。

總括而言：升至 1350 的幅度實在太大，不能用圍繞恆定速率 1080 的泊松波動來解釋。但在稱之為「令人震驚」之前，我們需要近年的年度數字（以區分突然上升與長期趨勢），以及人口和交通數字（以把數目化為比率）。我們還須查證各項定義沒有改變。

驗算

各個概率已透過精確地求泊松級數之和，以及進行 2×10^5 次試驗的模擬來核對。在 (iii) 中，當 $\lambda = 1080/365$ 時，$P(X \geq 7) = 0.0315$（模擬結果：0.0317）。在 (ii) 中，若每年有 6 宗墜機、每宗 30 人死亡，則 $P(N \leq 4) = 0.285$。若讓每宗墜機的死亡人數在 1 至 59 之間均勻變動，模擬所得的標準差為 84，而 $P(\text{死亡人數} \leq 125) \approx 0.28$。對泊松計數有一條簡便的經驗法則：與平均值相差約兩個平均值平方根以內的數值，都不足為奇。就一天而言，$3 + 2 \times 1.7 \approx 6.4$，所以 7 只是剛剛超出這個範圍，而且有 365 天可供挑選，這樣的日子必然會出現。就一年而言，$1080 + 2 \times 33 = 1146$，與 1350 相距甚遠。

常見錯誤

- 在 (i) 中比較總數而不是比率：風險永遠需要一個分母。
- 在 (ii) 中使用泊松標準差 $\sqrt{180}$。空難中的死亡是成羣出現的，所以離散程度大得多。
- 在 (iv) 中說到「8 個標準差，所以是真正的上升」便停下來，而沒有追問：與 16 年平均數比較這一點、人口或定義的變化，能否解釋這個升幅。

STEP 之外

第 (ii) 部分談的是過度離散（overdispersion）。一個由泊松數目的羣組合成、而每個羣組大小隨機的總數，是一個複合泊松（compound Poisson）變量，它的方差（variance）是 $\lambda \, \mathrm{E}[S^2]$，而不是 λ。第 (iv) 部分則通往假設檢定（hypothesis testing），以及一個困難得多的問題：在統計上顯著的變化，究竟意味着甚麼。兩者都會在大學概率與統計課程的初段出現。Blitzstein 與 Hwang 合著的 *Introduction to Probability*（概率論入門教材），是閱讀泊松模型及其局限的好去處。

Beyond STEP 詳細解答：https://beyondstep.keithmcnulty.org/zh-hant/。解答文字採用 CC0 1.0；題目本身版權屬 Cambridge University Press & Assessment 所有（©），本文件不予轉載。由機器從英文翻譯而成；尚未經審閱。