詳細解答

STEP III 1989，第 16 題

概率與統計 • 概率與假設檢驗

STEP III 1989，第 16 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/89-s3-q16

關鍵思路

三個「是／否」特徵把人口分成八格，而那八項陳述不過是關於各格比例的線性方程。先填綠眼的各格（那裏的條件最具體），再填淡褐眼的各格。之後的統計部分全靠一個數，即一個魯里坦尼亞人是窄頭的概率，並問在這個數值之下，觀察所得是否令人意外。

解答

試卷的印刷本有三處筆誤。我們把 (iv) 中的「arrow-headed」讀作 *narrow-headed*（窄頭），把 (vi) 中的「fair-headed」讀作 *fair-haired*（金髮），把 (viii) 中的「hazel-haired」讀作 *hazel-eyed*（淡褐眼）。這是唯一合理的讀法，因為三個特徵分別是頭髮（金或深色）、眼睛（綠或淡褐）和頭形（窄或闊）。

窄頭者所佔的比例

隨機選出一個魯里坦尼亞人。以 F、D 表示金髮和深色頭髮，以 G、H 表示綠眼和淡褐眼，以 N、B 表示窄頭和闊頭。已知 $\mathrm{P}(F) = \frac{1}{4}$ 及 $\mathrm{P}(G) = \frac{1}{5}$。

綠眼的各格。由 (v)，$\mathrm{P}(N \mid G) = \mathrm{P}(B \mid G) = \frac{1}{2}$，所以

$$\mathrm{P}(G \cap N) = \mathrm{P}(G \cap B) = \frac{1}{10}.$$

由 (vi)，綠眼的闊頭者平均分為金髮和深色頭髮兩半：

$$\mathrm{P}(F \cap G \cap B) = \mathrm{P}(D \cap G \cap B) = \frac{1}{20}.$$

由 (iv)，$\mathrm{P}(F \cap G \cap N) = 0$，所以 $\mathrm{P}(D \cap G \cap N) = \frac{1}{10}$。

淡褐眼的各格。設

$$a = \mathrm{P}(F \cap H \cap B), \quad b = \mathrm{P}(F \cap H \cap N), \quad c = \mathrm{P}(D \cap H \cap B), \quad d = \mathrm{P}(D \cap H \cap N).$$

- 金髮：綠眼的金髮者佔 $0 + \frac{1}{20}$，所以 $a + b = \frac{1}{4} - \frac{1}{20} = \frac{1}{5}$。
- 淡褐眼：$a + b + c + d = \frac{4}{5}$，所以 $c + d = \frac{3}{5}$。
- 陳述 (vii)：$\mathrm{P}(B \cap D) = \frac{1}{20} + c = \frac{1}{2}$，所以 $c = \frac{9}{20}$，因此 $d = \frac{3}{20}$。
- 陳述 (viii)：$\mathrm{P}(F \cap B \mid H) = \mathrm{P}(D \cap N \mid H)$；兩邊乘以 $\mathrm{P}(H)$ 得 $a = d = \frac{3}{20}$，於是 $b = \frac{1}{20}$。

全部八個比例，以二十分之一為單位，列出如下：

	$G \cap N$	$G \cap B$	$H \cap N$	$H \cap B$
F	0	1	1	3
D	2	1	3	9

它們全部非負，而且加起來是 20，所以這個模型並無矛盾。把窄頭的兩欄相加，

$$\mathrm{P}(N) = \frac{0+2+1+3}{20} = \frac{3}{10}.$$

所以按照這個模型，有 30% 的魯里坦尼亞人是窄頭的。

六個闊頭的相識

假如我這六個相識是一個隨機樣本（random sample），而且彼此獨立（independent），那麼六人全是闊頭的機會是

$$\left(\tfrac{7}{10}\right)^6 = 0.1176\ldots \approx 0.12.$$

想一想 一件概率為 12% 的事件，算得上是否定模型的有力證據嗎？

算不上。一件大約每 8 次就有 1 次發生的事並不罕見；在 5% 這類慣用的顯著性水平（significance level）下，它都不顯著。這些數據還有另一個弱點：一個人的相識並不是隨機樣本。他們可能來自同一家庭、同一地區或同一社交圈子，所以未必彼此獨立，也未必能代表整個人口。這項觀察頂多只提供極輕微的懷疑理由。

一個 200 人的隨機樣本

設 X 為一個 200 人的隨機樣本中窄頭者的數目。人口很大，所以各次抽取實際上是獨立的，而在這個模型之下，$X \sim \mathrm{B}(200, \frac{3}{10})$，且

$$\mathrm{E}(X) = 60, \qquad \mathrm{Var}(X) = 200 \times 0.3 \times 0.7 = 42, \qquad \sqrt{42} \approx 6.481.$$

由於 np 和 $n(1-p)$ 都很大，我們採用正態近似（normal approximation）$X \approx \mathrm{N}(60, 42)$，並加上連續性校正（continuity correction）：

$$\mathrm{P}(X \leq 50) \approx \Phi\left(\frac{50.5-60}{\sqrt{42}}\right) = \Phi(-1.466) \approx 0.071.$$

所以在這個模型之下，窄頭者不多於 50 人的情況大約有 7% 的機會出現。

想一想 我們現在應該否定這個模型嗎？

在 5% 的水平下不應該：以 $p = 0.3$ 對 $p < 0.3$ 作單尾檢驗（one-tailed test），所得的 p 值約為 0.07。不過，這個結果確實令人對模型有些懷疑。樣本比例 $\frac{50}{200} = 0.25$ 比預測值低了約 1.5 個標準差（standard deviation）。它的方向也與我那些相識的情況一致，他們當中的窄頭者同樣比預期少。在相信模型所給的 30% 這個數字之前，多收集一些數據才是明智的做法。（假如檢驗是雙尾的，p 值便約為 0.14，證據就更弱了。）

驗算

- 八個條件每一個都可以直接從表中讀出，例如 (vii)：$1 + 9 = 10$ 個二十分之一 $= \frac{1}{2}$。
- 用電腦核對，可確認各格的比例是唯一解，$\mathrm{P}(N) = \frac{3}{10}$，以及 $(0.7)^6 = 0.117649$。精確的二項概率 $\mathrm{P}(X \leq 50) = 0.0695$ 與正態近似值 0.0713 相當接近。

常見錯誤

- 把 (v) 或 (vi) 讀成關於整個人口的陳述，而不是條件（conditional）陳述：在綠眼的一組之內「同樣可能」（as likely as），意思是把 $\frac{1}{5}$ 平分成兩半。
- 忘記連續性校正。用 50 而不用 50.5 會得到 $\Phi(-1.543) \approx 0.061$。
- 把相識當作隨機樣本而不加任何說明。

STEP 之外

最後一部分是一個假設檢驗（hypothesis test），而那兩項證據引出一個很自然的問題：幾項薄弱的證據應該怎樣合併起來？大學的統計學以似然（likelihood）來回答這個問題（對貝葉斯學派（Bayesian）而言，還有後驗概率（posterior probability）），它們衡量 p 的每一個值對全部數據合起來解釋得有多好。Grimmett 與 Welsh 的 *Probability: An Introduction*（概率論入門）提供所需的概率背景，而 Blitzstein 與 Hwang 的 *Introduction to Probability*（概率導論）則對貝葉斯觀點有一段淺白易讀的介紹。

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。