详细解答

STEP III 1989，第 16 题

概率与统计 · 概率与假设检验

> **STEP III 1989，第 16 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/89-s3-q16

关键思路

三个“是／否”特征把人口分成八格，而那八项陈述不过是关于各格比例的线性方程。先填绿眼的各格（那里的条件最具体），再填淡褐眼的各格。之后的统计部分全靠一个数，即一个鲁里坦尼亚人是窄头的概率，并问在这个数值之下，观察所得是否令人意外。

解答

试卷的印刷本有三处笔误。我们把 (iv) 中的“arrow-headed”读作 *narrow-headed*（窄头），把 (vi) 中的“fair-headed”读作 *fair-haired*（金发），把 (viii) 中的“hazel-haired”读作 *hazel-eyed*（淡褐眼）。这是唯一合理的读法，因为三个特征分别是头发（金或深色）、眼睛（绿或淡褐）和头形（窄或阔）。

窄头者所占的比例

随机选出一个鲁里坦尼亚人。以 F、D 表示金发和深色头发，以 G、H 表示绿眼和淡褐眼，以 N、B 表示窄头和阔头。已知 $\mathrm{P}(F) = \frac{1}{4}$ 及 $\mathrm{P}(G) = \frac{1}{5}$。

绿眼的各格。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)，绿眼的阔头者平均分为金发和深色头发两半：

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

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。设

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

- 金发：绿眼的金发者占 $0 + \frac{1}{20}$，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 $c + d = \frac{3}{5}$。
- 陈述 (vii)：$\mathrm{P}(B \cap D) = \frac{1}{20} + c = \frac{1}{2}$，所以 $c = \frac{9}{20}$，因此 $d = \frac{3}{20}$。
- 陈述 (viii)：$\mathrm{P}(F \cap B \mid H) = \mathrm{P}(D \cap N \mid H)$；两边乘以 $\mathrm{P}(H)$ 得 $a = d = \frac{3}{20}$，于是 $b = \frac{1}{20}$。

全部八个比例，以二十分之一为单位，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

它们全部非负，而且加起来是 20，所以这个模型并无矛盾。把窄头的两栏相加，

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

所以按照这个模型，有 30% 的鲁里坦尼亚人是窄头的。

六个阔头的相识

假如我这六个相识是一个随机样本（random sample），而且彼此独立（independent），那么六人全是阔头的机会是

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

想一想 一件概率为 12% 的事件，算得上是否定模型的有力证据吗?

算不上。一件大约每 8 次就有 1 次发生的事并不罕见；在 5% 这类惯用的显著性水平（significance level）下，它都不显著。这些数据还有另一个弱点：一个人的相识并不是随机样本。他们可能来自同一家庭、同一地区或同一社交圈子，所以未必彼此独立，也未必能代表整个人口。这项观察顶多只提供极轻微的怀疑理由。

一个 200 人的随机样本

设 X 为一个 200 人的随机样本中窄头者的数目。人口很大，所以各次抽取实际上是独立的，而在这个模型之下，$X \sim \mathrm{B}(200, \frac{3}{10})$，且

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

由于 np 和 $n(1-p)$ 都很大，我们采用正态近似（normal approximation）$X \approx \mathrm{N}(60, 42)$，并加上连续性校正（continuity correction）：

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

所以在这个模型之下，窄头者不多于 50 人的情况大约有 7% 的机会出现。

想一想 我们现在应该否定这个模型吗?

在 5% 的水平下不应该：以 $p = 0.3$ 对 $p < 0.3$ 作单尾检验（one-tailed test），所得的 p 值约为 0.07。不过，这个结果确实令人对模型有些怀疑。样本比例 $\frac{50}{200} = 0.25$ 比预测值低了约 1.5 个标准差（standard deviation）。它的方向也与我那些相识的情况一致，他们当中的窄头者同样比预期少。在相信模型所给的 30% 这个数字之前，多收集一些数据才是明智的做法。（假如检验是双尾的，p 值便约为 0.14，证据就更弱了。）

验算

- 八个条件每一个都可以直接从表中读出，例如 (vii)：$1 + 9 = 10$ 个二十分之一 $= \frac{1}{2}$。
- 用电脑核对，可确认各格的比例是唯一解，$\mathrm{P}(N) = \frac{3}{10}$，以及 $(0.7)^6 = 0.117649$。精确的二项概率 $\mathrm{P}(X \leq 50) = 0.0695$ 与正态近似值 0.0713 相当接近。

常见错误

- 把 (v) 或 (vi) 读成关于整个人口的陈述，而不是条件（conditional）陈述：在绿眼的一组之内“同样可能”（as likely as），意思是把 $\frac{1}{5}$ 平分成两半。
- 忘记连续性校正。用 50 而不用 50.5 会得到 $\Phi(-1.543) \approx 0.061$。
- 把相识当作随机样本而不加任何说明。

STEP 之外

最后一部分是一个假设检验（hypothesis test），而那两项证据引出一个很自然的问题：几项薄弱的证据应该怎样合并起来? 大学的统计学以似然（likelihood）来回答这个问题（对贝叶斯学派（Bayesian）而言，还有后验概率（posterior probability）），它们衡量 p 的每一个值对全部数据合起来解释得有多好。Grimmett 与 Welsh 的 *Probability: An Introduction*（概率论入门）提供所需的概率背景，而 Blitzstein 与 Hwang 的 *Introduction to Probability*（概率导论）则对贝叶斯观点有一段浅白易读的介绍。

Beyond STEP 详细解答：https://beyondstep.keithmcnulty.org/zh-hans/。解答文字采用 CC0 1.0；题目本身版权属 Cambridge University Press & Assessment 所有（©），本文件不予转载。由机器从英文翻译而成；尚未经审阅。