13.1 解读新闻标题

工作纸

题目

STEP I 1992，第 14 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/92-s1-q14

提示

请先自行尝试解题。如果遇到困难，请逐一阅读提示，每读完一个便回到题目再试。标有*停一停，试一试*的提示是关键步骤。每个提示之后均留有空位供你演算。

> **想一想** 假如死亡事件以稳定的比率独立而随机地发生，每年的总数应有多大的变动？若平均每年为 1080，每年与平均值的典型偏差是多少？

阶段 1 “坐飞机比步行安全六倍”

涵盖：第 (i) 部分

> **想一想** 比值 $1080/180 = 6$ 比较的是总死亡人数。这与比较风险是同一回事吗？一个公平的风险量度需要考虑哪些因素？

阶段 2 “我们的天空更安全：1990 年仅 125 人丧生”

涵盖：第 (ii) 部分

想一想 空难死亡是独立（independent）的罕见事件，还是会成批出现？你预期每年的总数会有多大的波动？单凭一年的数字能看出趋势吗？

阶段 3 “道路惨剧有增无减：星期二 7 人丧生”

涵盖：第 (iii) 部分

停一停，试一试 平均每天有多少名行人丧生？若每天的死亡人数服从以此为平均值（mean）的泊松分布（Poisson distribution），你预期一年之中有多少天会有 7 人或以上丧生？单凭一天的数字，又能否说明某事正在增加？

阶段 4 “升幅惊人：1990 年 1350 名行人丧生”

涵盖：第 (iv) 部分

? **想一想** 在平均值为 1080 的泊松模型下，1350 比平均值高出多少个标准差（standard deviation）？若这并非偶然，还有甚么其他原因可以解释？想一想 16 年的平均数可能掩盖了甚么。

背后的数学

信号与杂讯

每一个统计问题，归根究底都是要把*信号*（signal，即真实的效应或趋势）与*杂讯*（noise，即随机波动）分开。泊松模型为独立事件计数中的杂讯提供了一把尺：标准差为 $\sqrt{\mu}$。罕见事件的*相对*波动很大：计数为 9 时，波动约为 ± 3，即其大小的三分之一；而计数为 10 000 时，波动只有约 1%。正因如此，小数目（某一天的死亡人数、某一年的坠机次数）才会屡屡被人过度解读，而小镇上的癌症“群组”最终也多半只是偶然。

均值回归

还有另一个陷阱。道路安全摄影机往往安装在刚刚经历了异常恶劣一年的地点。翌年，意外减少，功劳就归于摄影机。但异常恶劣的一年过后，本来就很可能迎来较为平常的一年：这就是*均值回归*（regression to the mean），最早由高尔顿（Francis Galton）在 1880 年代提出。要把摄影机的效果与均值回归分开，就必须与没有安装摄影机的类似地点作比较，这正是对照实验（controlled experiment）背后的理念。

南丁格尔的统计学

南丁格尔（Florence Nightingale）以克里米亚战争中的护士身分闻名于世，同时也是统计学的先驱。她在 1858 年绘制的“玫瑰图”显示，死于可预防疾病的士兵远多于死于

创伤的士兵，并说服了政府改革军队的卫生状况。她坚持使用比率（每年每千名士兵的死亡人数）而非原始总数，并坚持同类才可相比，这些正是本题所考核的要点。她是首位获选为皇家统计学会（Royal Statistical Society）会士的女性。

延伸练习

1 **泊松分布的尾部。**设 X 服从平均值为 2.96 的泊松分布，验证 $\mathrm{P}(X \geq 7) \approx 0.031$。你预期每年有多少天会有 9 名或以上的行人死亡？

2 **成群出现的死亡。**假设坠机事故按泊松过程（Poisson process）发生，比率为每年 1.2 宗，而每宗夺去 150 人的性命。求每年空难死亡人数的平均值和标准差。一年之内有 125 人或以下死亡，是否不寻常？有没有可能某一年恰好有 125 人死亡？

3 **显著的升幅？**假如 1974–1989 年的数字一直平稳地维持在 1080 ± 33，试（利用正态近似（normal approximation））求单一年份出现 1350 或以上的概率。

4 **拟一个更好的标题。**为四个标题逐一改写，各拟一个数据真正支持的标题。

延伸阅读

- **David Spiegelhalter, *The Art of Statistics***（Pelican）。讲述如何从数据中学习，并附有大量取自新闻的例子。
- **Tim Harford, *How to Make the World Add Up***（Bridge Street）。思考传媒中的数字的十条法则。
- **Darrell Huff, *How to Lie with Statistics***（Penguin）。1954 年出版的短篇经典，至今仍不过时。

完整的引导式解答见第 13.1 节：

https://beyondstep.keithmcnulty.org/zh-hans/ch13/headlines.html

每星期一道 STEP 题目，完整讲解：The STEP Up，以电子邮件免费发送，订阅网址：keithmcnulty.substack.com