详细解答

STEP I 1992，第 14 题

概率与统计 · 统计论断的诠释

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

本题在 *Beyond STEP* 第 13.1 节解读新闻标题中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch13/headlines.html

关键思路

每则标题都是拿着一个数字，就一跃跳到结论。对每一则标题，都要问两个问题。第一，它是否在拿同类的东西作比较：同一样事物，用同一种方法量度，并按承受风险的程度（exposure）作出调整？第二，这个差异会不会纯属偶然？对于第二个问题，一把很自然的尺是泊松模型（Poisson model），它描述罕有而互相独立的事件的发生次数，其标准差（standard deviation）等于平均值的平方根。懂得这个模型何时适用、何时不适用，正是 (ii) 至 (iv) 的核心。

解答

以下所说的“平均”，一律指 1974–1989 年的年平均数：每年 1080 名行人死亡，180 人死于空难。

第 (i) 部分

“六倍”之说来自比率 $1080/180 = 6$，但这个比率比较的是原始总数，而不是风险。风险是指每单位风险承担量的死亡人数：按每程、每小时或每公里计算。在 Poldavia，几乎人人每天都在马路附近步行，而乘飞机的人相对很少，而且只是偶一为之。要比较两者

的安全程度，我们需要的是类似

$$\frac{\text{行人死亡人数}}{\text{步行的总时数（或总公里数）}} \quad \text{对比} \quad \frac{\text{空难死亡人数}}{\text{飞行的总时数（或总公里数）}},$$

这样的比较，而标题却没有给出任何一个分母。视乎所用的量度方法，答案甚至可能倒转过来：按每公里计，飞行看来非常安全；但按每小时或每程计，就未必那么安全了。

此外还有其他问题。这里所谓的“步行”，其实是指徒步时被车辆撞倒，这个数字完全没有反映行人在僻静地方的情况。空难死亡人数除了乘客，还包括机组人员。而且一段时期的平均数根本无法“证明”1991 年的安全情况；它们充其量只是关于 1974–1989 年的证据。

第 (ii) 部分

乍看之下，125 远低于平均数 180。这个跌幅是否超出了偶然的范围？

假如死亡事件是互相独立、逐宗发生的，那么采用平均值为 180 的泊松模型便算合理。它的标准差是 $\sqrt{180} \approx 13.4$，因此 125 大约比平均值低 4.1 个标准差，纯粹出于偶然的可能性极低（$P(X \leq 125) \approx 10^{-5}$）。

但空难死亡并不是逐宗发生的，而是成群出现：一次坠机就可以夺去数十以至数百人的性命。较自然的模型是：*致命坠机*的次数服从泊松分布，而每次坠机的死亡人数是随机的。这样一来，变异程度便大大增加。举例说，假设平均每年有 6 宗致命坠机，每宗有 30 人死亡。那么全年死亡人数是 $30N$，其中 $N \sim \text{Po}(6)$，其标准差为 $30\sqrt{6} \approx 73$，而

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

按这些（仅作说明用的）数字，一年有 125 人死亡实在平常不过：只需要比平常少一两宗坠机便会出现。如果连每宗坠机的死亡人数也容许变动，离散程度还会更大。

所以单凭一年的数字，我们所知甚少。我们还想知道 1990 年的飞行量有多少（如果航空交通减少了，死亡人数本来就应该较少），以及好几年的数字，才能看出是否真的有趋势。标题的说法并无根据。

第 (iii) 部分

这里有两个不同的毛病。

单凭一天不能显示上升。“上升”是关于一段时间内趋势的说法，而单独一天的数字根本无从比较。我们需要一段长时间内每天（或每年）的数字。

*七人根本不算异常。*我们手上唯一的道路数字是行人的：每年 1080 人，即

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

用泊松分布来模拟行人死亡相当合适：这类事件罕有，大多是每宗意外一人死亡，而且大致互相独立。设 $X \sim \mathrm{Po}(2.96)$ 为某一天死亡的行人数目。那么

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

所以我们预期一年之中大约有 $365 \times 0.031 \approx 11$ 天，会有最少 7 名行人死亡。况且标题说的是“死亡人数”，那还包括司机、乘客和骑单车的人。每天道路死亡总人数的平均率必定高于 2.96，于是 7 就更加不足为奇了。

想一想 某一天的死亡人数是平均数的两倍有多，难道不令人震惊吗?

数目小的时候并不然。当平均值约为 3 时，泊松标准差是 $\sqrt{3} \approx 1.7$，与平均值相比已经很大，所以每天之间按比例计出现大幅波动是正常的。专挑最差的一天，也是一种选择效应（selection effect）：一年有 365 天，报章总能找到差的一天。那天是星期几也有关系，因为交通流量和外出步行的人数会随星期几和季节而变化。速率恒定的泊松模型只是一个粗略的初步近似。

第 (iv) 部分

这一则标题确实有值得认真回应之处。先看统计，再谈诠释。

*1350 是否超出了偶然的范围？*基于 (iii) 所述的理由，泊松模型在这里是个不错的起点。平均值为 1080 时，标准差是 $\sqrt{1080} \approx 32.9$，而

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

在这个模型下，一个比平均值高出 8 个标准差的数值，其概率的数量级为 10^{-15}。即使实际上每年之间的离散程度是泊松值的两三倍（比方说因为严冬），相对于 1080 这个稳定的速率，1350 仍然非常异常。所以极有可能有些东西改变了。问题是：改变的是甚么?

*与甚么比较？*标题是拿 1990 年与一个 16 年的平均数比较，而不是与 1989 年比较。假如行人死亡人数在这段时期内一直稳步上升，那么平均数 1080 描述的就是这段时期的中段（约 1981–82 年），而 1990 年理应远高于这个数。例如，由 1974 年约 855 人起，每年稳定增加 30 人，平均数便是 1080，而 1990 年则约为 1335 人。在这种情况下，1990 年根本算不上甚么“急升”，只是长期趋势的延续。这个趋势或许值得上头条，但应该是另一则标题。我们需要逐年的数字，才能分辨这两种说法。

*风险承担量有没有改变？*人口增加、车辆增加或步行增加，都会令死亡人数上升，却不代表按每程计步行变得更危险。公平的比较应该用比率，例如每 100 000 人口的死亡人数，或每步行一公里的死亡人数。

*量度方法有没有改变？*死亡个案的记录或分类方法一旦改变，数字便会出现表面上的跳

升。例子包括：意外发生后多久之内死亡仍算作道路死亡的准则有变、谁算是行人的定义有变，或者呈报情况有所改善。任何这类改变都必须先予排除。

用词是否恰当？“伤亡”通常是指死亡*或受伤*的人。所引述的数字只计死亡，因此标题夸大了这个数字所能说明的事。受伤的人数多得多，而且可能呈现截然不同的规律。

总括而言：升至 1350 的幅度实在太大，不能用围绕恒定速率 1080 的泊松波动来解释。但在称之为“令人震惊”之前，我们需要近年的年度数字（以区分突然上升与长期趋势），以及人口和交通数字（以把数目化为比率）。我们还须查证各项定义没有改变。

验算

各个概率已通过精确地求泊松级数之和，以及进行 2×10^5 次试验的模拟来核对。在 (iii) 中，当 $\lambda = 1080/365$ 时，$P(X \geq 7) = 0.0315$（模拟结果：0.0317）。在 (ii) 中，若每年有 6 宗坠机、每宗 30 人死亡，则 $P(N \leq 4) = 0.285$。若让每宗坠机的死亡人数在 1 至 59 之间均匀变动，模拟所得的标准差为 84，而 $P(\text{死亡人数} \leq 125) \approx 0.28$。对泊松计数有一条简便的经验法则：与平均值相差约两个平均值平方根以内的数值，都不足为奇。就一天而言，$3 + 2 \times 1.7 \approx 6.4$，所以 7 只是刚刚超出这个范围，而且有 365 天可供挑选，这样的日子必然会出现。就一年而言，$1080 + 2 \times 33 = 1146$，与 1350 相距甚远。

常见错误

- 在 (i) 中比较总数而不是比率：风险永远需要一个分母。
- 在 (ii) 中使用泊松标准差 $\sqrt{180}$。空难中的死亡是成群出现的，所以离散程度大得多。
- 在 (iv) 中说到“8 个标准差，所以是真正的上升”便停下来，而没有追问：与 16 年平均数比较这一点、人口或定义的变化，能否解释这个升幅。

STEP 之外

第 (ii) 部分谈的是*过度离散*（overdispersion）。一个由泊松数目的群组合成、而每个群组大小随机的总数，是一个*复合泊松*（compound Poisson）变量，它的方差（variance）是 $\lambda\,\mathrm{E}[S^2]$，而不是 λ。第 (iv) 部分则通往假设检定（hypothesis testing），以及一个困难得多的问题：在统计上显著的变化，究竟意味着甚么。两者都会在大学概率与统计课程的初段出现。Blitzstein 与 Hwang 合著的 *Introduction to Probability*（概率论入门教材），是阅读泊松模型及其局限的好去处。

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。