13.1 신문 머리기사 읽기

학습지

문제

STEP I 1992, 14번 문제

더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:

step.maths.org/questions/92-s1-q14

힌트

먼저 혼자 힘으로 문제를 풀어 보자. 막히면 힌트를 **하나씩** 읽고, 힌트를 하나 읽을 때마다 다시 문제로 돌아가자. **잠깐, 직접 해 보기**라고 표시된 힌트가 핵심 단계이다. 각 힌트 뒤에는 풀이를 적을 공간이 있다.

> **생각해 보기** 사망 사고가 일정한 비율로 서로 독립적이고 무작위로 일어난다면, 연간 총계는 얼마나 변동해야 하는가? 평균이 한 해에 1080이라면, 해마다 나타나는 전형적인 편차는 얼마인가?

1단계 "걷는 것보다 비행기가 여섯 배 더 안전하다"

다루는 내용: 소문항 (i)

> **생각해 보기** 비 $1080/180 = 6$은 전체 사망자 수를 비교한 것이다. 이것이 위험도를 비교하는 것과 같은가? 위험도를 공정하게 측정하려면 무엇을 고려해야 하는가?

2단계 **"하늘은 더 안전해졌다: 1990년 사망자는 단 125명"**

다루는 내용: 소문항 (ii)

생각해 보기 항공 사고 사망은 서로 독립인(independent) 드문 사건인가, 아니면 한꺼번에 몰려서 일어나는가? 연간 총계가 얼마나 변동할 것이라고 예상하는가? 한 해의 자료로 추세를 보여 줄 수 있는가?

3단계 **"도로 참사 증가: 화요일 하루에만 7명 사망"**

다루는 내용: 소문항 (iii)

잠깐, 직접 해 보기 하루 평균 보행자 사망자는 몇 명인가? 하루 사망자 수가 그 값을 평균으로 하는 푸아송 분포(Poisson distribution)를 따른다면, 1년 중 사망자가 7명 이상인 날은 며칠쯤 될 것이라고 예상하는가? 그리고 단 하루의 자료로 무언가가 증가하고 있음을 보여 줄 수 있는가?

4단계 **"우려스러운 증가: 1990년 보행자 1350명 사망"**

다루는 내용: 소문항 (iv)

> ? **생각해 보기** 평균이 1080인 푸아송 모형에서 1350은 평균보다 표준편차(standard deviation)의 몇 배만큼 큰가? 우연이 아니라면 달리 어떤 설명이 가능한가? 16년에 걸친 평균이 무엇을 가릴 수 있는지 생각해 보자.

숨은 수학

신호와 잡음

모든 통계적 질문은 근본적으로 **신호**(*signal*)(실제 효과나 추세)와 **잡음**(*noise*)(무작위 변동)을 분리하는 문제이다. 푸아송 모형은 서로 독립인 사건의 개수에 담긴 잡음을 재는 잣대, 곧 표준편차 $\sqrt{\mu}$를 준다. 드문 사건은 **상대적** 변동이 크다. 개수 9는 약 ± 3, 즉 그 크기의 3분의 1만큼 변동하지만, 개수 10 000은 약 1%만큼만 변동한다. 그래서 작은 수(하루의 사망자 수, 한 해의 추락 사고 수)가 그토록 자주 과잉 해석되며, 작은 마을의 암 "집단 발생"이 대개 우연으로 밝혀지는 것이다.

평균으로의 회귀

함정이 하나 더 있다. 도로 안전 카메라는 흔히 막 유난히 나쁜 한 해를 보낸 장소에 설치된다. 이듬해 사고가 줄어들면 카메라가 공을 인정받는다. 그러나 유난히 나쁜 해 다음에는 어차피 더 평범한 해가 올 가능성이 크다. 이것이 1880년대에 프랜시스 골턴(Francis Galton)이 처음 기술한 **평균으로의 회귀**(*regression to the mean*)이다. 카메라의 효과를 평균으로의 회귀와 분리하려면 카메라가 설치되지 않은 비슷한 장소와 비교해야 하며, 이것이 대조 실험(controlled experiment)의 바탕에 있는 생각이다.

플로렌스 나이팅게일의 통계

크림 전쟁의 간호사로 유명한 플로렌스 나이팅게일(Florence Nightingale)은 선구적인 통계학자이기도 했다. 1858년 그녀의 "장미 도표"는 부상보다 예방 가능한 질병으로 죽어 가는 병사가 훨씬 더 많다는 것을 보여 주었고, 정부가 군대의 위생을 개혁하도록 설득했다. 그녀는 원래의 총계가 아니라 비율(1년간 병사 천 명당 사망자 수)을 쓸 것과 같은 것끼리 비교할 것을 고집했는데, 바로 이 문제가 시험하는 점들이다. 그녀는 왕립통계학회(Royal Statistical Society)의 펠로로 선출된 최초의 여성이었다.

더 도전하기

1 **푸아송 꼬리.** X가 평균이 2.96인 푸아송 분포를 따를 때 $\mathrm{P}(X \geq 7) \approx 0.031$임을 확인하라. 보행자 사망자가 9명 이상인 날은 1년에 며칠일 것으로 예상하는가?

2 **무리 지어 일어나는 사망.** 항공기 추락 사고가 연간 1.2회의 비율로 푸아송 과정(Poisson process)을 따라 일어나고, 사고마다 150명이 사망한다고 하자. 연간 항공 사망자 수의 평균과 표준편차를 구하라. 사망자가 125명 이하인 해는 이례적인가? 사망자가 정확히 125명인 해가 있을 수 있는가?

3 **유의한 증가?** 1974–1989년의 수치가 1080 ± 33으로 일정했다면, (정규 근사(normal approximation)를 이용하여) 한 해에 1350명 이상이 사망할 확률을 구하라.

4 **더 나은 헤드라인을 써라.** 네 헤드라인 각각에 대해, 데이터가 실제로 뒷받침하는 헤드라인을 써라.

더 읽을거리

- **David Spiegelhalter, *The Art of Statistics*** (Pelican). 뉴스에서 가져온 많은 예와 함께, 데이터로부터 배우는 방법을 다룬다.
- **Tim Harford, *How to Make the World Add Up*** (Bridge Street). 언론에 나오는 숫자에 대해 생각하는 열 가지 규칙.
- **Darrell Huff, *How to Lie with Statistics*** (Penguin). 1954년에 나온 짧은 고전으로, 지금도 여전히 유효하다.

안내에 따른 전체 풀이는 제 13.1절에 있다:

https://beyondstep.keithmcnulty.org/ko/ch13/headlines.html

매주 STEP 문제 하나를 끝까지 풀어 본다. The STEP Up은 이메일로 무료로 받아 볼 수 있다: keithmcnulty.substack.com