

牌号 7 · 单位 8 · 评估指南

概率和抽样

家长或教师指南——观察要点

单元 8 构建 概率 和 采样直觉。经典错误：混淆实验概率和理论概率、将小样本视为确凿证据以及有偏见的抽样（只询问你的朋友他们最喜欢的电影是什么）。更深层次的问题是：样本什么时候“足够好”？什么时候差异才“有意义”？观察学生是否将差异与变异性（MAD）进行比较。

问题 1 — 似然词汇

- (a) 不可能。(b) 可能性与不可能性相同。(c) 肯定的（在大多数地方）。(d) 可能性与不可能性相同 ($26/52 = 1/2$)。
- ★优秀的学生给出精确的词汇。(b) 和 (d) 都应 = $1/2$ 。

问题 2 — 旋转器

- (a) $3/8 = 37.5\%$ 。(b) $4/8 = 1/2$ 。(c) $7/8$ 。(d) 约 10 次。
- ★强：对 分数 到 百分比 的转换充满信心。(d) 测试期望值推理。

问题 3 — 大理石包

- 总计 = 10。 $P(\text{红色}) = 5/10 = 1/2$ 。 $P(\text{蓝色}) = 3/10$ 。 $P(\text{红色或蓝色}) = 8/10 = 4/5$ 。
- 所有相加到 1，因为每次抽奖必须是这些颜色之一（完整的样本空间）。
- ★“必须等于 1”的见解是关键采样完整性思想。

问题 4 — 两个硬币

- 样本空间：HH、HT、TH、TT。4 结果。
- $P(\text{HH}) = 1/4$ 。 $P(\text{至少一次反面}) = 3/4$ 。
- ★注意只列出 HH、HT、TT（3 结果）的学生 - 忘记了顺序很重要。
- 或者说 $P(\text{至少一次反面}) = 1/2$ （将 HT 和 TH 视为一种结果的错误）。

问题 5 — 两个骰子

- 总计：36 结果 (6×6)。
- $P(\text{和} = 7) = 6/36 = 1/6$ 。(1+6, 2+5, 3+4, 4+3, 5+2, 6+1.)
- $P(\text{和} = 12) = 1/36$ （仅 6+6）。
- $P(\text{和} > 9) = 6/36 = 1/6$ （10、11、12 总和：4+6, 5+5, 6+4、5+6, 6+5、6+6 → 6 结果的组合）。
- ★较好表现：用表格或树状图有条理地列举。较弱表现：遗漏结果或重复计数。典型错误：认为不同的和等可能出现（如说 $P(\text{和}=2)=P(\text{和}=7)$ ——这是错的）。
- ★

问题 6 — 林的 13 在 20 翻转中头球

- (a) 实验 $P = 13/20 = 0.65$ 。
- (b) 理论 = 0.5。
- (c) 尚无不公平证据。20 翻转是一个小样本；获得 13 头完全在正常范围内。通过更多次翻转，实验概率应该接近 0.5。
- ★这是关键思想：小样本是有噪音的。强：明确提及样本量。
- 弱：“是的，这不公平”或“绝对公平”——两者都太明确了。

问题 7 — 采样偏差

- 第一批50 学生不具有代表性——步行的学生可能住得更近并且到达得更早；开车或乘坐公共汽车的学生较晚到达。
- 更好的方法：随机抽样（e.g., 随机选择50 学生ID）、分层抽样（每个年级均等抽样）或在一天中的不同时间抽样。
- ★强：指出抽样方法会产生偏差（而不仅仅是“样本很小”）。

问题 8 — 估计人口

- 样品： $45/60 = 75\%$ 。
- 估计： 75% 学生中的 $1200 = 900$ 。
- ★强：按比例缩放。弱：只给出 45 作为答案（忘记 $60 \neq 1200$ ）。

问题 9 — 有意义的差异？

- 均值差： $84 - 78 = 6$ 。
- MAD 约为4.5（5 和4 的平均值）。
- 差值(6) 大于MAD，但不大于 $2 \times \text{MAD}$ （这将是9）。
- 可能有意义，但意义不大。
- ★强：明确地将差异与变异性进行比较。弱：“是的，6 很多”，没有比较。
- IM 中的经验法则：差值 $> 2 \times \text{MAD}$ 表明有意义。

问题 10 — Han 的“被操纵的模具”

- $P(5) = 1/6$ 。预期： $10/6 \approx 1.67$ 10 卷中的五元。
- Han 获得了6——远远超出预期。有暗示性但不能证明。
- ★仅使用 10 掷骰，甚至罕见事件也会发生。需要更多的骰子才能自信地结束骰子被操纵了。
- 良好表现：出现 6+ 次五点（共掷 10 次）的概率很小，但仍有可能。既认识到结果令人意外，也认识到存在不确定性。
- 弱：“是的，骰子被操纵了”或“没有汉是愚蠢的”——两者都太极端了。

—结束—