

第 5 题 2026 · 问题 1 自我选择偏差与调查设计的改善

对应官方问题册：第 2 页（PDF 页码）。请与原题对照阅读。

本大题给出一所大学对全体 10,000 名学生实施的自愿回答式在线问卷结果，问两个百分比、一个 150 字以内的判断题和一个 200 字以内的改善方案。设问 1 的计算本身很简单，但它的全部意义在于把两个学部的回答率差异摆到台面上，设问 2、设问 3 都要以这两个数字为依据。先算再论，顺序不能颠倒。

设问 1(a) 两个学部的回答率

解答

回答率 = (回答者数) ÷ (学生数)。

$$\text{理工学部: } \frac{800}{4,000} \times 100 = 20.0 \%,$$

$$\text{人文学部: } \frac{200}{6,000} \times 100 = 3.33 \cdots \% \rightarrow 3.3 \%.$$

答 理工学部 20.0 % / 人文学部 3.3 %

解说

分母是学生数 回答率的分母是该学部的全体学生数，不是回答者总数 1,000。本题特意让学生数（4,000 对 6,000）与回答者数（800 对 200）的大小关系相反，用错分母就会得到完全相反的印象。

按指定位数处理 题目要求“小数第 2 位四舍五入，求到小数第 1 位”。20 % 应写成 20.0 %，3.33... 应写成 3.3 %。

这两个数字是全题的支点 学生数少的理工学部回答率是人文学部的约 6 倍。回答者 1,000 人中理工学部占 800 人，即八成，而理工学部只占全体学生的四成。这个失衡就是设问 2 要说明的偏差来源。

设问 1(b) 回答者中的满意率

解答

满意率 = (回答“满意”的人数) ÷ (该学部的回答者数)。

$$\text{理工学部: } \frac{750}{800} \times 100 = 93.75 \% \rightarrow 93.8 \%,$$

$$\text{人文学部: } \frac{100}{200} \times 100 = 50.0 \%.$$

答 理工学部 93.8 % / 人文学部 50.0 %

解说

分母换成了回答者数 (a) 的分母是学生数, (b) 的分母是回答者数。题目在两小问里刻意换了分母, 是为了让读者意识到“85 %”这个整体数字算的是回答者中的比例。

93.75 的进位 第 2 位是 5, 四舍五入进位得 93.8。写成 93.7 是常见的机械错误。

把两组数并排看 回答率高的学部满意率也高 (20.0 % 与 93.8 %), 回答率低的学部满意率也低 (3.3 % 与 50.0 %)。两者同向, 正是“愿意回答的人本来就更满意”这一自我选择机制的直接痕迹。

还可以算一个数 全体 10,000 名学生中, 明确回答“满意”的只有 850 人, 即 8.5 %。这个数字在设问 2 里极有说服力。

设问 2 自我选择偏差与大学判断的妥当性

解答

问卷为自愿回答, 对系统关心高、评价好的学生更容易回答, 因而回答者不是全体的缩影。实际上理工学部的回答率为 20.0 %、人文学部仅 3.3 %, 回答者 1,000 人中八成来自只占全体四成的理工学部, 且该学部的满意率高达 93.8 %。整体 85 % 是被这一偏斜拉高的结果。全体 10,000 人中确认满意的仅 850 人 (8.5 %), 未回答的 9,000 人的评价完全未知。因此“被大多数学生支持”的判断不妥当。

答 不妥当。回答率在两个学部间相差约 6 倍, 回答者被回答率高且满意率高的理工学部占据八成, 85 % 是自我选择偏差的产物; 全体中确认满意的只有 8.5 %。(150 字以内)

解说

必须“踏まえ”设问 1 的结果 题目写明“设问 1 的计算結果を踏まえ”。不引用 20.0 %、3.3 %、93.8 % 中的任何一个, 就没有满足题目条件。

8.5 % 是最强的一句 把分母换回全体学生数, 85 % 立刻变成 8.5 %。这一句用一个数字就说清了“回答者中的比例”与“全体中的比例”的差距。

偏差的方向要说出来 自我选择偏差不一定使结果偏高。本题中因为“关心高的人更易回答”, 方向是偏高; 若是投诉型调查 (不满的人才回答), 方向就相反。指出方向才算说清机制。

字数控制 150 字以内。结论 1 句、依据 2 句 (回答率的差、满意率的差)、补充 1 句 (8.5 %), 刚好。不要在这里写改善方案, 那是设问 3。

设问 3 减轻偏差、提高数据可信度的改善计划

解答

把“谁回答”从学生自己的意愿中拿走, 是全部改善的核心。

首先, 改自愿回答为按学部、学年分层的随机抽样, 对被抽中者反复催答, 使各层的回答率接近且可比; 回答仍有缺失时, 按各层的学生数加权汇总, 消除理工学部被过度代表的影响。其次, 把回答时间安排在必修课内或校内系统的登录流程中, 使回答与否不再取决于关心程度。最后, 从未回答者中再随机抽取少数进行追访, 比较回答者与未回答者的满意度, 据此估计残余偏差的大小并加以修正。

答 按学部与学年分层随机抽样并催答, 以各层学生数加权汇总; 把回答机会嵌入必修课或登录流程, 使回答与否不取决于关心程度; 再对未回答者随机追访以估计残余偏差。(200 字以内)

解说

要针对“自我选择”这一机制 增加问题数量、改进问卷措辞都不能解决本题的偏差，因为问题出在谁回答而不是问什么。对策必须落在抽样与回收上。

加权是低成本的补救 即使回答率无法拉平，只要知道各层的学生数，就可以按 4,000 : 6,000 重新加权，把整体满意率由 85 % 修正为 $0.4 \times 93.8 + 0.6 \times 50.0 \doteq 67.5$ %。写出这一步，说明读者知道偏差可以被部分校正。

未回答者追访是判断残余偏差的唯一手段 加权只能校正“已观测到的层间差”，不能校正层内部“回答者与未回答者的差”。提出追访调查，才覆盖了后一种偏差。

注意伦理与可行性 强制回答、以成绩为条件的诱导会带来新的偏差与伦理问题。写成“确保回答时间”“提供轻微谢礼”这类不改变回答内容的措施更稳妥。

第6题 2026 · 问题3 导数的定义与排队人数的概率分布

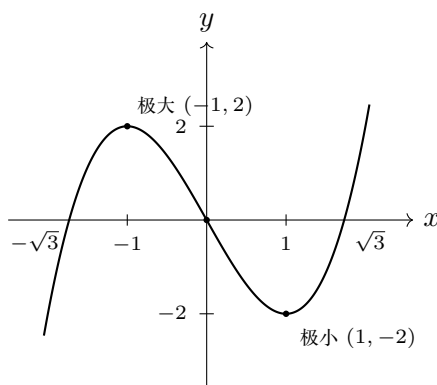
对应官方问题册：第5页（PDF页码）。请与原题对照阅读。

本大题前半考导数的定义本身：画一条三次曲线，再按定义导出 x^n 的微分，最后要求用一个具体例子说明微分在现实中的用处。后半是一道有限容量排队的概率题，由递推 $p_n = \rho p_{n-1}$ 与归一化条件出发，一路走到 $K \rightarrow \infty$ 的极限，实质是几何分布的推导。两半都要求写出过程，只写结论不给分。

设问 1(a) $f(x) = x^3 - 3x$ 的图像

解答

$f'(x) = 3x^2 - 3 = 3(x-1)(x+1)$ 。故 f 在 $x < -1$ 上递增、 $-1 < x < 1$ 上递减、 $x > 1$ 上递增，在 $x = -1$ 取极大值 $f(-1) = -1 + 3 = 2$ ，在 $x = 1$ 取极小值 $f(1) = 1 - 3 = -2$ 。 $f(-x) = -f(x)$ ，图像关于原点对称； $f(x) = x(x^2 - 3) = 0$ 给出与 x 轴的交点 $x = 0, \pm\sqrt{3}$ 。 $f''(x) = 6x$ ，故 $x = 0$ 是拐点。



答 如上图。极大 $(-1, 2)$ 、极小 $(1, -2)$ 、拐点与原点重合，与 x 轴交于 $0, \pm\sqrt{3}$ 。

解说

画图题的评分点是标注 曲线画得好看不加分，标出极值坐标、 x 轴交点与对称性才加分。先做增减表，再把表上的每一行落到图上的一个标记。

三个交点的求法 $x^3 - 3x = x(x^2 - 3)$ ，一眼可分解。若漏掉 $\pm\sqrt{3}$ ，图像与 x 轴的关系就无法确定。

奇函数性质省一半工夫 确认 $f(-x) = -f(x)$ 后，只需分析 $x \geq 0$ 的一侧，另一侧按原点对称画出即可。

与设问 9 的呼应 2025 年度设问 9 处理的正是同族的三次曲线 $y = x^3 - 6x^2 + a$ 。两年都要求先做增减表，可见这是社会工学类反复考的基本功。

设问 1(b) 按定义导出 x^n 的微分系数

解答

$f(x) = x^n$ (n 为自然数)。由因式分解

$$x^n - x_0^n = (x - x_0)(x^{n-1} + x^{n-2}x_0 + \cdots + x x_0^{n-2} + x_0^{n-1})$$

(右侧括号内共 n 项), 当 $x \neq x_0$ 时

$$\frac{f(x) - f(x_0)}{x - x_0} = x^{n-1} + x^{n-2}x_0 + \cdots + x_0^{n-1} = \sum_{k=0}^{n-1} x^{n-1-k} x_0^k.$$

右侧是 x 的多项式, 故可直接令 $x \rightarrow x_0$ 。每一项都趋于 x_0^{n-1} , 共 n 项, 因此

$$f'(x_0) = \lim_{x \rightarrow x_0} \frac{x^n - x_0^n}{x - x_0} = n x_0^{n-1}.$$

答 $f'(x_0) = n x_0^{n-1}$

解说

题目禁止用公式 写 “ $(x^n)' = nx^{n-1}$, 所以 $f'(x_0) = nx_0^{n-1}$ ” 是循环论证, 得零分。必须从题给的定义式出发, 先把 $x - x_0$ 约掉。

因式分解要验证项数 括号内是 $x^{n-1-k}x_0^k$ ($k = 0, \dots, n-1$), 恰好 n 项。项数就是最终系数 n 的来源, 答题时应把 “共 n 项” 写出来。

也可用二项式定理 令 $x = x_0 + h$, 则 $\frac{(x_0 + h)^n - x_0^n}{h} = nx_0^{n-1} + \binom{n}{2}x_0^{n-2}h + \cdots$, $h \rightarrow 0$ 时同样得到 nx_0^{n-1} 。两种写法都是正解, 因式分解法不必展开二项系数, 更简短。

极限可以直接代入的理由 约分后得到的是多项式, 多项式在每一点都连续, 所以 $x \rightarrow x_0$ 可以直接代值。写一句 “右侧为多项式, 故可直接代入” 能让论证严密。

检验 $n = 1$ 给出 $f' = 1$, $n = 2$ 给出 $f' = 2x_0$, 与已知一致。

设问 1(c) 微分在现实社会中的用处

解答

微分给出的是 “每增加一单位投入, 结果变化多少” 这一瞬时变化率, 因此可以用来找出使某个量最优的点。

例如, 某商品定价为 p 时的需求量为 $q(p)$, 则销售额为 $R(p) = pq(p)$ 。把 R 对 p 微分并解 $R'(p) = 0$, 就能求出使销售额最大的价格; $R'(p) > 0$ 的区间意味着还可以提价, $R'(p) < 0$ 的区间意味着提价反而减收。仅靠若干个价位的试卖只能得到离散的比较, 而微分给出应向哪个方向、以多大幅度调整的判断依据。

答 微分给出瞬时变化率, 可用于求最优点。例如把销售额对价格微分并令导数为零, 即可求出使销售额最大的价格, 并判断当前应提价还是降价。

解说

题目要求 “举一个具体例” 只写 “可用于最优化” 不给分。要指定一个量 (销售额)、一个变量 (价格), 并说明导数为零意味着什么。

其他可用的例子 物理的速度与加速度、经济学的边际成本与边际收益、人口或感染者数的增长速度、机器学习沿梯度更新参数。社会工学类的题材以经济与都市为主, 选经济类的例子最贴合。

“瞬时” 是微分区别于差分的地方 用两点之差算出的平均变化率随所取区间而变, 而导数是区间趋于零时的极限, 与取法无关。写出这一点能显示对定义的理解。

设问 2(a) 用 p_0 与 ρ 表示 p_n **解答**由 $p_n = \rho p_{n-1}$ ($n = 1, 2, \dots, K$) 反复代入:

$$p_n = \rho p_{n-1} = \rho^2 p_{n-2} = \cdots = \rho^n p_0.$$

答 $p_n = \rho^n p_0$ ($n = 1, 2, \dots, K$)**解说**

这是等比数列 $\{p_n\}$ 是首项 p_0 、公比 ρ 的等比数列。认出这一点后, (b) 的归一化条件就变成等比数列求和。

ρ 的含义 $\rho = p_n/p_{n-1}$ 是“队列人数再多一人”的相对概率。 $0 < \rho < 1$ 保证人数越多概率越小, 符合直觉。

范围包含 $n = 0$ $n = 0$ 时 $\rho^0 p_0 = p_0$, 公式自动成立, 所以下一问求和时可以直接从 $n = 0$ 写到 $n = K$ 。

设问 2(b) 用 ρ 与 K 表示 p_0 **解答**把 (a) 代入归一化条件 $\sum_{n=0}^K p_n = 1$:

$$\sum_{n=0}^K \rho^n p_0 = p_0 \cdot \frac{1 - \rho^{K+1}}{1 - \rho} = 1 \quad (0 < \rho < 1 \text{ 故 } \rho \neq 1),$$

$$p_0 = \frac{1 - \rho}{1 - \rho^{K+1}}.$$

答 $p_0 = \frac{1 - \rho}{1 - \rho^{K+1}}$ **解说**

项数是 $K + 1$ 而不是 K n 从 0 取到 K , 共 $K + 1$ 项, 故等比数列和的指数是 $K + 1$ 。这一处的差错会一直传到 (c)。

$\rho \neq 1$ 由题设保证 题目给出 $0 < \rho < 1$, 可以直接用等比数列求和公式。若允许 $\rho = 1$, 则各 p_n 相等, $p_0 = 1/(K + 1)$, 需另行讨论。

检验 $K = 0$ (队列最多 0 人) 时 $p_0 = \frac{1 - \rho}{1 - \rho} = 1$, 符合“必然为 0 人”。 $K = 1$ 时 $p_0 = \frac{1 - \rho}{1 - \rho^2} = \frac{1}{1 + \rho}$, $p_1 = \frac{\rho}{1 + \rho}$, 两者之和为 1。

设问 2(c) $K \rightarrow \infty$ 时 p_0 的极限 π_0 **解答** $0 < \rho < 1$ 故 $\rho^{K+1} \rightarrow 0$ ($K \rightarrow \infty$), 因此

$$\pi_0 = \lim_{K \rightarrow \infty} \frac{1 - \rho}{1 - \rho^{K+1}} = \frac{1 - \rho}{1 - 0} = 1 - \rho.$$

答 $\pi_0 = 1 - \rho$ **解说**

条件 $0 < \rho < 1$ 在这里才真正起作用 若 $\rho \geq 1$, ρ^{K+1} 不趋于 0, 极限不存在 (且原级数发散)。题目把 $0 < \rho < 1$ 写进条件, 正是为了保证这一步成立。

结果的形式极简 K 完全消失, 只剩 ρ 。这说明当容量足够大时, 队列长度的分布不再依赖容量上限, 是本小问最值得注意的地方。

与几何分布的关系 此时 $p_n \rightarrow (1 - \rho)\rho^n$, 正是以 $1 - \rho$ 为参数的几何分布。在排队论中 ρ 称为利用率, 这一结果就是 M/M/1 队列的稳态分布。

设问 2(d) π_0 在调查队列人数分布时的用处**解答**

$\pi_0 = 1 - \rho$ 是“队列为空”的概率, 因此 $\rho = 1 - \pi_0$ 可以由观测直接估计: 在随机时点反复观察收银台, 记录无人排队的时间比例, 就得到 π_0 , 进而得到 ρ 。

一旦 ρ 确定, 整个分布 $p_n = (1 - \rho)\rho^n$ 随之确定, 不必逐个统计“恰好 n 人”的频率, 也不必知道容量 K 。由此还可以算出平均排队人数

$$E[X] = \sum_{n=0}^{\infty} n(1 - \rho)\rho^n = \frac{\rho}{1 - \rho},$$

并看出 $\rho \rightarrow 1$ 时 $\pi_0 \rightarrow 0$, $E[X] \rightarrow \infty$, 即利用率接近 1 时队列会急剧变长。这为增设收银台的判断提供了定量依据。

答 π_0 是队列为空的概率, 由 $\rho = 1 - \pi_0$ 可从观测直接估计 ρ , 从而确定整个分布 $p_n = (1 - \rho)\rho^n$ 与平均人数 $\rho/(1 - \rho)$, 并判断利用率接近 1 时的混杂风险。

解说

关键是“一个可观测量决定整个分布” 分布本来含 ρ 与 K 两个参数, $K \rightarrow \infty$ 后只剩 ρ , 而 ρ 又可由最容易观测的量 (空闲比例) 得到。这条链路就是本小问要求的“怎样有用”。

空闲比例是最容易采集的数据 数“现在排了几个人”需要在每个时点计数, 而判断“现在有没有人”只需一个是或否, 观测成本低得多, 误差也小。

$E[X]$ 的求法 $\sum n\rho^n = \frac{\rho}{(1 - \rho)^2}$, 乘以 $(1 - \rho)$ 即得 $\frac{\rho}{1 - \rho}$ 。不写这一步也可以, 但给出平均人数会让“有什么用”更具体。

注意近似的前提 $\pi_0 = 1 - \rho$ 只在 K 足够大时成立。若收银台前的空间实际很小 (K 为 3 或 4), 就必须

用 (b) 的有限 K 公式，否则会低估空闲概率。写出这一限制能显示论证的分寸。