回归结果的解释边界：两个案例分析任务

作答说明

请独立完成。我们关注的是您的推理过程，不只是最终结论。对于每一项判断，请尽量说明理由。除非题目明确给出，否则不要假定研究者收集了其他信息。

案例 A：葡萄酒消费与心脏病死亡率

研究者收集了 20 个国家的资料。每个国家的数据包括：

- 成年人年均葡萄酒消费量；
- 冠心病年死亡率；
- 人均国内生产总值；
- 平均吸烟率。

以“国家”为观测单位建立多元线性回归模型。在控制人均国内生产总值和平均吸烟率后，葡萄酒消费量每增加 1 升/人年，冠心病年死亡率平均降低 3.2 例/10 万人，95%置信区间为−5.1 至−1.3。

研究者据此写道：

增加葡萄酒消费能够降低一个人患冠心病的风险，因此成年人可以通过适量增加葡萄酒摄入保护心脏。

请回答：

1. 这项研究中的观测单位是什么？研究者最终结论所谈论的单位又是什么？二者是否一致？
2. 回归结果直接支持的是以下哪类结论？请选择并说明理由：
    - 描述性结论
    - 关联性结论
    - 预测性结论
    - 因果性结论
3. 统计显著的负回归系数是否足以证明“增加葡萄酒消费会降低个人患病风险”？为什么？
4. 除人均国内生产总值和吸烟率外，请提出至少三个可能同时影响葡萄酒消费和冠心病死亡率的因素或机制。
5. 即使把这些因素加入回归模型，是否就一定能够得到因果结论？请说明。
6. 该研究可能存在哪些问题？可从以下方面考虑，但不限于这些方面：
    - 混杂变量；
    - 反向因果；

- 测量误差；
- 国家平均值与个人情况的差异；
- 国家之间的医疗、饮食或生活方式差异；
- 模型设定和样本量。

7. 如果研究者希望更可靠地回答因果问题，还需要什么数据、研究设计或分析方法？
8. 请把原研究者的结论改写成一段不超过 80 字、证据强度适当、可向公众发布的表述。
9. 如果这项结论被媒体或健康机构直接用来建议公众饮酒，可能产生哪些后果？统计分析者在发布结论前负有什么责任？

案例 B：在线练习平台与统计课程通过率

研究者收集了 42 所大学一年级统计课程的资料。每所大学的数据包括：

- 学生平均每周使用在线练习平台的时间；
- 课程总体通过率；
- 学生平均入学成绩；
- 平均班级人数。

以“大学”为观测单位建立多元线性回归模型。在控制平均入学成绩和平均班级人数后，学校平均平台使用时间每增加 30 分钟，其统计课程通过率平均提高 2.4 个百分点，95%置信区间为 0.8 至 4.0 个百分点。

研究者据此写道：

要求每名学生每周多使用平台 30 分钟，会提高该学生通过统计课程的概率。

请回答：

1. 这项研究中的观测单位是什么？研究者最终结论所谈论的单位又是什么？二者是否一致？
2. 回归结果直接支持的是描述性、关联性、预测性还是因果性结论？请说明理由。
3. 统计显著的正回归系数是否足以证明强制增加平台使用时间会提高个人通过概率？为什么？
4. 请提出至少三个可能同时影响平台平均使用时间和课程通过率的因素或机制。
5. 把这些因素加入回归模型后，是否就一定能够得到因果结论？
6. 该研究可能存在哪些混杂、选择、测量、层级或模型设定问题？
7. 如果希望判断平台使用本身是否提高学习成绩，应怎样改进研究设计？
8. 请把研究者的结论改写成一段不超过 80 字、证据强度适当的表述。
9. 比较案例 A 和案例 B：两者最重要的共同推理问题是什么？您在案例 A 中形成的判断原则，哪些可以迁移到案例 B？