实验 5：指标观测与最小复现实验

1. 本节定位

这是实验线的收束课。目标不是追求漂亮结果，而是训练你把前四节实验中的代码定位、日志、表格和解释整理成一份合格的最小复现实验包。

2. 核心问题

怎样把一次课堂运行整理成别人可以复查、可以复跑、也可以否定的实验记录?

3. 学习目标

完成本节后，你应能够:

- 定义一个可否定的小型系统问题。
- 设计基线与对照组。
- 形成完整的运行记录、结果表和解释文档。
- 清楚说明“观察到现象”和“证明机制收益”的边界。

4. 开始前

4.1 先修要求

- 已完成前四节实验。
- 手头已有代码地图、阶段日志、batch 日志或状态观测表。
- 已了解基线、对照组、warmup、重复实验和结论边界。

4.2 本节材料

- 理论配套：build/tutorials/Tutorial_10_实验方法与验证.pdf
- 参考代码：src/code/nano-vllm-hust/bench.py
- 学生练习：src/code/nano-vllm-hust/student_exercises/bench.py
- 作业包：src/experiments/nano-vLLM实验课/experiment_5_指标观测与最小复现实验/assignment_spring-2027/
- 结果汇总脚本：src/code/nano-vllm-hust/scripts/teaching/collect_benchmark_results.py

5. 提交内容

本节提交物必须包含:

1. 一份最小复现实验说明。

2. 一张结果表。
3. 一段 500-800 字分析，写清结论边界。
4. 如有需要，附原始日志或截图。

6. 时间安排建议

- 10 分钟：说明最小复现实验包结构。
- 20 分钟：确定本次要验证的小问题。
- 25 分钟：运行基线与对照。
- 20 分钟：整理结果表和异常记录。
- 10 分钟：写结论边界。

7. 实验任务

任务 1：提出可证伪问题

从前四节实验中选一个小问题，例如：

- 长 prompt 是否显著抬高 TTFT。
- 某个 batch limit 是否改善吞吐但恶化短请求等待。
- 上下文变长是否通过状态驻留抬高 queue wait。

问题必须具体、可检验，不能写成“看看系统表现”。

任务 2：设计基线与对照

至少包含：

- 一个基线
- 一个改动配置或改动策略
- 一组固定 workload
- 三次重复运行或等价稳态统计

任务 3：确定实验记录结构

实验包至少应包含：

- `README.md`：实验问题、命令、环境说明
- `results.csv`：核心指标
- `notes.md`：异常、失败、解释
- 原始日志摘录或截图

任务 4：运行并记录结果

请填写下表：

experiment	workload	policy/config	run_id	TTFT	throughput	queue_wait	notes

任务 5：写出边界说明

每份报告必须明确写出：

- 哪个结果支持当前解释。
- 哪个结果会直接推翻当前解释。
- 哪个变量没有被完全控制住。

8. 证据要求

以下情况不能视为合格证据：

- 只有结果表，没有运行命令。
- 只有最好的一次结果，没有对照或重复。
- 只写“性能提升了”，没有写边界与失败条件。
- 把观察结果直接写成机制已经被证明。

9. 提交核对

- 已附最小复现实验说明。
- 已附结果表与运行命令。
- 已附边界说明。
- 已说明异常值或失败运行如何处理。

10. 评分关注点

- 记录是否完整。
- baseline 与对照是否清楚。
- 是否有诚实的边界说明。
- 结果是否回到控制点和系统对象。

11. 与后续实验的衔接

本节提交物应直接进入课程项目开题或结题材料，而不是另起一套格式。建议在课程项目第一页引用本节实验编号和结果表编号。

12. 提交模板

12.1 基本信息

姓名 / 组员:
学号:
实验日期:
模型与环境:

12.2 可证伪问题

我要验证的问题:
哪个结果支持它:
哪个结果会推翻它:

12.3 实验记录结构

`README.md`:
`results.csv`:
`notes.md`:
原始日志 / 截图:

12.4 结果表

experiment	workload	policy/config	run_id	TTFT	throughput	queue_wait	notes

12.5 边界说明

支持当前解释的结果:
会直接推翻当前解释的结果:
未完全控制住的变量:

12.6 总结说明

请用 500-800 字整理：你的基线、对照、异常值与结论边界分别是什么?