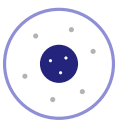
Collecting Data

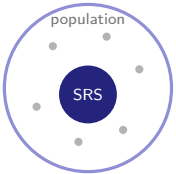
AP Statistics ▪ Unit 3

AP Statistics



What we will cover

1. Trusting the data
2. Observational vs experiment
3. Random sampling
4. When sampling goes wrong
5. Designing experiments
6. Scope of inference



Trusting the data

The *method* of collection decides what we may claim. We study a **sample** 样本 to learn about a **population** 总体.

- **bias** 偏差 tilts a sample systematically off the truth

> Good data comes first – a clever calculation cannot rescue a badly collected sample.

clinical trial lab

Observational vs experiment

- an **observational study** 观察性研究 watches – shows only **association** 关联
- an **experiment** 实验 applies a **treatment** 处理 – can prove **cause** 因果关系

A **confounding variable** 混杂变量 tied to both explanatory and response is why observation can't claim cause.

random dice

Random sampling

Let chance, not the researcher, choose – four random designs:

SRS 简单随机抽样: every group of n equally likely

Stratified 分层: SRS within similar **strata** 层

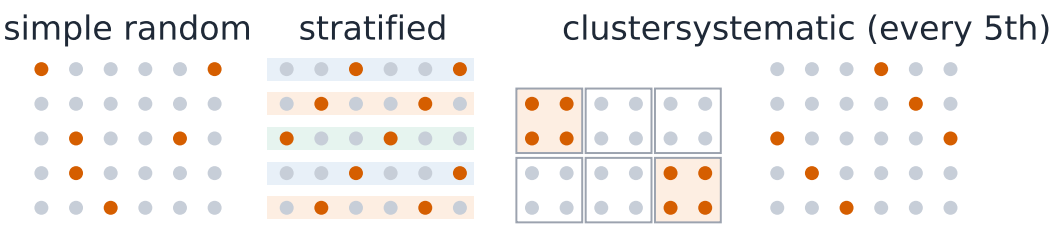
Cluster 整群: pick whole **clusters** 群

Systematic: every kth individual

sampling bias crowd

Random selection lets us **generalize** 推广 the sample result to the whole population.

Random sampling

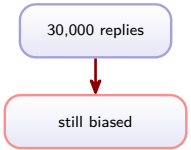


When sampling goes wrong

Even a "random" plan can bias:

- **undercoverage** 覆盖不足: groups left out
- **voluntary response** 自愿回应: only strong feelings
- **nonresponse** 无回应, **response bias** 回应偏差, wording

> A bigger sample does not fix bias – it just makes a bigger biased sample. Method beats size.



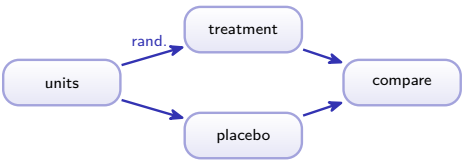
SRS of 1000 beats it

Designing experiments

Three principles:

- **control** 控制: hold other variables equal
- **randomization** 随机化: balances unknowns
- **replication** 重复: many units per treatment

> A **placebo** 安慰剂 + **double-blind** 双盲 design separate the real effect from belief.

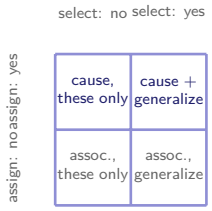


Scope of inference

Two uses of randomness set what you may claim:

- random **assignment** 随机分配 ⇒ cause
- random **selection** 随机选择 ⇒ generalize

> **Blocking** 区组化 / **matched pairs** 配对设计 remove a known source of variability.



Worked example – spotting the bias

A school emails a survey to all students; 80 of 1000 reply, and 70 say they want a later start. Identify the flaw.

- Everyone could reply, so this is a voluntary response sample, not a random one.
- Only students who care strongly bother – and those wanting a later start care most.
- So the 87.5% is biased upward; you cannot generalise to all 1000.
- Fix: take a simple random sample and follow up non-responders $\Rightarrow$ a representative estimate.

Name the bias and say which direction it pushes the estimate – that is where the marks are. A big sample does not fix bias; only good selection does.

Unit 3 – key takeaways

1 Study type

observation shows association; experiment proves cause

2 Sampling

SRS, stratified, cluster, systematic; bias $\neq$ fixed by size

3 Experiments

control, randomize, replicate; placebo + blinding

4 Scope

assignment $\rightarrow$ cause; selection $\rightarrow$ generalize

Check yourself

1. Which study type can prove cause and effect?
2. What sampling method takes an SRS within each stratum?
3. Does a larger sample remove bias?
4. What must be random to generalize to a population?

cause	both
assoc.	gen.

Answers **1.** a randomized experiment **2.** stratified **3.** no **4.** random selection