Unit 2 · Sampling and Experimentation Cheat Sheet

Unit 2 focus: populations and samples, random sampling methods and sources of bias

1 KEY VOCABULARY

population	The entire group we want information about.	all 1,200 students at a school
sample	The part of the population we actually measure.	80 students surveyed
census	Data from every member of the population.	asking all 1,200
parameter	A number describing the population (usually unknown).	p, μ
statistic	A number computed from a sample.	$\hat{p}$, $\bar{x}$

2 SAMPLING METHODS

Method	How it works
Simple random (SRS)	every group of n is equally likely to be chosen
Stratified	split into similar groups (strata); take an SRS from *each*
Cluster	split into groups that each look like the population; take *all* of a few groups
Systematic	random start, then every kth individual
Convenience / voluntary	easy-to-reach or self-selected people: biased

3 CHOOSING AN SRS WITH RANDOM DIGITS

Label every individual with the same number of digits, read pairs from the table, skip out-of-range numbers and repeats.

Example: Pick 3 of 40 students (label 01–40). Digits: 07 51 40 13 07 22
Read pairs: 07 ✔, 51 skip (> 40), 40 ✔, 13 ✔
Stop: we have 3 students

Answer: students 07, 40 and 13

4 STRATIFIED OR CLUSTER?

Survey 20 freshmen, 20 sophomores, 20 juniors, 20 seniors (random within each grade)	**stratified**
Randomly choose 4 homerooms and survey everyone in them	**cluster**
Number all students and pick 80 with a random generator	**SRS**
Survey every 15th student entering, starting at a random one of the first 15	**systematic**

5 SOURCES OF BIAS

Bias	Example
Undercoverage: some groups left out of the frame	phone survey misses people with no phone
Nonresponse: chosen people don't answer	only 30% return a mailed survey
Response bias: answers are not truthful	asking about cheating face to face
Wording: the question pushes an answer	"Don't you agree that …?"

A bigger sample does **not** fix bias; random selection does.

6 QUICK REMINDERS & COMMON MISTAKES

✔ Do remember

- Name the population and the sample in context.
- Describe the random method well enough to repeat it.
- Say the *direction* of bias (over- or underestimate).

✘ Common mistakes

- "Random" = haphazard: use a random number generator or table
- Stratified = cluster: strata are similar inside; clusters are mini-populations
- Large sample removes bias: it only reduces variability

Unit 2 · Sampling and Experimentation Cheat Sheet

More practice: observational studies, experiments, experimental designs and the scope of inference

1 OBSERVATIONAL STUDY OR EXPERIMENT?

Type	What happens
Observational study	observe and measure; no treatment is imposed. Shows association only.
Experiment	deliberately impose treatments on units. With random assignment it can show **cause and effect**.

Confounding: two variables whose effects on the response cannot be separated (e.g. coffee drinkers also smoke more).

2 EXPERIMENT VOCABULARY

experimental units	Who or what receives the treatments.	60 tomato plants
factor / levels	Explanatory variable / its values.	fertilizer: 0, 5, 10 g
treatment	A specific condition applied.	5 g of fertilizer
response variable	The outcome measured.	plant height
placebo	A fake treatment with no active ingredient.	sugar pill

3 PRINCIPLES OF EXPERIMENTAL DESIGN

Comparison	compare two or more treatments (often a control group)
Random assignment	assign treatments by chance so groups are alike
Control	keep other variables the same for every group
Replication	use enough units in each group

Blinding: single-blind (subjects don't know), double-blind (subjects *and* evaluators don't know).

4 DESIGNS

Design	How units get treatments
Completely randomized	all units randomly assigned to treatments
Randomized block	group similar units into blocks (e.g. by gender), then randomize within each block
Matched pairs	pairs of similar units, or each unit gets both treatments in random order

Block to reduce variability from a variable you *know* affects the response.

5 COUNTING TREATMENTS

Number of treatments = product of the number of levels of each factor.

Example: 3 doses × 2 times of day; 60 volunteers
Treatments: $3 \times 2 = 6$
Per group: $60 \div 6 = 10$
Assign: number 1–60, randomly pick 10 for each treatment

Answer: 6 treatments, 10 volunteers each

6 SCOPE OF INFERENCE

	Random assignment	No random assignment
Random sample	cause and effect; generalize to population	association only; generalize
Not random sample	cause and effect for these subjects only	association, these subjects only

✘ Correlation is not causation. ✔ Random *sampling* → generalize; random *assignment* → causation.