Housekeeping.

- Writing assignment due 11:59 p.m. today ON CANVAS
- Written assignment due at beginning of class on Wednesday (hard copy, not Canvas)
- Quiz on Wednesday   • Tuesday study 5-6 p.m. B205
- Book project: choose from among:

    - The Signal & the Noise
    - Moneyball
    - Dataclysm
    - How not to be Wrong

See links on Canvas for descriptions of each.
We will have our first book project assignment ~~well be likely~~ due on Monday, Feb. 13 — so get your chosen book soon!

Last time :   • Simple random sampling
              • ~~Cluster~~ sampling
                stratified

QUESTIONS?

- **Cluster sampling:** Again, divide the population into groups (this time, "groups" are called "clusters"). Randomly select a ~~group~~ cluster, and put **all** its members into the sample. Repeat until you have as many sample members as you wanted.

Works best when each group is a <u>microcosm</u> of the whole — the opposite of strata! (Strata are usually <u>homogeneous</u>: within each stratum, the members share some feature — and you want <u>diversity</u> in the sample, so you should choose from different strata!) With clusters, the groups are <u>heterogeneous</u>! that is, within each cluster, the members are usually different from each other ("as different", on avg., as people in the whole population). That's why it's safe to take an entire cluster for the sample.

Examples.

- You want a sample of 4 from our class of 26. It's probably enough (?) to ~~choose~~ have the students sit in rows of 4, and just choose one of those rows. (What could go wrong here ??)

- To get a sample of 5,000 employees in ice cream shops, just pick a few ice cream shops and put **all** their employees in the sample, until you reach 5,000 employees.

· Cluster sampling is typically used when the population is very large, ~~when~~ when it "organically" clusters itself, and when your resources are limited (for example, you might be unable to sample from different areas in the city, so you just choose your neighborhood & hope for the best).

Q. Why does stratified sampling require more resources than cluster sampling does?

Example 1.12 :

What type of sampling was used?

(a) Coach takes 6 players from among boys aged 8-10; 7 players from 11-12, & 3 players from 13-14.

(b) HS admin interviews 50 female HS teachers & 50 male HS teachers.

(c) Medical rsrchr. interviews every 3rd cancer patient from a list at the local hospital.

(e) A student interviews 5 of his classmates to determine how many pairs of shoes students own on average.

Systematic random sampling.

Randomly select a starting point, and take every nth entry from the population $(n = \frac{\text{population size}}{\text{desired sample size}})$.

Example. Your phone book contains 20,000 listings, and you want to sample 400. Well, $\frac{20,000}{400} = \frac{200}{4} = \frac{100}{2} = 50$, so start with the 1st entry & choose every 50th entry thereafter — this will give you a sample of 400.

Example 1.11 (b) is systematic r.s.

Convenience / Opportunity Sampling.

Recall example from Cartoon Guide: the author of a book sent 100,000 questionnaires to randomly chosen women, and got responses from only 4.5% of those who were sent the questionnaire. The author used those responses anyway.

This is opportunity sampling or convenience sampling, which USES RESULTS THAT ARE READILY AVAILABLE.

Sometimes you get good results w/ convenience sample — sometimes not.

Sampling with/without replacement.

WITH REPLACEMENT: Once a member/number is picked, (he/she/it) goes back into the population, and might be picked again.

WITHOUT REPLACEMENT: A member of the population may be chosen only once.

Example. If a college has 10,000 undergraduates, and you're picking 1,000 for a simple random sample, then...

WITHOUT REPL.
$\left\{\begin{array}{l}\end{array}\right.$
- The first student t/b selected has a $\frac{1}{10,000}$ chance of being selected.

- The second has $\frac{1}{9,999}$

  $\vdots$

- The thousandth student has $\frac{1}{9,000}$ chance of being selected.

WITH REPL. $\left\{\begin{array}{l}\end{array}\right.$ • Each student has $\frac{1}{10,000}$ chance of being selected.