Cht 1 Data Types & Sampling
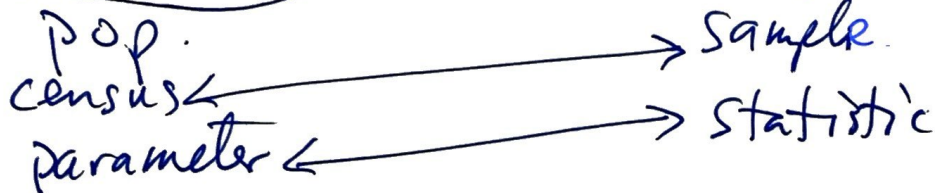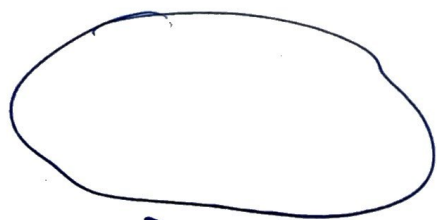
1.1 Introduction to Sampling

Def: Statistics is the study of procedures for collecting, describing and then infering information from a sample onto the population from where the sample came from.

- **population**: the entire collection of objects under study.

- **parameters**: numbers used to describe the population's attributes. These attributes are collected from all members in the popul'n.

- **Sample**: a subset of the population

- **statistic**: a number describing an attribute from the sample

POP.

census

parameter

A sample

→ Sample.

→ Statistic

Describing a population from data aquired from only a sample will highly depend on the quality of the sample itself.
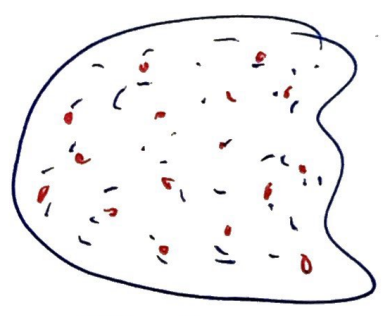

EX

1.5 million tax payers : census ⇒ everyone

→ parameter : % of those who want the wealthy to pay more taxes

Boston city

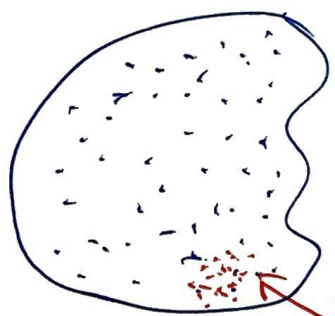Bay

subject

Determine the city's sentiment on taxation

Too expensive to conduct a census, so we need to sample & describe and then infere onto the pop.
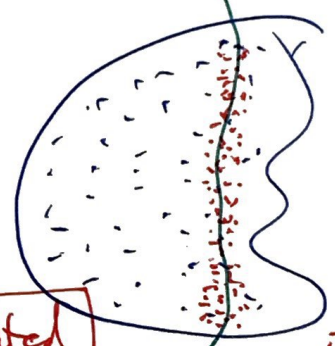
---

* Sampling

I 95

Simple Random Sample

Best.

SRS

Convienience Sampling
{ close by my house }

Gated Community

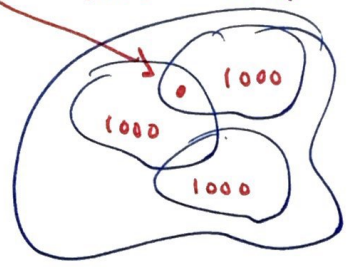Worst

Cluster Sampling

"regional survey"

almost the worst

• Sample size : n or N

• **SRS:** Here every group of "n" objects in your population has equal chance of being selected in a survey an any other object in a group of size "n".

ex let n = 1000

SRS : Any individual object in a collection of n objects has equal chance of being selected in any other group of n objects.

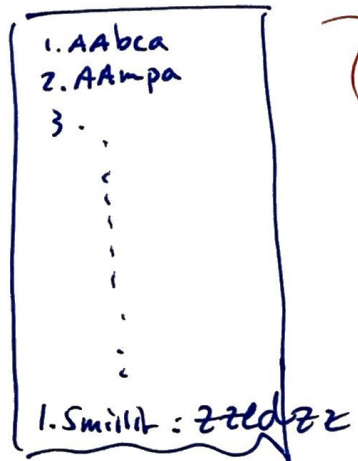not in the third group.

ended up in both sample.

• = sampled individual

• = another's sample of individuals

⊛ How might we execute such sampling?
→ Let's list from the Boston City Tax department all taxpayers from 1 to 1.5 million.

" List or
Frame

| 1. AAbca |
| 2. AAmpa |
| 3. . |
| ⋮ |
| 1. Smillit : ZZZdZZ |

Randomly select 1000 individuals and contact them to execute a survey with them.

To randomly select from the 1.5 million we would use a <u>Random Number Generator</u> such as random.org.

[EX] From a list of numbers from 1 to 100 randomly select 20.   <u>random.org</u>

    random.org → integer sets { no duplicate #'s}

        → random integer {allows duplicates}

[EX] <u>Book page numbering</u> : $n = 13$ between $\underset{00}{\uparrow}$ to $\underline{99}$. ↑ 100

    Flip through the pages and record the last two integers

    p 3 05   p 1 27   p 4 27   p 2 63

    p 1 17   p 53   p 11   p 4 09

    p 3 47   p 2 89   p 61   p 4 87   p 3 91

| 05, 27, 27, 63, 17, 53, 11, 09, 47, 89, 61, 87, 91 |

EX We have 300 employees @ ACME inc.
we desire to poll 10 of them and ask if they have
a personal, non-corporate, IRA.

```
1. Andrew E.
2. Andres J.
3. Barry O.
4.
   ⋮
299. Xiotle R.
300 zus. M.
```

Fram of
ACME
employees

• take the top 10 : **Convenient Sampling.**

$1, 2, 3, 4, 5, 6, 7, 8, 9, 10$ Sample

pro: cheap
con: not random

• survey every 30th

$30, 60, 90, 120, 150, 180, 210,$
$240, 270, 30$

Systematic Sample

pro: low cost
con: not random

Ⓧ Lets sort the list (frame) according to Salary.

strata
$\begin{cases}
0 \to 30,000 \$ & \to \text{List } 0 \\
31,000 \to 90,000 & \to \text{List } 1 \\
91,000 \to 120,000k & \to \text{List } 2 \\
121,000 \to \text{up.} & \to \text{List } 3
\end{cases}$

Now in each of these select, in a round robin fashion
two individuals, 4*2 = 8 then pick an additi
2 pple, one each from the lowest two groups

Stratefied Sampling

pro: representative across
income
con: not truely random.

- **Voluntary Sampling** erect a table outside the main entrance and post a sign that says "do you have an IRA?" If so come see us. You record their answer.
  - pro: cheap
  - con: voluntary bias response.

- **SRS** use Random.org and ask that it randomly selects 10 numbers from a list of 1 to 300. Then you contact those 10 selected & poll them.
  - pro: best random sample
  - con: more expensive.

- Notes posted on mymathmantra.com/math
  → Fall 2024

- will record 1.2 Friday.

  1.1 & 1.2 HW due Monday on $8\frac{1}{2} \times 11$" white paper

  - two stack.