Let's do a "survey". Who has an iPhone? 13 of 20 people.

x_1 = 0, x_2 = 0, x_3 = 1, ..., x_20 = 1

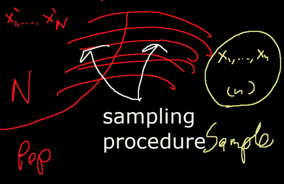
data — survey element number — values (in this case 1 = yes, 0 = no)

Do we believe this survey has a uniformly random "sample" of n = 20 elements from a superset called the "population"? If so, what is the population? Let's say yes, there is a population. This assumption is the "population model sampling assumption".

Is the population...
* all people on Earth? NO
* all people in America? NO
* all college students in NYC? NO
* students and faculty of QC? Maybe...

It's tough to define precisely the population once you have a sample. The more classical situation is you first define the population and then sample from it (how to sample you'll see on the homework).

The population size is N and the sample size is n and n << N. If the sample size is all Americans, N = 333 million people so n << N.



Can we use the sample data to tell us "something" about the population? Hopefully yes. This is called "inference". The sample data is used to "infer" properties about the population. Numeric properties of the population are called "population parameters".

"Infer" means to make an educated guess from the particular --> the universal AKA "induction". "Deduction" means to use logic usually from universal --> particular. Induction is difficult. **You never really know you're right.**

[deduction] You know all swans are white --> any 5 swans are white
contrasted to...
[induction] You observe 5 white swans --> all swans are white
Is your deduction correct? Yes. Is your induction correct? Maybe.

How is inference done on samples? First you compute "statistics" which are functions of the data:

$$\hat{\hat{\theta}} = w(x_1, x_2, \ldots, x_n)$$

where thetahathat is a scalar number e.g. the "average" / "sample mean"

$$\hat{\hat{\theta}} = \bar{x} = w(\vec{x}) = \frac{1}{n}\sum_{i=1}^{n} x_i \overset{\text{our survey}}{=} \frac{1}{20}(13) = 0.65$$

$= \hat{p}$ If the survey has binary values (Bernoulli), then we call the average "p-hat" or "sample proportion".

What can you "infer" from using this statistic? Usually, an unknown parameter which we denote θ. In our class example maybe:

$$\theta := \frac{X}{N}$$

X ← total number of iPhones among population
N ← size of the population

The values of $\theta \in \Theta$, the "parameter space" in our case:

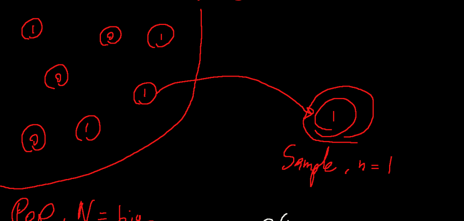
$$\Theta = \left\{0, \frac{1}{N}, \frac{2}{N}, \ldots, \frac{N-1}{N}, 1\right\}$$

Convention: is Greek letters are "unknowable" parameters / quantities and English are knowable / computed quantities.

$\hat{\hat{\theta}}$ in the case of sample proportion or average is a "point estimate" for θ. "Point" means one single numeric value best guess for θ where θ is a single numeric value itself.

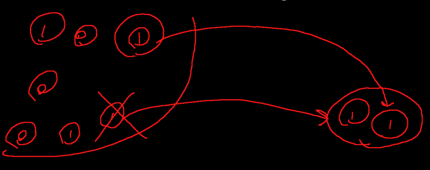
"Statistical Inference" is using statistics to make inference. There are three main goals:
(1) Point estimation
(2) Confidence set creation: give me a reasonale set of values for the value of θ.
(3) Theory testing (testing a theory about the true value of θ)

Let's discuss sampling a bit more. Let's let n = 1.



Each element should be sampled "at random". Really, uniformly sampled i.e. the probability of each population element is 1/N

$$P(X_1 = x_1 = 1) = \frac{X}{N} = \theta$$



For n = 2, on the second sampling, we can't get the first element.

$$P(X_2 = 1 \mid X_1 = 1) = \frac{X-1}{N-1} < \theta$$