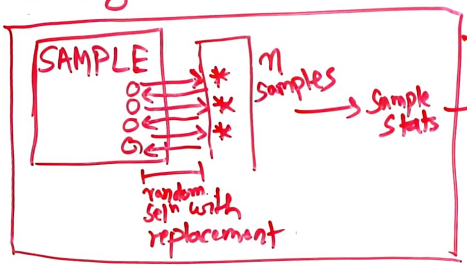
Population
follow some underlying distribution
Sample
Classical Statistics
assumptions on population statistics + statistics of sample
Modern Statistics
Bootstrap method

Regression to the mean
↳ Upon multiple measurement of the same stats. Extreme values usually Precede normal values
↳ Eg: Shorter child of taller parents

Bootstrapping
↳ Sampling with replacement



* Doesn't Require any assumptions on Population Stats
* Get Confidence Intervals based on Sample stats distr

Note :⇒ Doesn't create new info or data.
↳ Just a proxy to show what would be the distribution of Population statistics Under the assumption that the Random Sample drawn is a good Representation of Popula

Distribution of Sample Statistics
↳ CLT ⇒ Central Limit Theorem
↳ distribution of Sample means
↳ Normal Distribution
Given ↳ Population dist?: Normal like
↳ Sample size: Large Enough
↳ Standard Error :⇒ σ of mean distribution
$SE = \frac{\sigma}{\sqrt{n}}$ n: sample size
↳ Validity Comes from CLT assumption

Sampling :-
↳ Data Quality is Data Quantity
↳ Always more important
↳ larger dataset is always desirable for sparse datasets
* Random Sampling
↳ Stratified sampling (for better representation in the dataset of the population)
Shouldn't have Sampling Bias

Selection Bias :⇒ only selecting favorable results
↳ Data snooping :⇒ hunting through data until something interesting comes up
↳ Vast Search effect :⇒ Repeated exp → atleast one favorable outcome