Central Limit Theorem

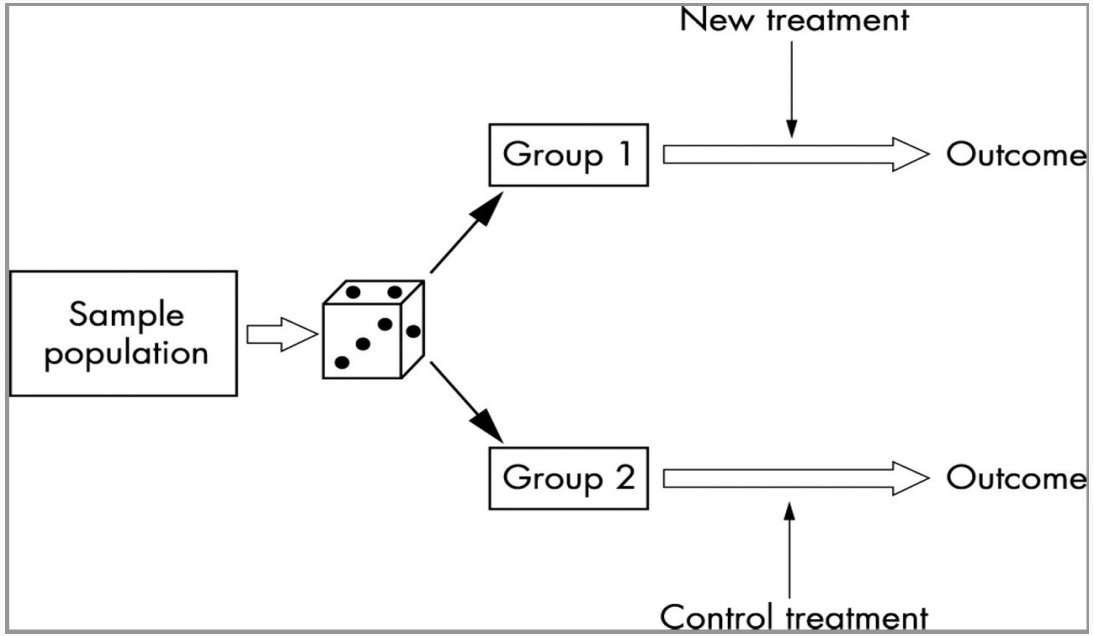
Akshay Pamnani, Gurusankar Gopalakrishnan, Yuan Yao

CLT - A Quick Recap

- What is the Central Limit Theorem?
- Things to consider :
  - Underlying Distribution
  - Random samples
  - Sample size
  - Number of samples
- CLT forms the basis for hypothesis testing

Randomized Control Trials

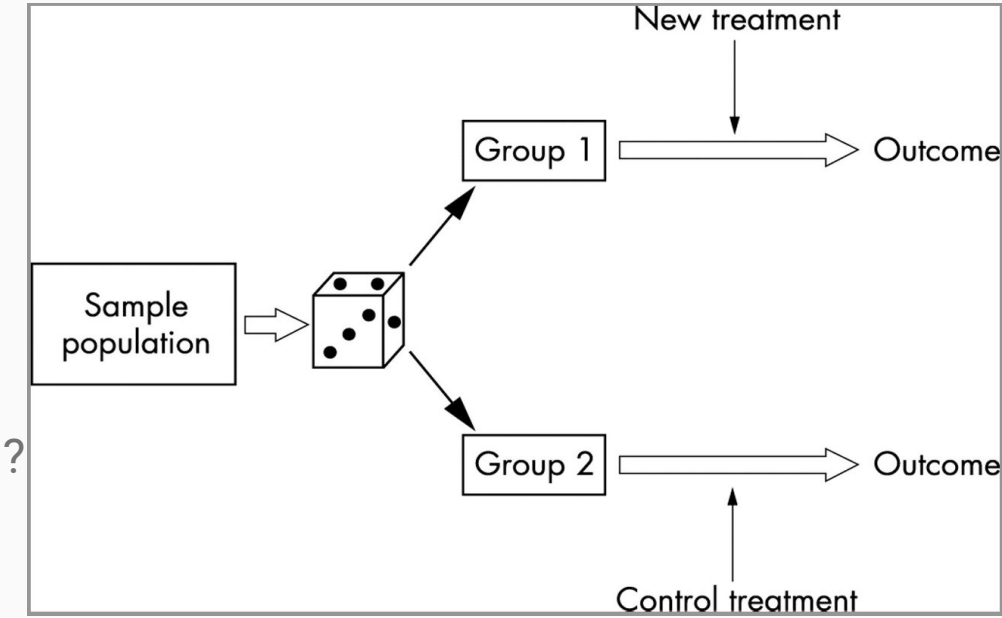
- What are they?
- Why do we need RCTs?
- Gold Standard of evidence



Source: https://emj.bmj.com/content/20/2/164

Explaining RCT with an example

- Example: Drug called ‘Antisugar’
- After two months treatment
- Group1: Mean Hb1Ac=6.1
- Group2: Mean Hb1Ac=6.8
- How do we decide if the drug worked?

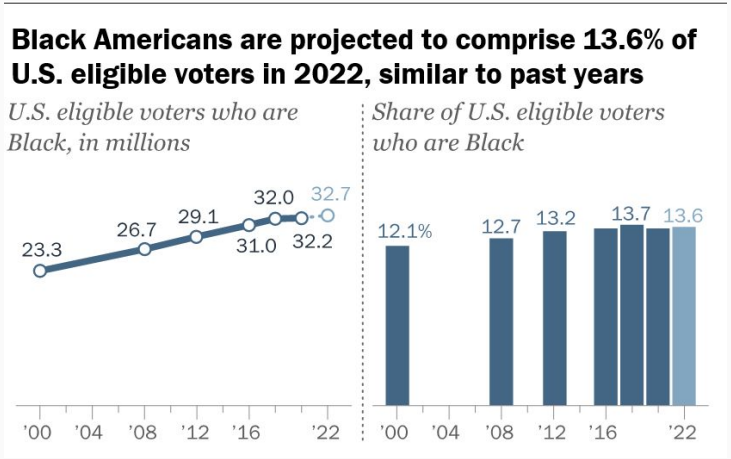


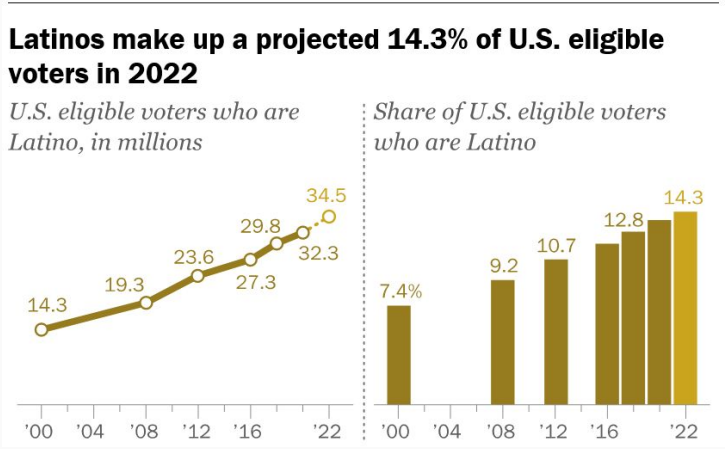
Source: https://emj.bmj.com/content/20/2/164

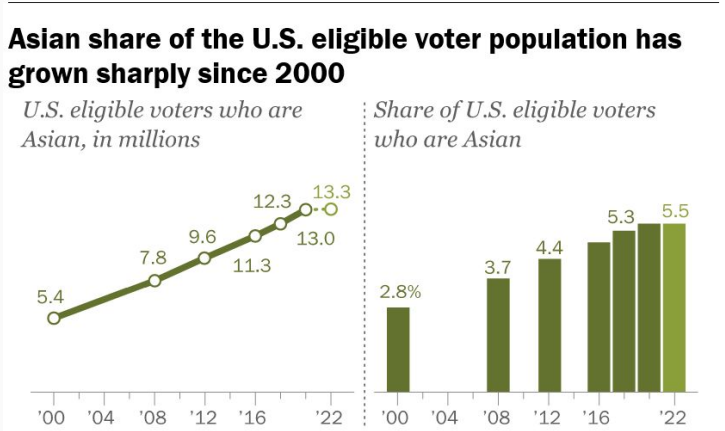
Why CLT is essential to Randomized Controlled Trials?

- Without CLT, we will not be able to establish efficacy of the drug, at required significance level
- Medicine is based on weighing pros and cons, CLT provides a great framework to measure this
- To conclude, in a world without CLT, it would be very difficult to measure efficacy to drugs and take decisions on prescribing them.

US Elections - Can CLT help?







- "Projections" - Estimation using CLT
- Survey data -> Sampling data
- Multiple Surveys -> Multiple Samples

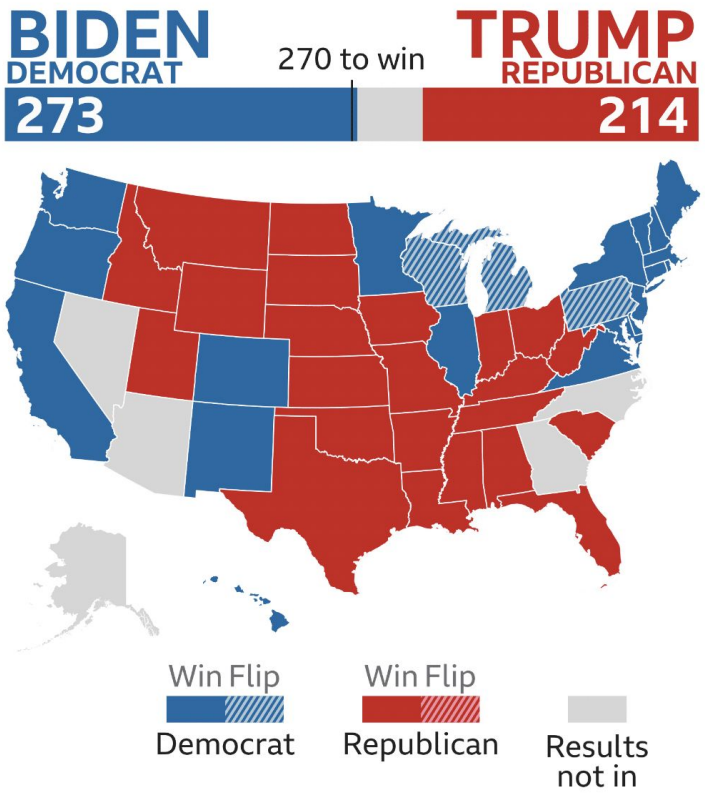
Source: Pew Research Center analysis of 2020, 2018, 2016, 2012 and 2008 American Community Surveys, and 2000 decennial census (IPUMS). Pew Research Center projection for Nov. 1, 2022.

PEW RESEARCH CENTER

Election Poll Predictions - Why are they bad?

- 2016 - Trump victory against all claims from Election Polls.
- The nationwide polling average gave Clinton about a 3-point lead overall, and the state-by-state polls indicated that she would win at least 300 electoral votes.
- Why? Did CLT fail?
  - Pollsters herd mentality -> violates random sampling
  - Silent Vote -> people might not want to disclose true intention
  - Hard to reach every demographic -> not a true representation of the population.

Exit Polls are more reliable?



Source: NEP/Edison via Reuters, 16:42 GMT (11:42 EST)

BBC

- Poll people as they come out
- “Truly random” - to some degree
- Covers almost everybody -> large sample sizes

Noise Reduction in Image

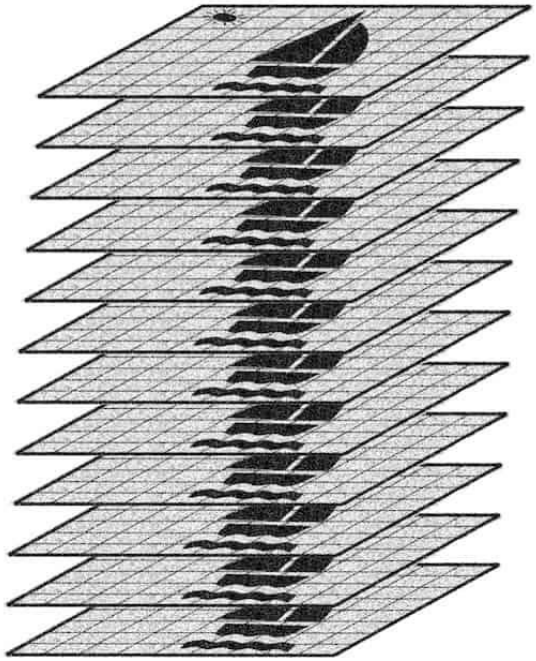
What is noise in image?

In any electronic or physical system, noise is inevitable and affect the image quality.

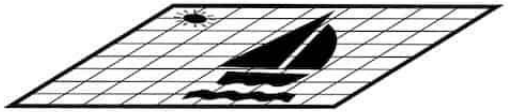
Image (e) has less noise, and the others have different levels of noises.

Source: https://vciba.springeropen.com/articles/10.1186/s42492-019-0016-7

Denoising Technique using the Idea of CLT



Stack of 12 exposures with random high noise (low signal to noise ratio)



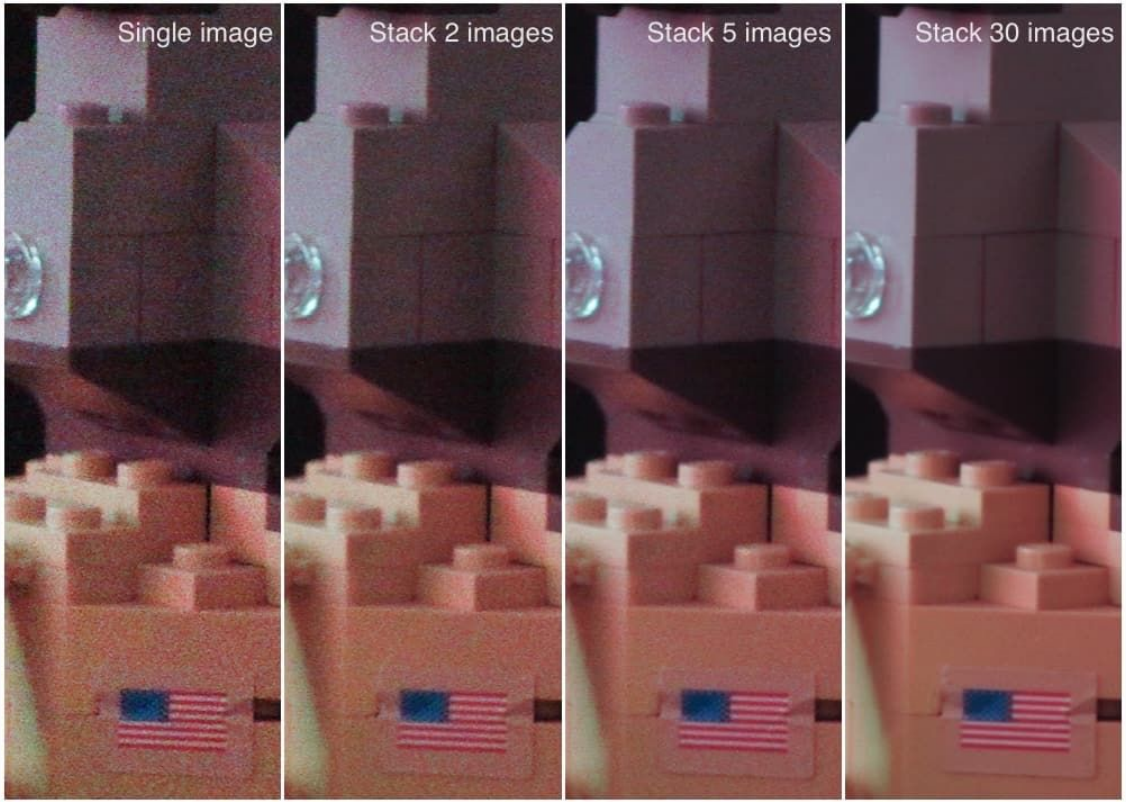
stacked (averaged) image has low noise and improved signal to noise ratio

Stacking a number of same images can help to denoise the object and improve the image quality.

The average values of each pixel will closer to its true value (mean), and the value of noise will decrease to around zero.

Source: https://nightskypix.com/astrophotography-stacking-software/

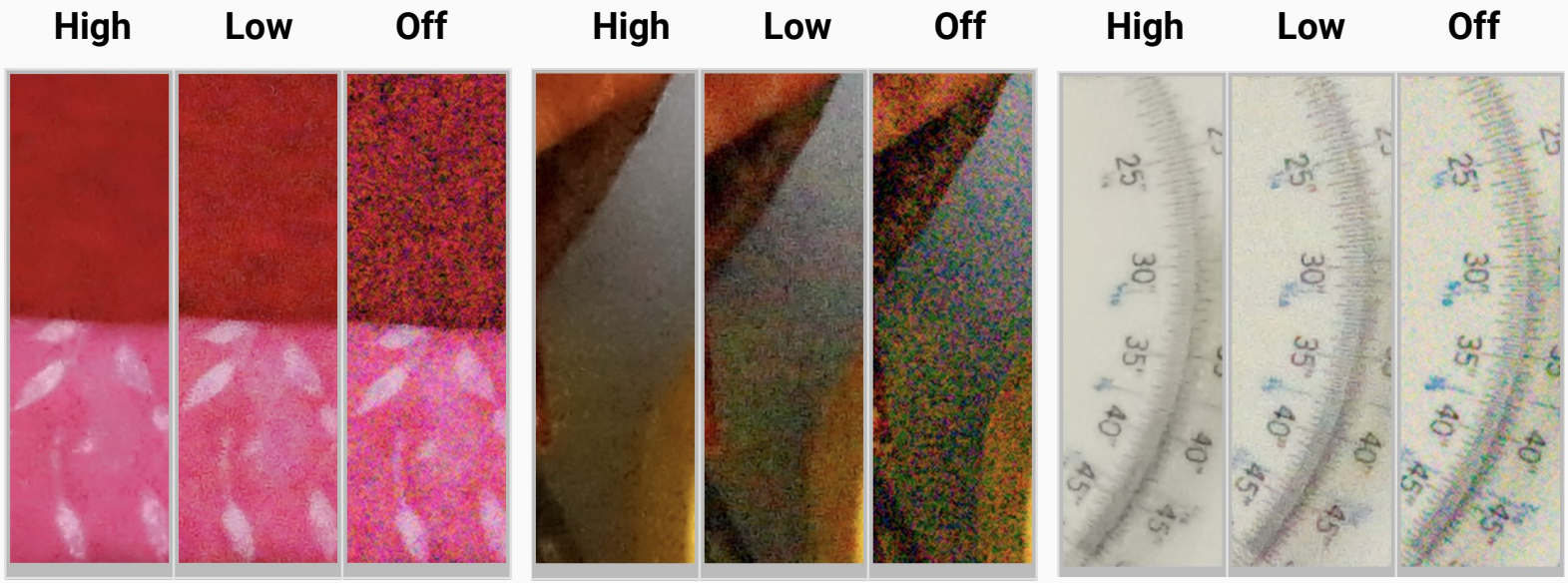
Denoising with Different Sample Sizes



An example with different number of stacking images.

The more images we stack, the higher image quality we can get.

Denoising Technique used in Camera



Many digital cameras use this technique for each single image you shoot!
It takes several images at the same time can calculate the mean values of each pixel.

Source: https://www.imaging-resource.com/PRODS/sony-a7/sony-a7HI_ISO_NR.HTM

In summary -

- CLT has a wide range of real-life applications
- Parametric Hypothesis testing is more robust due to CLT
- CLT is useful, but wrong data/incomplete data -> Wrong inferences
- **Fun Fact** : CLT is also widely involved in all parameter estimation process, e.g. EE System, Simulation, ML models, etc

Questions?