Eltecon Data Science Course by Emarsys

Measuring uncertainty

András Bérczi

October 14, 2020

Homeworks from last week

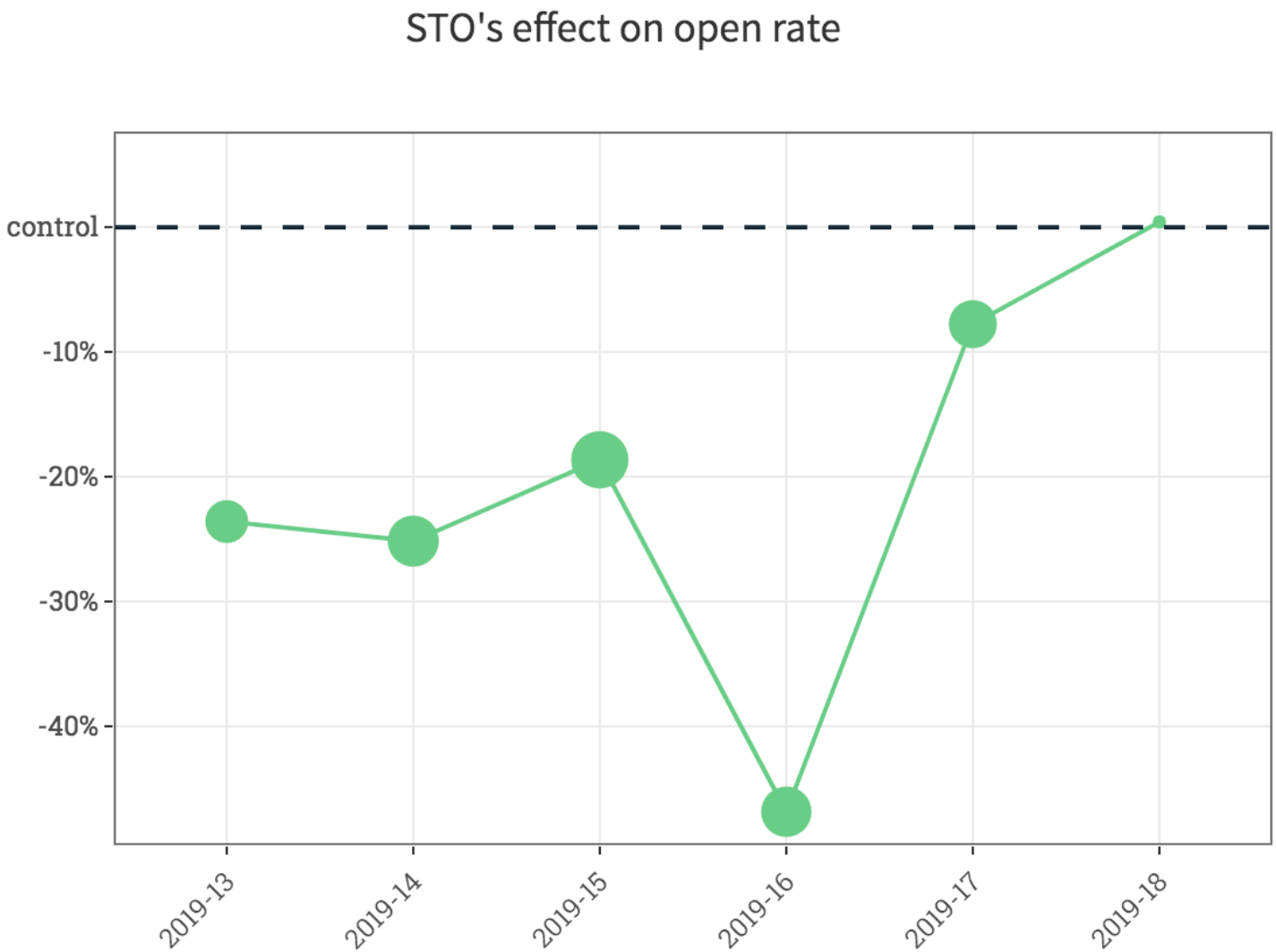
Any questions about final project?

Measuring uncertainty

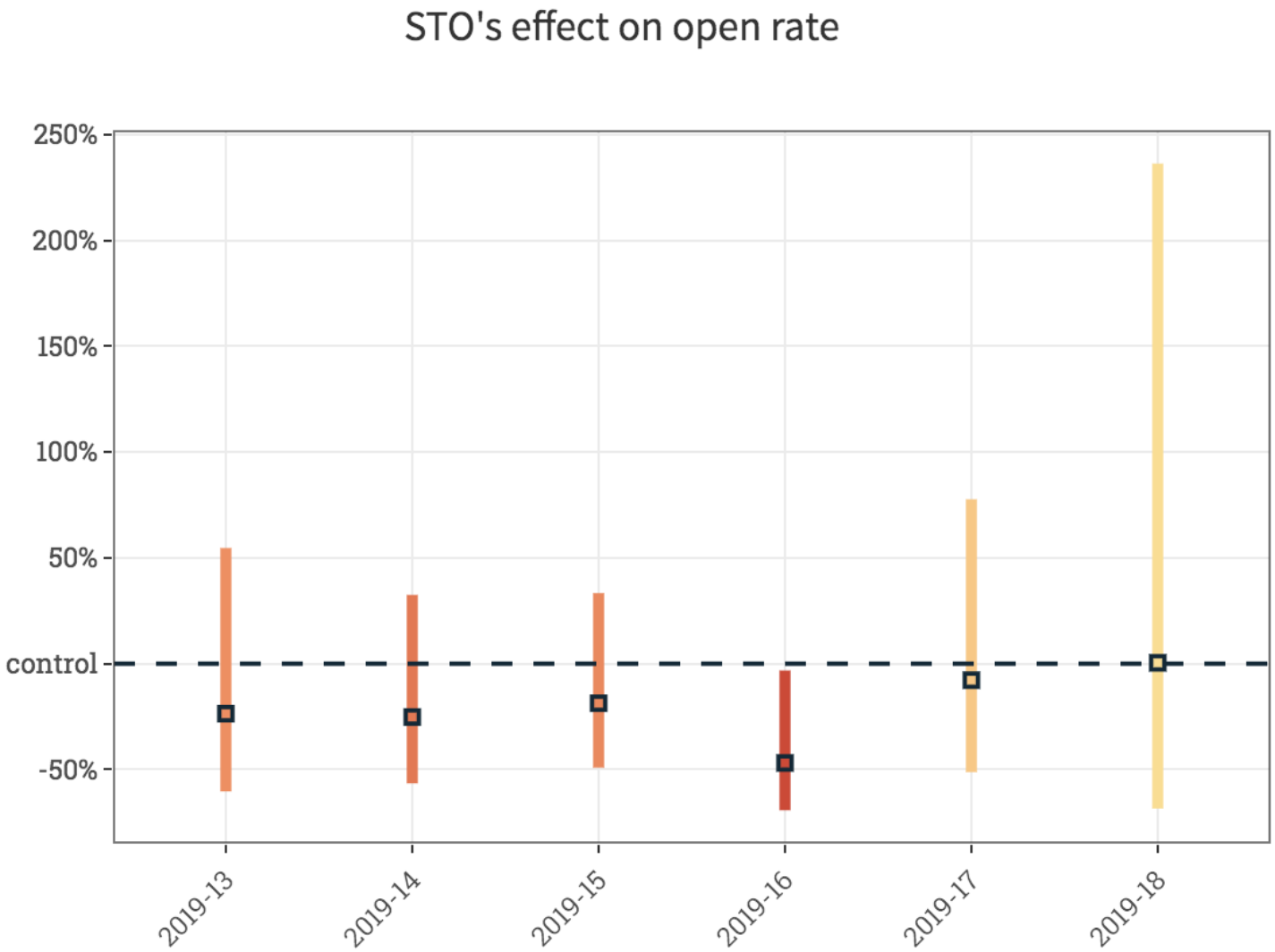
We can always measure something from our data...

...but how sure can we be about our measurement?

We can always measure something from our data...



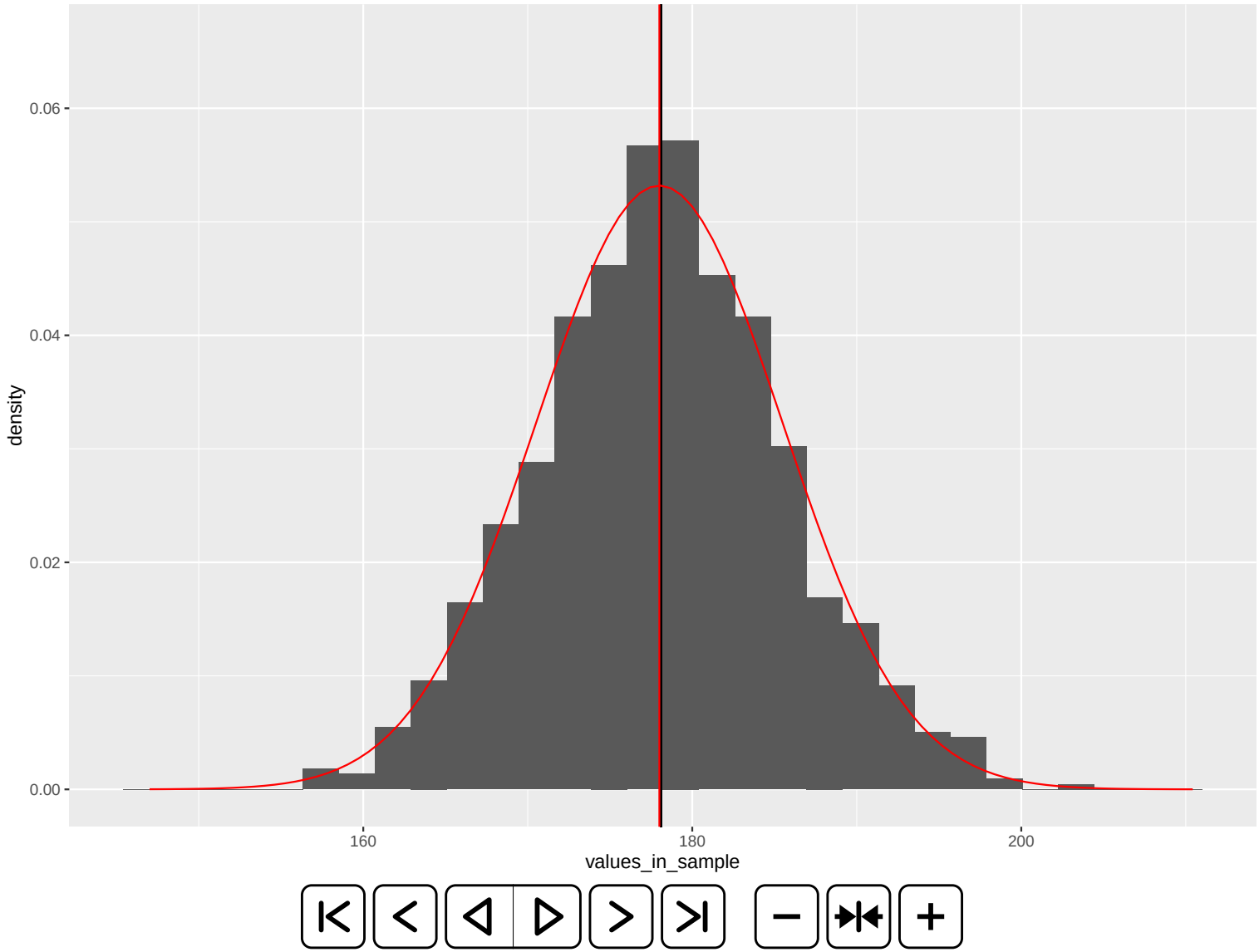
But not necessarily significant!



Why do have uncertainty in the measurement?

- If you knew the whole population, there wouldn't be uncertainty in your measurement
- But we only see 1 'segment' of the data = we have a sample of the population

Sampling from a population



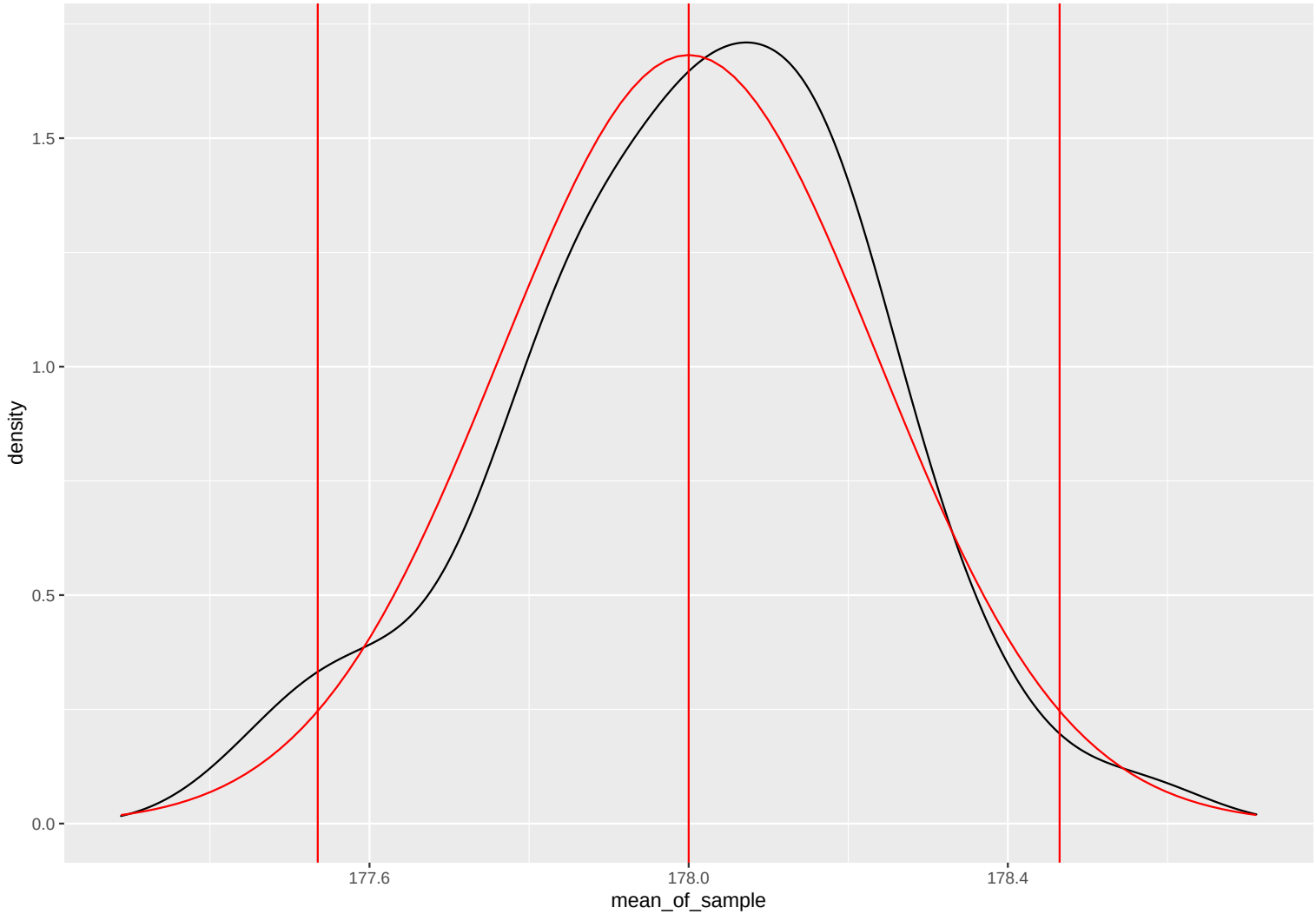
Law of Large Numbers

The average of the results obtained from a large number of trials should be close to the expected value and will tend to become closer to the expected value as more trials are performed. - Wikipedia

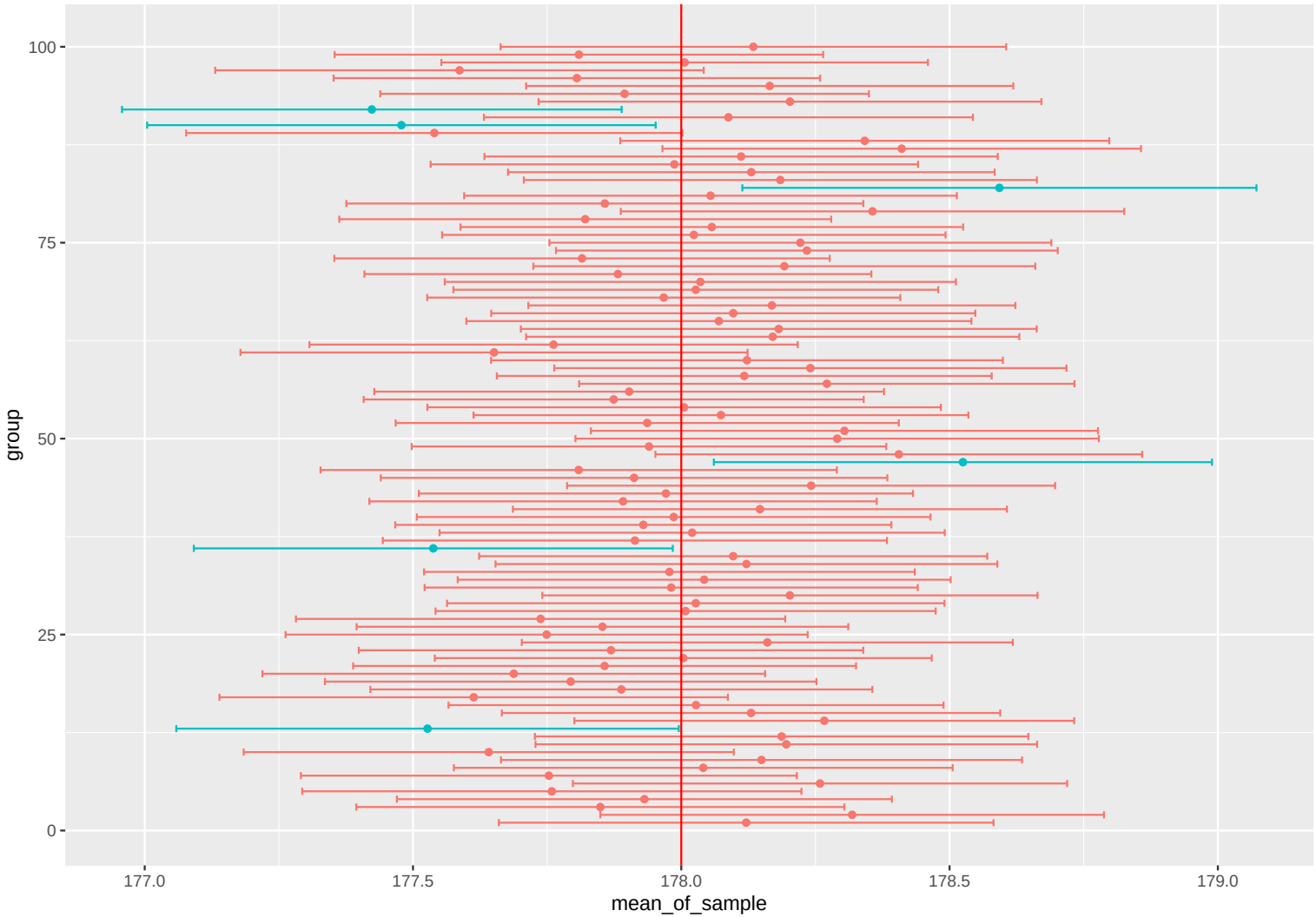
Central Limit Theorem

When independent random variables are added, their properly normalized sum tends toward a normal distribution (informally a bell curve) even if the original variables themselves are not normally distributed. - Wikipedia

Distribution of sample means - LLN + CLT



What we do when we check for CI



What are the key assumptions?

- i.i.d. sampling
- finite variance

How can we calculate uncertainty to our measurement?

- Based on variance of known distribution
- Monte-Carlo method
- Bootstrapping
- (and other methods as well of course)