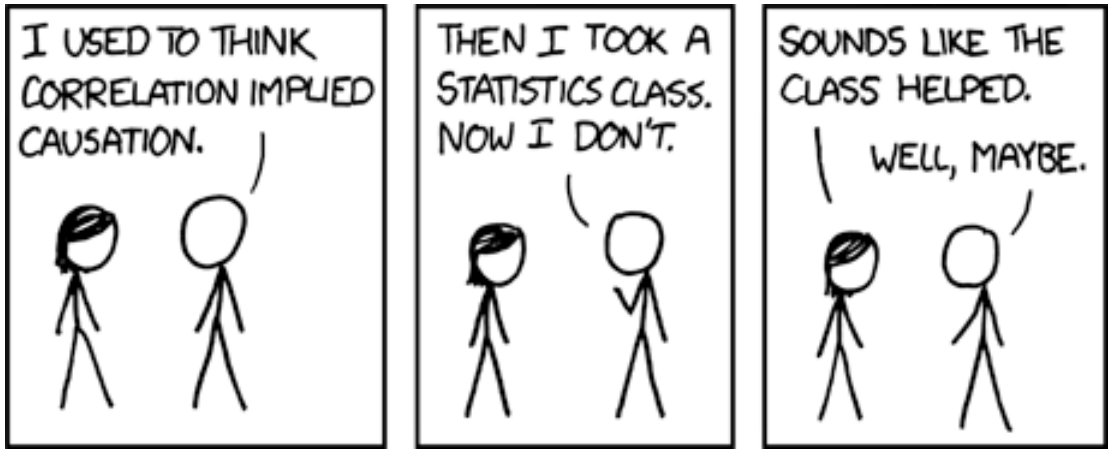
Research Methods for Political Science

MT week 1, lecture 2



Why statistics?
The problem of sampling

TRINITY COLLEGE DUBLIN
COLÁISTE NA TRÍONÓIDE, BAILE ÁTHA CLIATH
THE UNIVERSITY OF DUBLIN

Dr. Thomas Chadefaux

Assistant Professor in Political Science

thomas.chadefaux@tcd.ie

Why statistics?

- Discovering patterns in large datasets
- Working with quantitative observational data
- Working with samples

Sampling

- Learning about the many by studying only a few (relatively speaking)

- A random sample of 1000 people can tell you a lot about what the population as a whole thinks.

Population and sample

- Population: the collection of units (people, counties, parties, wars) to which we want to generalize a set of findings or a statistical model.

- Sample: a smaller collection of units from a population.

Probability

- An estimate of the likelihood that something will happen (or is the case)

- P(heads) = .5
- P(tails) = .5
- P(heads OR tails) = P(heads) + P(tails) = 1

Example: gender in Ireland

Census 2011:

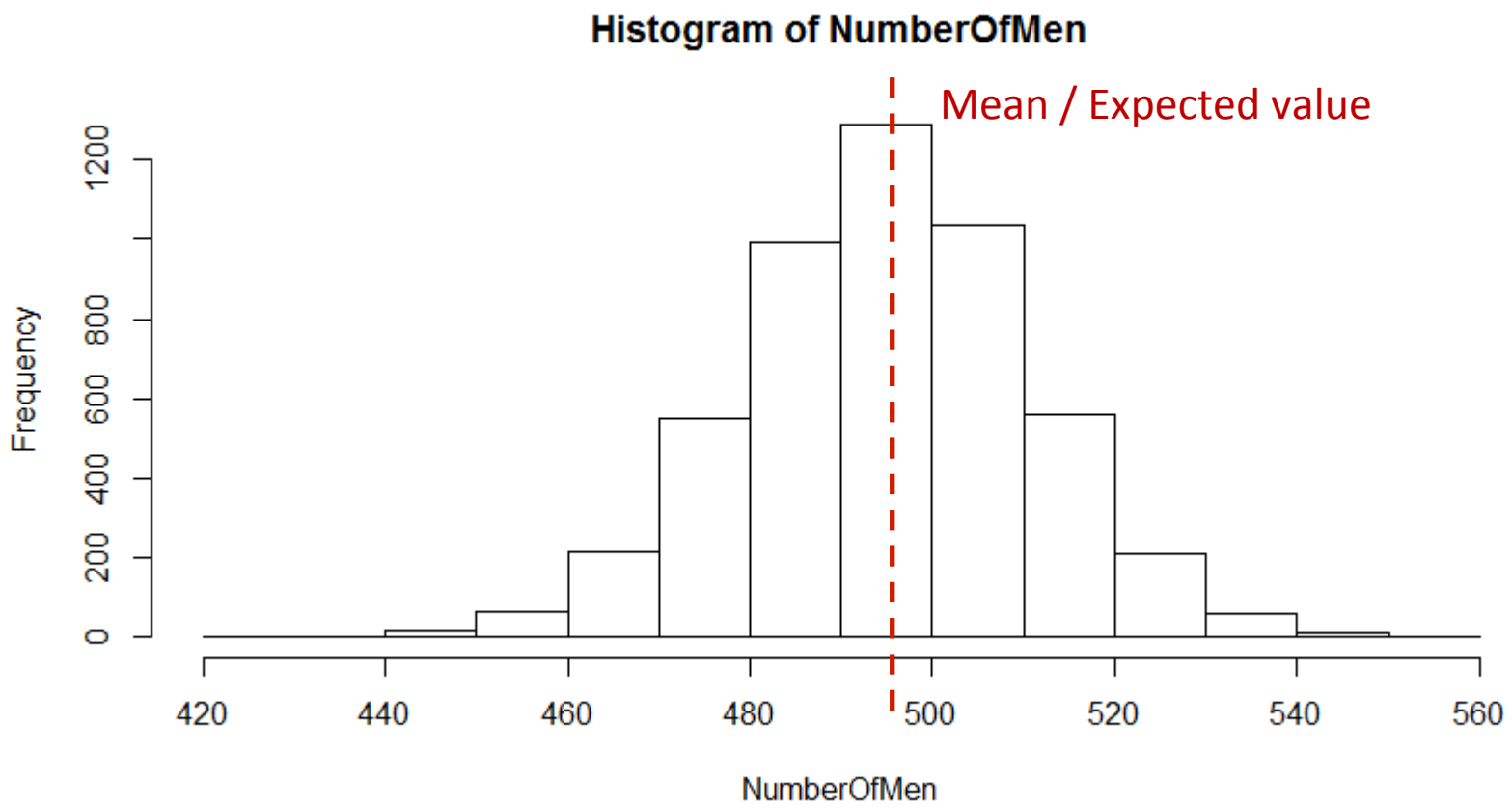
Female: 2262651 (50,5%)

Male: 2221672 (49,5%)

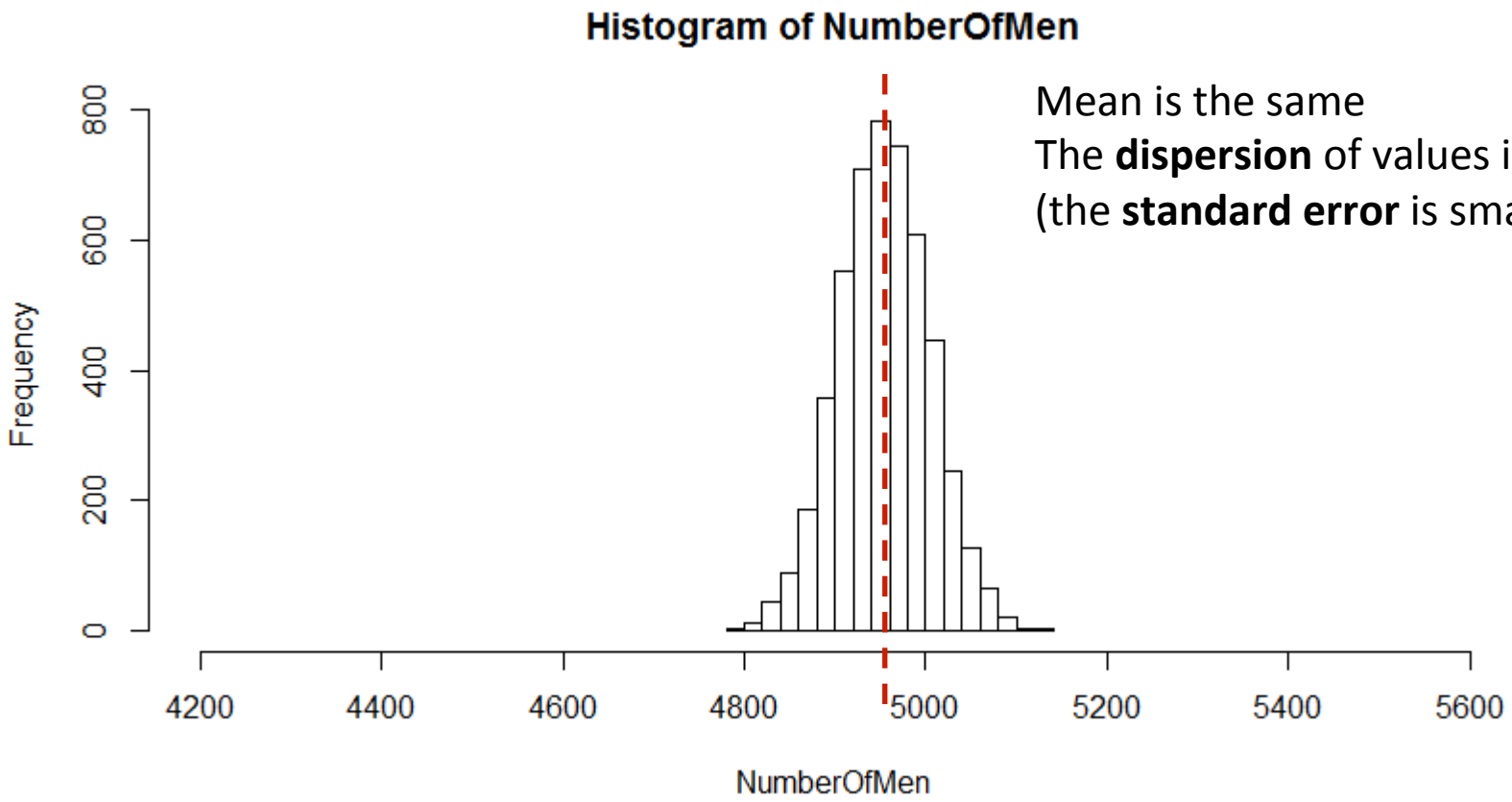
Take a sample of 1000 people

How many men?

5000 samples of 1000 people



5000 samples of 10,000 people



Mean is the same
The **dispersion** of values is smaller
(the **standard error** is smaller)

Random sample

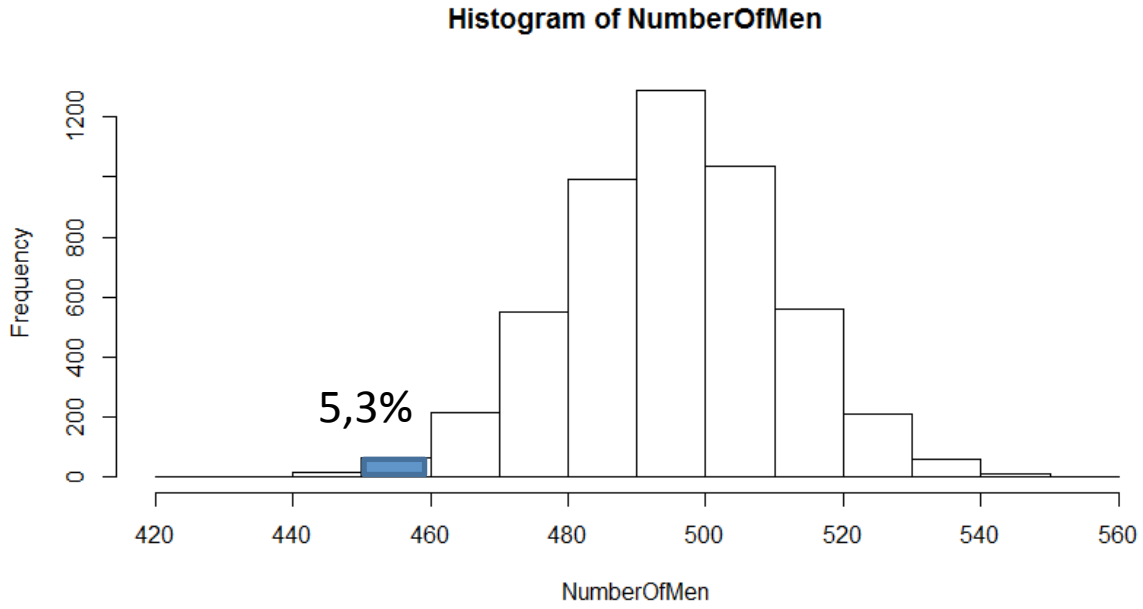
- The mean sample (the mean of the sampling distribution) will be equal to the population mean

(provided you get enough samples, an infinity actually)

- As the number of observations (**n**) in the samples increases, the *standard error* (**s**) of our estimate will decrease.

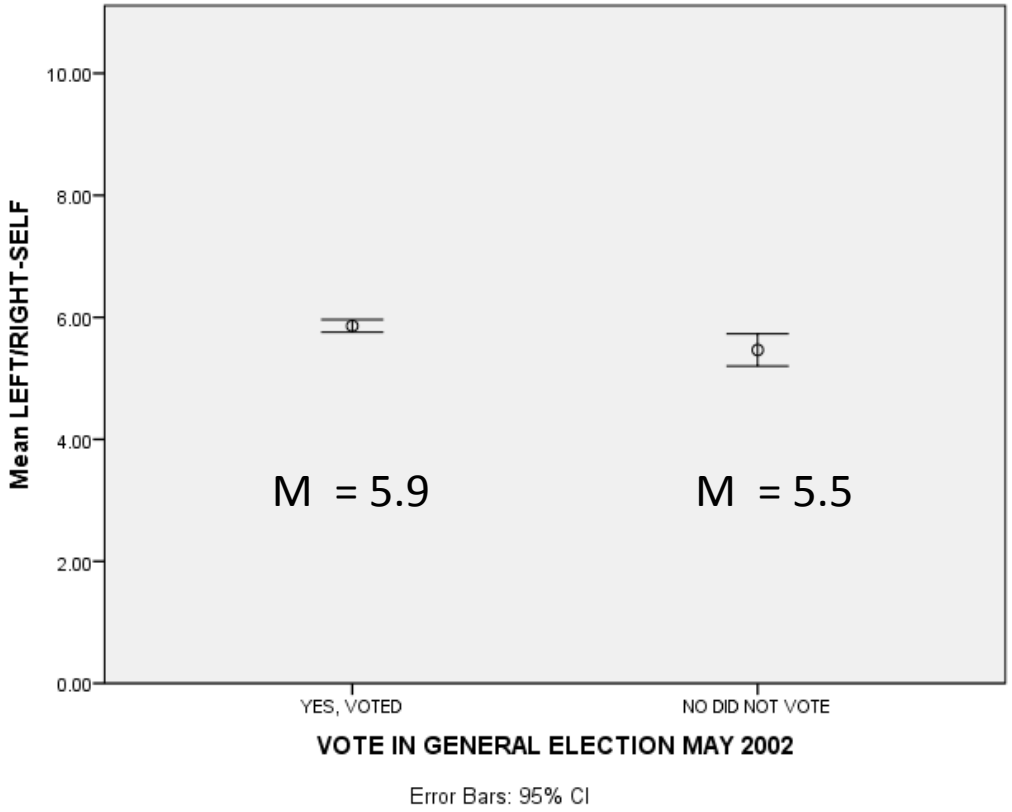
You don’t know which sample you have…

- E.g. random sample of 1000 people. We find only 45% is male. Do we have a problem?



Example: Left - Right

"In Politics people sometimes talk of left and right. And Where would you place yourself on a scale from 0 to 10 where '0' means the left and '10' means the right? Please tick (a) one box."



Example: Left-Right (II)

- Difference between voters and non-voters? Are non-voters more left-wing?

- *Depends on the number of people that answered the survey (N) and the distribution of their answers (s).*

- You need a statistical test to see whether these differences are *statistically significant*.

Why *random* sampling?

If the sample is random – so everyone has an equal chance of being selected – then we can apply standard statistical techniques. For example, we would know that the mean of the sampling distribution is equal to the population mean.

It ain't so easy…

- Often difficult to get a good list of the whole population (sampling frame)
  - E.g., How would you get a list of all combattants in Syria?
- Cost
- Access to data, e.g. participation in surveys, availability of data

- If sampling is non-random >> even more cautious