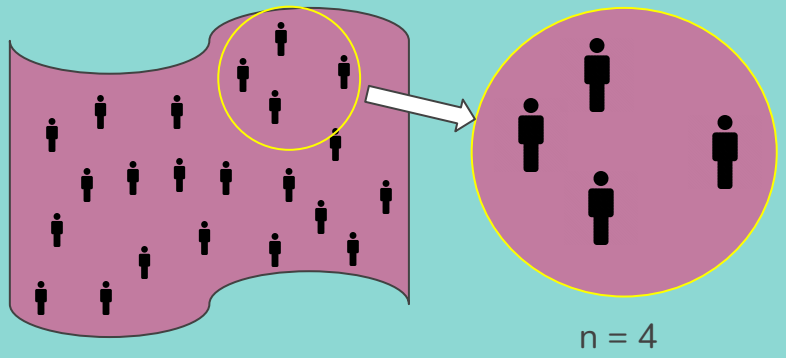
Spatial Statistics and Survey Sampling

Andrew Sousa
DRP Fall 2023

Differences between Statistics and Parameters

Statistic

Data collected by sampling part of the population

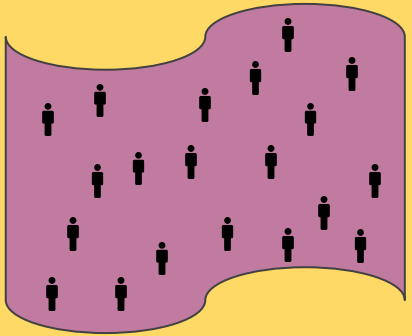


n = 4

"50% of our sample says they like pineapple on pizza"

Parameter

Data that is descriptive of the entire population

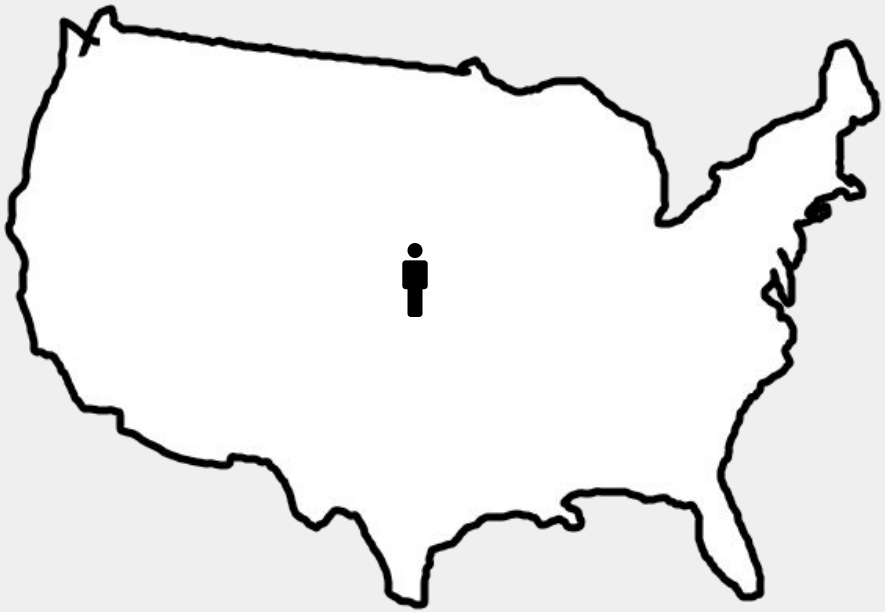


N = 20

"45% of people in the population say they like pineapple on pizza"

Why take samples instead of just the whole population?

Short answer: time, resources, and money



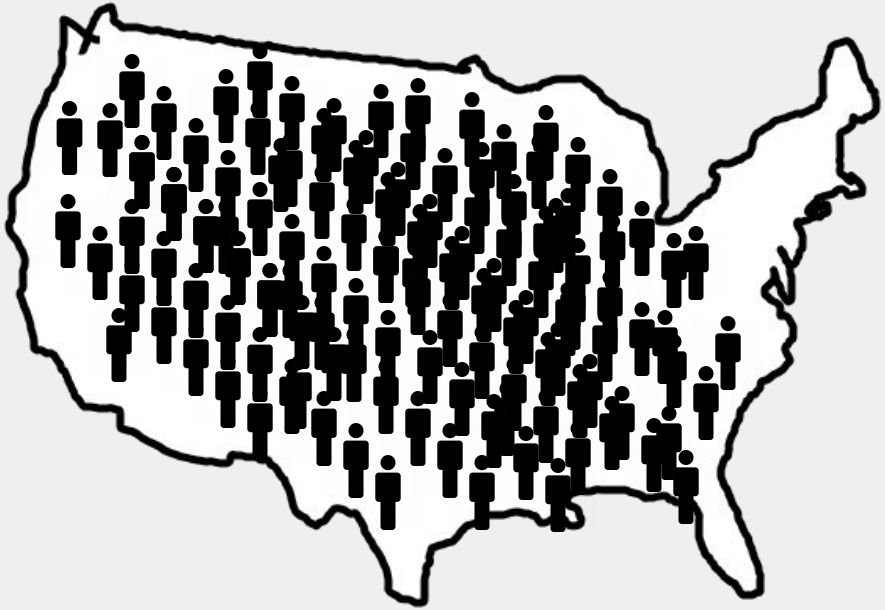
United States (Population: 335,734,899)*
example: each person costs $10 to survey

Sample Size: 1
Cost: $10

*United States Census Bureau

Why take samples instead of just the whole population?

Short answer: time, resources money



United States (Population: 335,734,899)*
Example: each person costs $10 to survey

Sample Size: 100,000
Cost: $1,000,000

*United States Census Bureau

Why take samples instead of just the whole population?

Short answer: money

United States (Population: 335,734,899)*
Example: each person costs $10 to survey

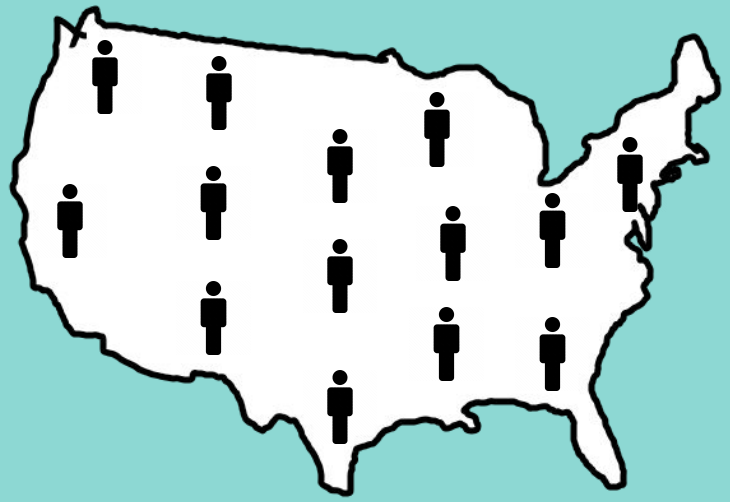
Sample Size: 335,734,899
Cost: $3,357,348,990

*United States Census Bureau

Differences between Random and Probability Sampling

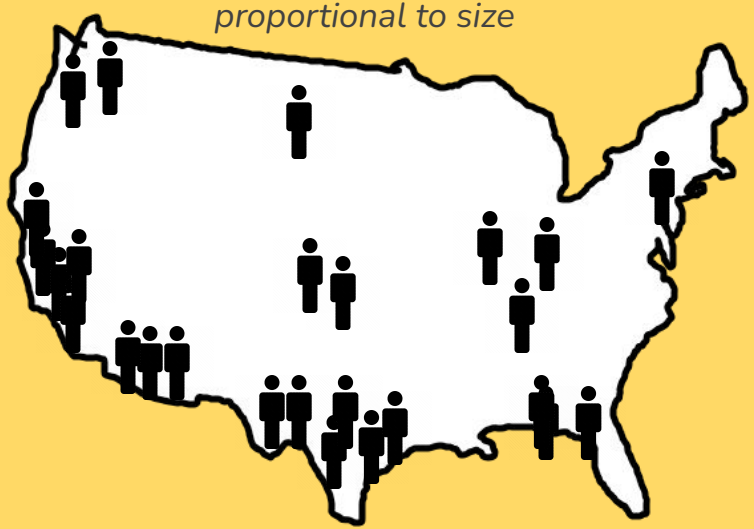
Random Sampling

Each person is randomly chosen and has the same odds of being chosen



Probability Sampling

Choosing samples randomly based on a probability of proportional to size



Probability Sampling

- Randomly selecting a sample based on some method, where each individual has an equal probability of being selected relative to their peers.

- Methods of probability sampling include techniques like:
  -Simple random sampling
  -Stratified Random Sampling
  -Cluster Random Sampling
  -Two-Stage Sampling

Horvitz-Thompson Estimator

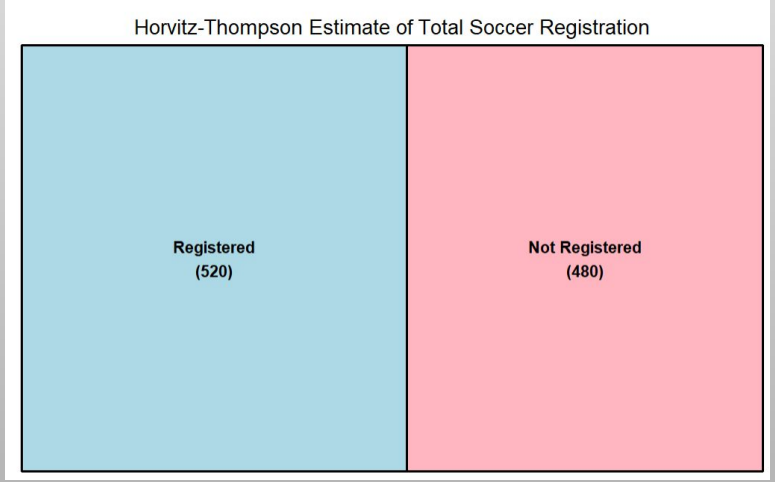
Uses observed values and probability of a unit being chosen (determined by the type of sampling method) to estimate the total amount or proportion of something in a population.

- Summation of “total additions” (unit weight * unit observed value), where weight is determine by probability of being chosen

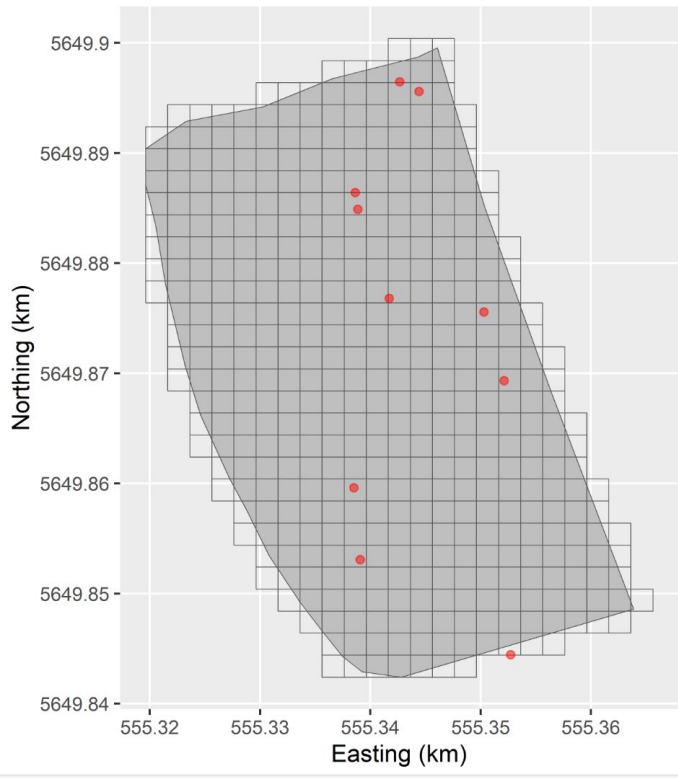
- Allows us to make accurate predictions about the population without having to sample every single person.

$$\hat{Y}_{HT} = \sum_{i=1}^{n} \pi_i^{-1} Y_i,$$

We ask 50 people in a city of 1000 if they are registered for soccer or not



Spatial Sampling



- Grid represents our complete area of inquiry, which is very, very large!
- To address this, we divide our area into square grids, randomly choose some, and then a point is randomly chosen from each grid.
- Using this technique alongside the Horvitz-Thompson estimator, we can make estimations about some observation about our population without having to sample every single part of the area.

Conclusions

1) Sampling and surveying can be a difficult process to fulfill accurately: choosing and implementing the correct design is crucial to get an accurate representation of the population.

2) With a sound sampling system in place, using the Horvitz-Thompson estimator is a strong and simple method to make accurate estimations about characteristics in our population.

References

Brus, Dick J. "Spatial Sampling with R." *Chapter 2 Introduction to Probability Sampling*, 3 Sept. 2023, dickbrus.github.io/SpatialSamplingwithR/IntroProbabilitySampling.html.

"U.S. and World Population Clock." *United States Census Bureau*, www.census.gov/popclock/. Accessed 20 Nov. 2023.