Statistics and its roles

Statistics is the *science* of collecting, organizing, and interpreting data. Its role is to:

- Summarize data
- Detect hidden relationships and trends in variables
- Guide analytical theory
- Guide decisions
- Predict the future
- Quantifies how confident we can be in conclusions

Basic terminology

- **Population:** the *complete* set of people or things being studied.
- **Sample:** a subset of the population from which the raw data are actually obtained.
- **Population parameter:** characteristic of the population that is being studied.
- **Sample statistics:** numbers or observations that summarize the raw data.

Example of a statistical study

A researcher wants to find out the average daily screen time of high school students in a city. They survey 200 students chosen randomly from all high schools in that city.

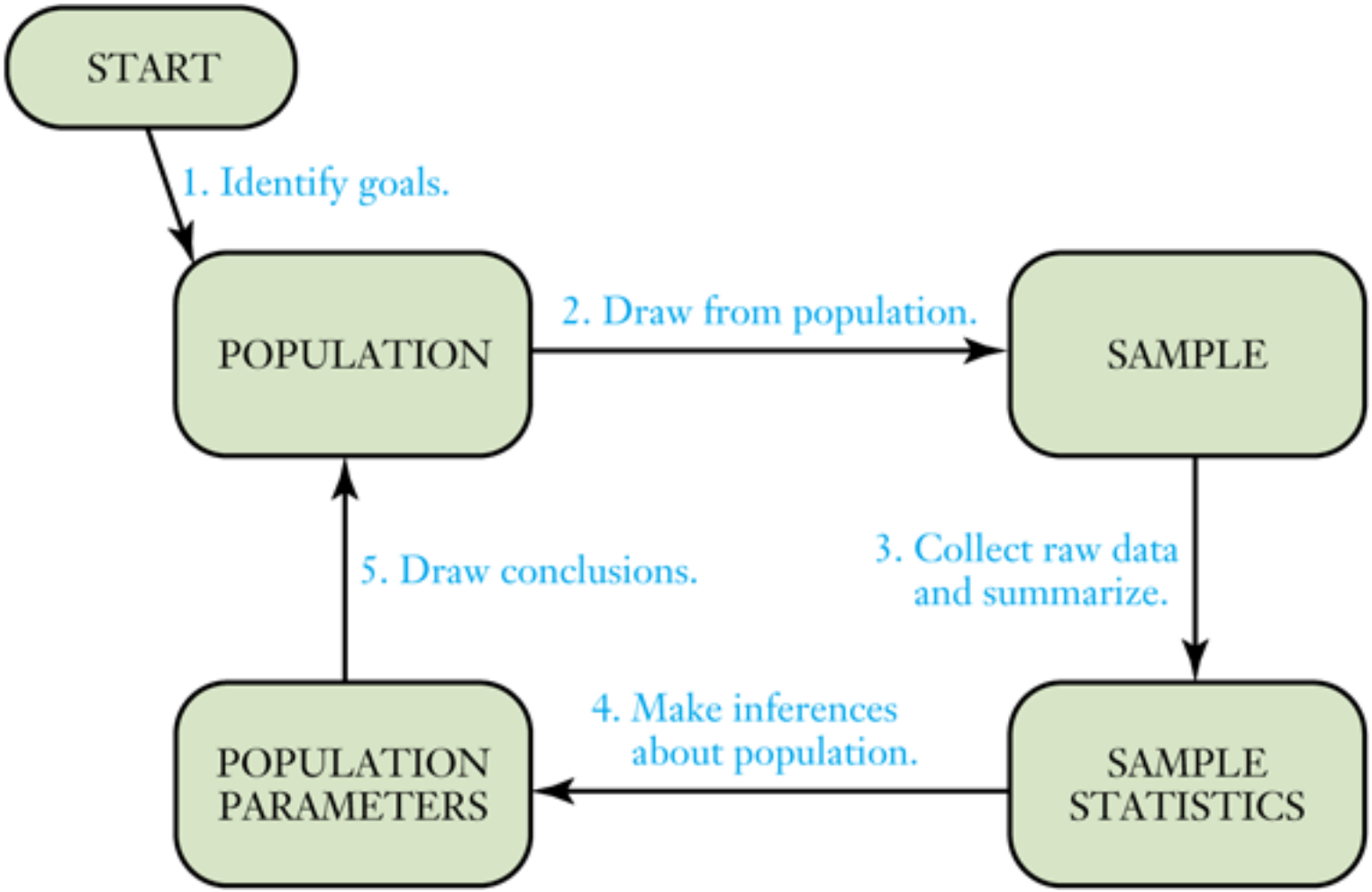
Population: all high school students in the city

Sample: the 200 students that were surveyed

Population parameter: average daily screen time of all high school students in the city

Sample statistic: average daily screen time calculated from the 200 surveyed students

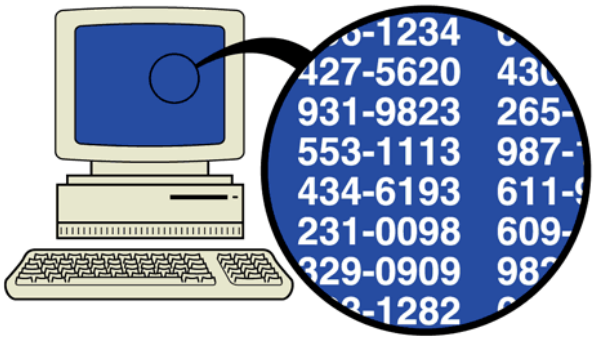
How Statistics works



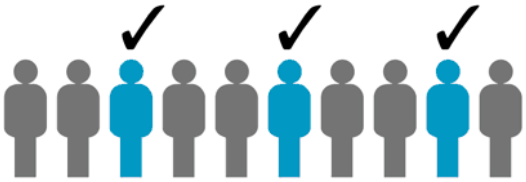
Choosing a sample

- **Representative sample:** a sample whose relevant characteristic matches that of the population.
- A statistical study suffers from **bias** if its design or conduct tends to favor certain results.
- Common sampling techniques:
  - Simple random sampling
  - Systematic sampling
  - Convenient sampling
  - Cluster sampling
  - Stratified sampling

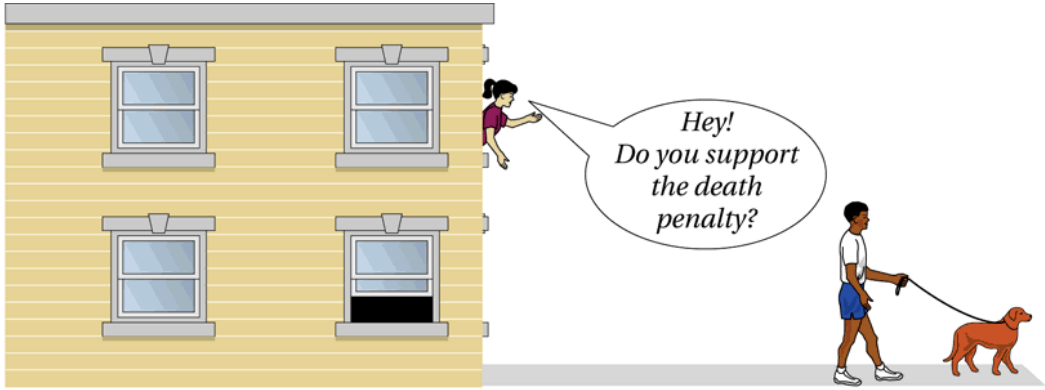
Common Sampling Techniques



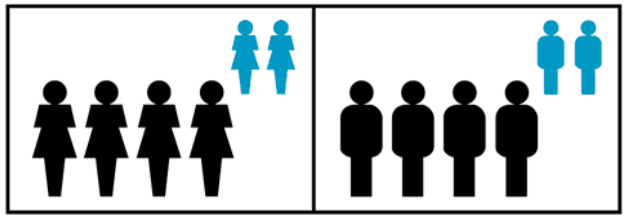
Simple Random Sampling:
Every sample of the same size has an equal chance of being selected. Computers are often used to generate random telephone numbers.



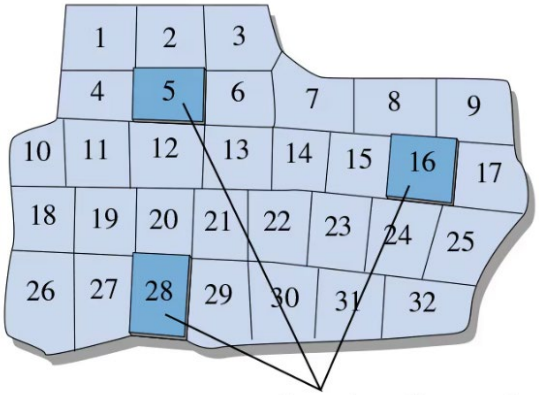
Systematic Sampling:
Select every *k*th member.



Convenience Sampling:
Use results that are readily available.



Stratified Sampling:
Partition the population into at least two strata, then draw a sample from each.



Interview all voters in shaded precincts.

Cluster Sampling:
Divide the population into clusters, randomly select some of those clusters, then choose all members of the selected clusters.

Types of Statistical Study

Based on the method of investigation, statistical studies are categorized into:

- **Observational study:** researchers observe or measure characteristics of the sample members without changing them.
- **Experiment:** researchers apply a treatment to some sample members and then look to see whether the treatment has any effects.

Types of Statistical Study - Example

Goal: determine whether regular exercise reduces the risk of heart disease.

Observational study: observe a large group of people with different exercise habits. Record how much each person exercises and whether they develop heart disease.

Experiment: assign participants to two groups: one group follows a prescribed exercise program, the other maintains their usual routine. Then compare the heart outcomes between groups.

Experiment

- **Treatment group:** the group of sample members who receive the treatment being tested.
- **Control group:** the group of sample members who do *not* receive the treatment being tested.
- **Single-blind:** the participants don't know if which group they are in, but the experimenters know.
- **Double-blind:** if neither the participants nor the experimenters know who are in the treatment group and who are in the control group.

Experiment

- **Placebo**: an inactive substance used to test real treatment effects.
- **Placebo effect:** the situation when patients improve due to *believing* they are receiving an effective treatment.
- **Example:** Researchers test whether a study lamp improves students' learning performance.
  - One group uses a lamp said to emit concentration-boosting light. Another group uses a regular lamp that looks identical but has no special feature.
  - If both groups perform equally well, the improvement may result from belief in the lamp's effect, not the light itself.

Observational study

- **Case-control study:** uses data from the past to divide the sample into a group of cases who engaged in the behavior under study and a group of controls who did not.
- The participants who engage in the behavior under study form the **cases**.
- The participants who do not engage in the behavior are the **controls**.