

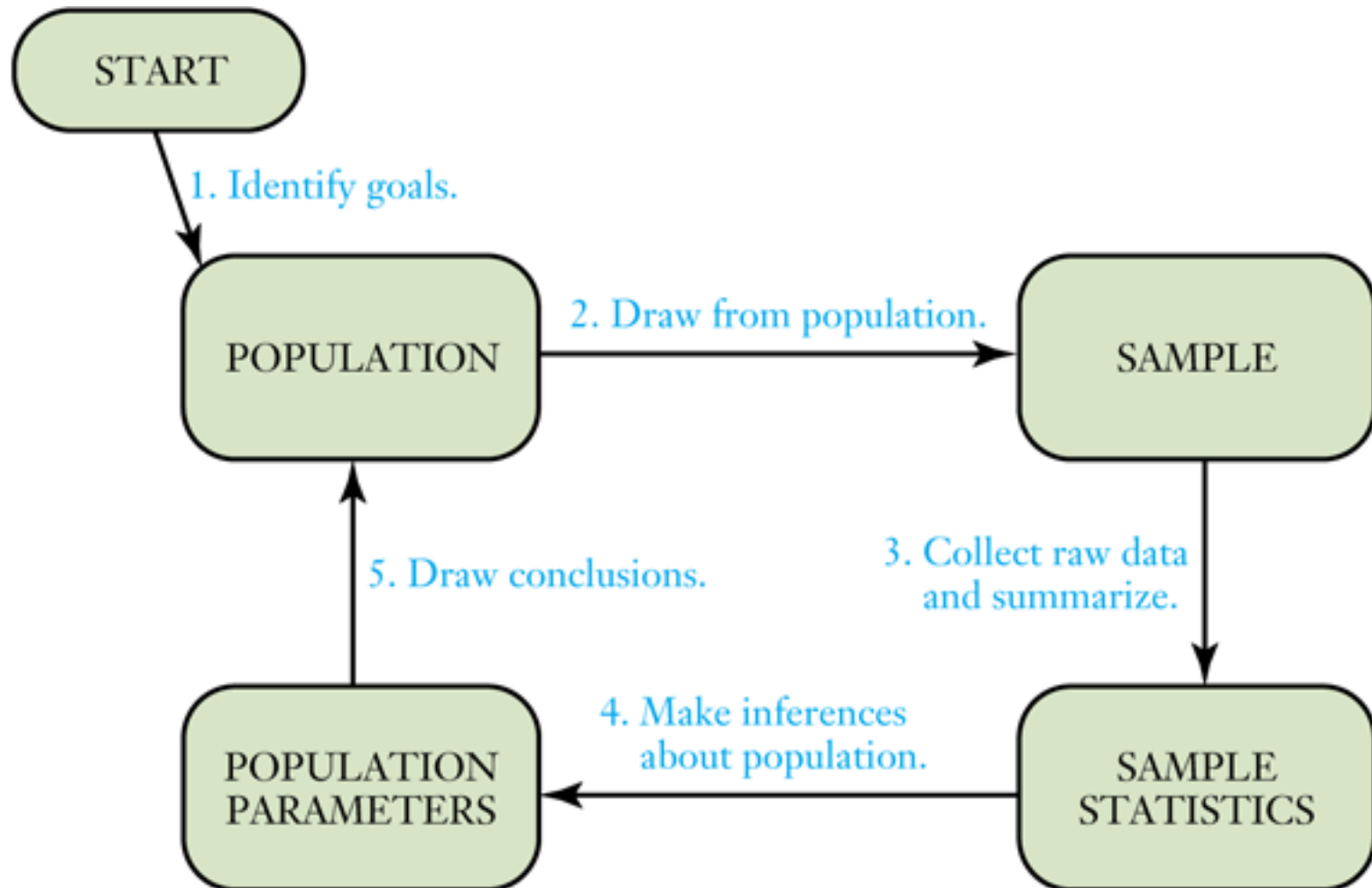
# Two Definitions of Statistics

- Statistics is the *science* of collecting, organizing, and interpreting data.
- Statistics are the *data* that describe or summarize something.

# Definitions

- The **population** in a statistical study is the *complete* set of people or things being studied.
- The **sample** is the subset of the population from which the raw data are actually obtained.
- **Population parameters** are specific characteristics of the population that a statistical study is designed to estimate.
- **Sample statistics** are numbers or observations that summarize the raw data.

# Elements of a Statistical Study



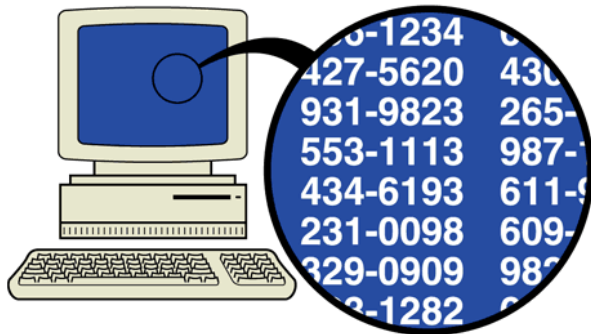
# Describe the study

- An AP/CBS telephone poll of 998 randomly selected Americans revealed that 6 in 10 people believe there has been progress in finding a cure for cancer in the last 30 years.
- What are the population, sample, population parameter, and sample statistics of this study?

# Definitions

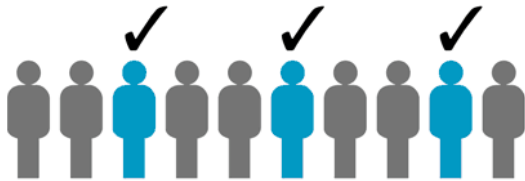
- A **representative sample** is a sample in which the relevant characteristics of the sample members match those of the population.
- A statistical study suffers from **bias** if its design or conduct tends to favor certain results.

# Common Sampling Techniques



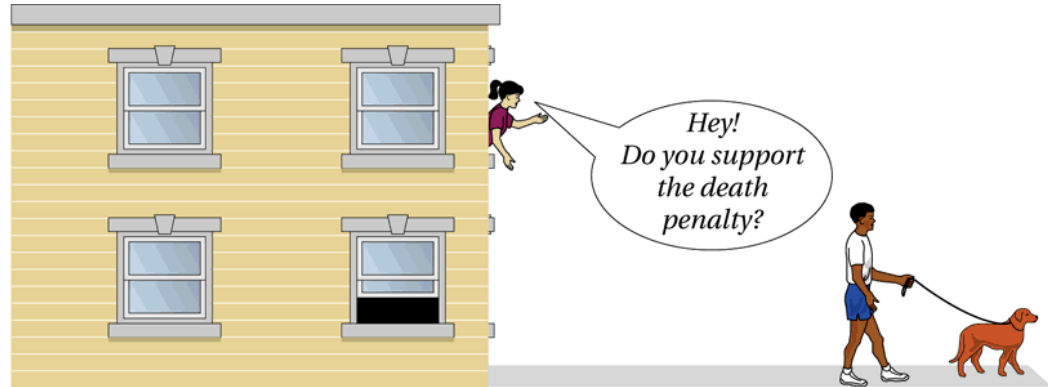
## Simple Random Sampling:

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



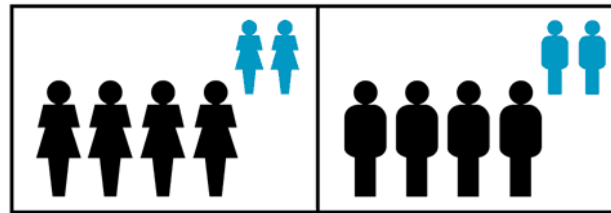
## Systematic Sampling:

Select every  $k$ th member.



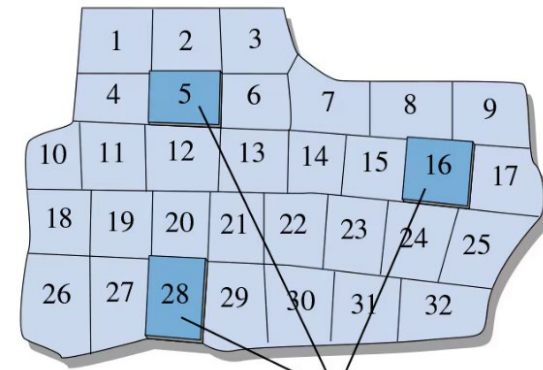
## Convenience Sampling:

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

- In an **observational study**, researchers observe or measure characteristics of the sample members but do not attempt to influence or modify these characteristics.
- In an **experiment**, researchers apply a treatment to some or all of the sample members and then look to see whether the treatment has any effects.

# Treatment and Control Groups

- The **treatment group** in an experiment is the group of sample members who receive the treatment being tested.
- The **control group** in an experiment is the group of sample members who do *not* receive the treatment being tested.



# Placebos and the Placebo Effect

- A **placebo** lacks the active ingredients of a treatment being tested in a study, but is identical in appearance to the treatment.
- The **placebo effect** refers to the situation in which patients improve simply because they believe they are receiving a useful treatment.

# Blinding in Experiments

- An experiment is **single-blind** if the participants do not know whether they are members of the treatment group or members of the control group, but the experimenters do know.
- An experiment is **double-blind** if neither the participants nor the experimenters know who belongs to the treatment group and who belongs to the control group.

# Definitions

- A **case-control study** is an observational study that resembles an experiment because the sample naturally divides into two (or more) groups.
- The participants who engage in the behavior under study form the **cases**.
- The participants who do not engage in the behavior are the **controls**.

# Definitions

- The **margin of error** is used to describe a **confidence interval** that is likely to contain the true population parameter.

A confidence interval is

from (sample statistic – margin of error)

to (sample statistic + margin of error).

# Confidence Interval

- A poll is conducted the day before a state election for Senator. There are only two candidates running for this office. The poll results show that 59% of the voters favor the Republican candidate, with a margin of error of 2 percentage points. Should the Republican expect to win?
- The results suggest that the Republican is likely to win a solid majority because he or she will most likely get between 57% and 61% of the vote.