STATISTICS
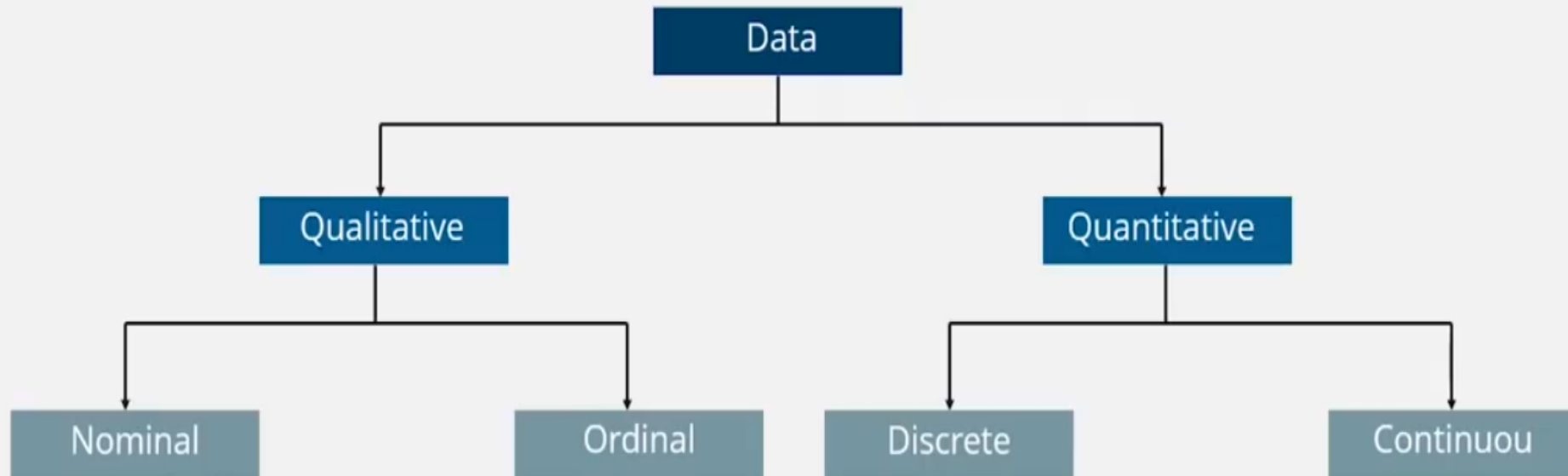
What is data?

Data refers to facts and statistics collected together for reference or analysis.

Categories of Data

Qualitative Data

Qualitative data deals with characteristics and descriptors that can't be easily measured, but can be observed subjectively.

Nominal Data

Data with no inherent order or ranking such as gender or race, such kind of data is called Nominal data

Gender
Male
Female
Male
Male

Ordinal Data

Data with an ordered series, such as shown in the table, such kind of data is called Ordinal data

Customer ID	Rating
001	Good
002	Average
003	Average
004	Bad

Quantitative Data

Quantitative data deals with numbers and things you can measure objectively.

Discrete Data	Continuous Data
Also known as categorical data, it can hold finite number of possible values.	Data that can hold infinite number of possible values.
Example: Number of students in a class	Example: Weight of a person

What is Statistics?

Statistics is an area of applied mathematics concerned with the data collection, analysis, interpretation and presentation.

Your company has created a new drug that may cure cancer. How would you conduct a test to confirm the drug's effectiveness?
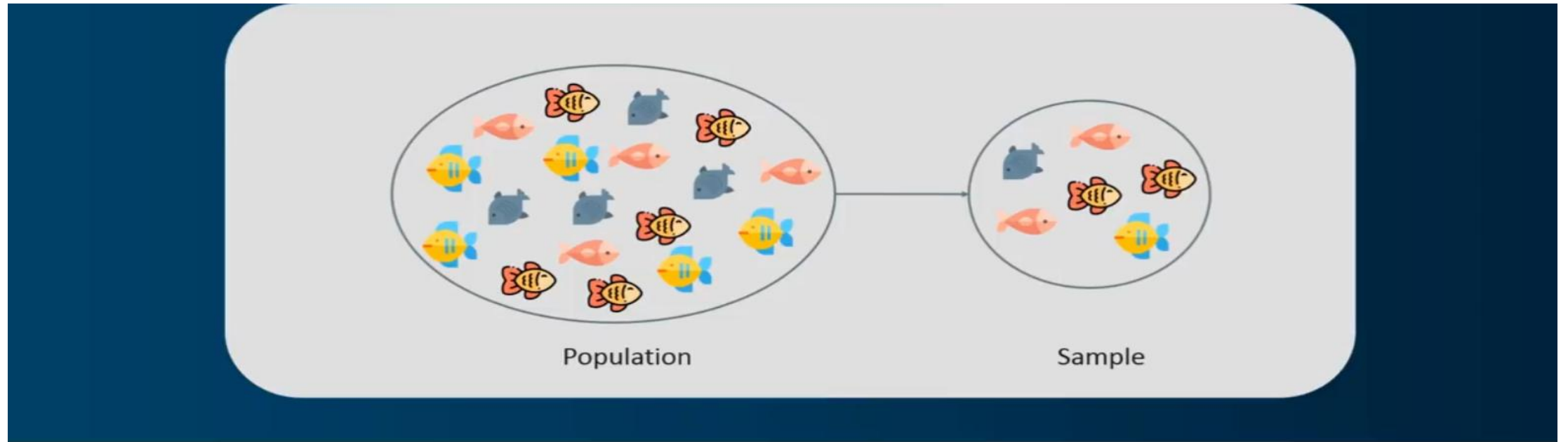
You and a friend are at a baseball game, and out of the blue he offers you a bet that neither team will hit a home run in that game. Should you take the bet?
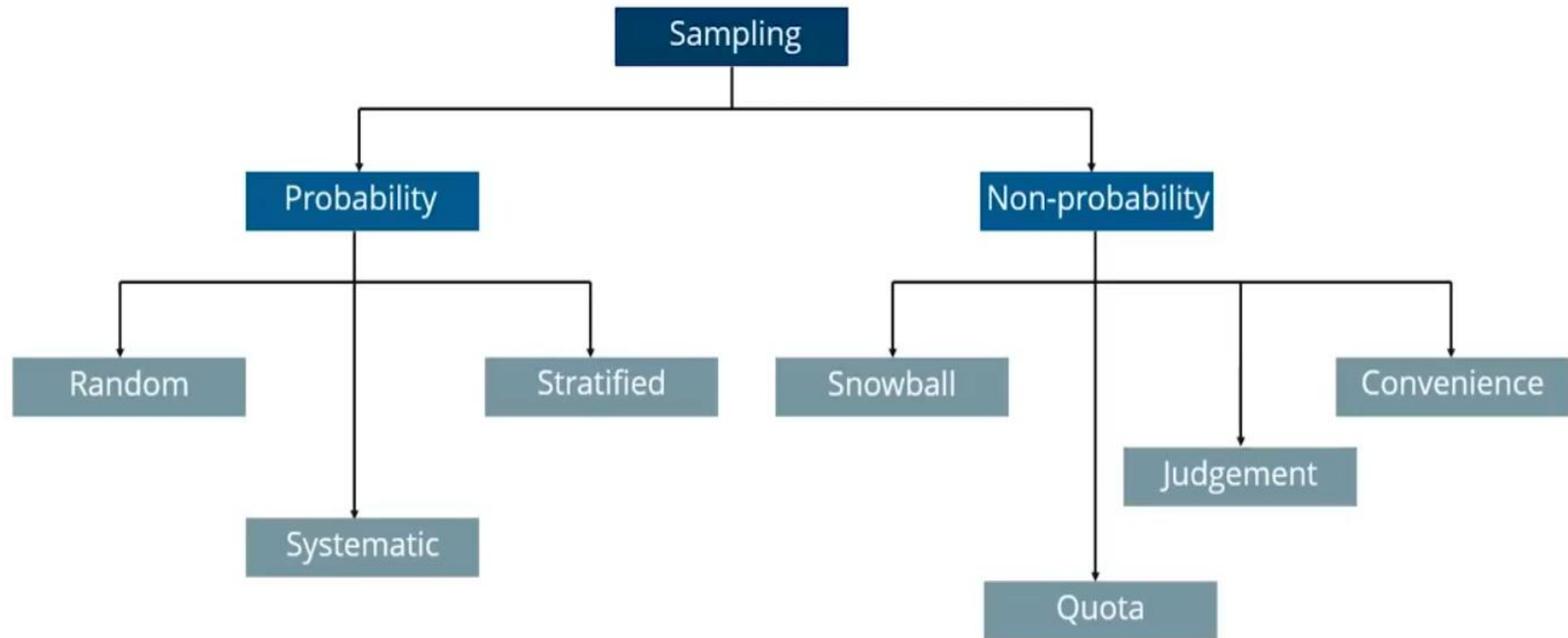
Terminologies in statistics

Population and Sample

Statistics Terminologies

Population: A collection or set of individuals or objects or events whose properties are to be analyzed.

Sample: A subset of population is called 'Sample'. A well chosen sample will contain most of the information about a particular population parameter

Sampling Techniques

Probability Sampling

- Probability sampling is based on the fact that **every member of a population has a known and equal chance of being selected**.

- For example, if you had a population of 100 people, each person would have odds of 1 out of 100 of being chosen. With non-probability sampling, those odds are not equal. For example, a person might have a better chance of being chosen if they live close to the researcher or have access to a computer. Probability sampling gives you the best chance to create a sample that is truly representative of the population.

Types of Probability Sampling

- Random sampling
- Stratified Sampling
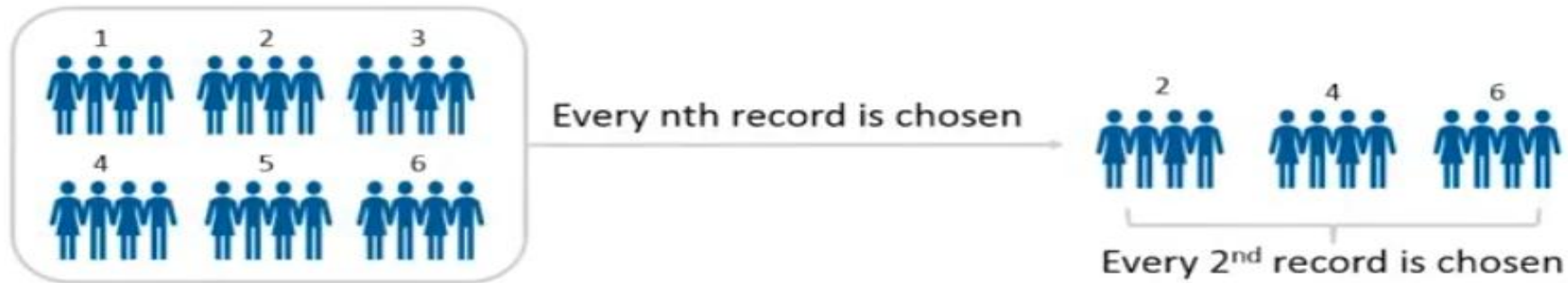- Systematic Sampling

Random Sampling

Each member of the population has equal chance of being selected in the sample.

- **random sampling** is a completely random method of selecting subjects. These can include assigning numbers to all subjects and then using a random number generator to choose random numbers.

- Classic ball and urn experiments are another example of this process (assuming the balls are sufficiently mixed). The members whose numbers are chosen are included in the sample.
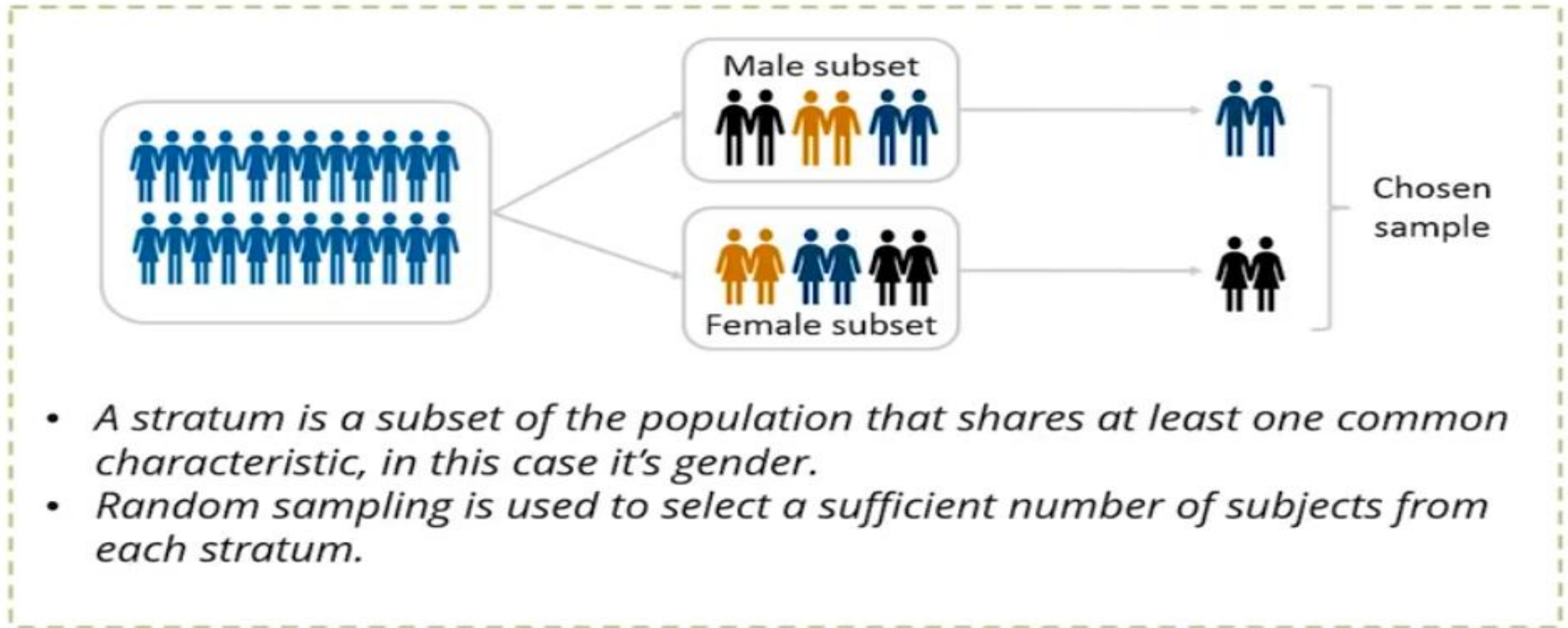
Systematic Sampling

Every nth record is chosen

Every 2nd record is chosen

In Systematic sampling every nth record is chosen from the population to be a part of the sample.

- **Systematic Sampling** means that you choose every "nth" participant from a complete list. For example, you could choose every 10th person listed.

Stratified Sampling

- A stratum is a subset of the population that shares at least one common characteristic, in this case it's gender.
- Random sampling is used to select a sufficient number of subjects from each stratum.

- **Stratified Sampling** involves splitting subjects into mutually exclusive groups and then using simple random sampling to choose members from groups.