

Introduction of Biostatistics

HERTANTO WAHYU SUBAGYO

Statistics

- **The science of :**
 - **collecting**
 - **summarizing**
 - **presenting**
 - **interpreting data,**
and of using them to test hypotheses.

Biostatistics

**Statistics in the area of biological and
health sciences**

The purpose of a biostatistics

- to provide the numbers / tables / graphics that contain information about a certain situation
- to present them in such a way that valid interpretations are possible

Increasing role of biostatistics

- Biostatistics provides a way of organizing information on a wider and more formal basis than relying on the exchange of anecdotes and personal experience
- More things are now measured quantitatively in medicine
- There is a great deal of intrinsic variation in most biological processes

To answer these questions, we rely on the methods of biostatistics

- Is the new drug effective + safe?
- Does the use of seat belt reduce the chance of death in motor vehicle accident?
- Where should government invest its resources if it wishes to reduce infant mortality?
- etc

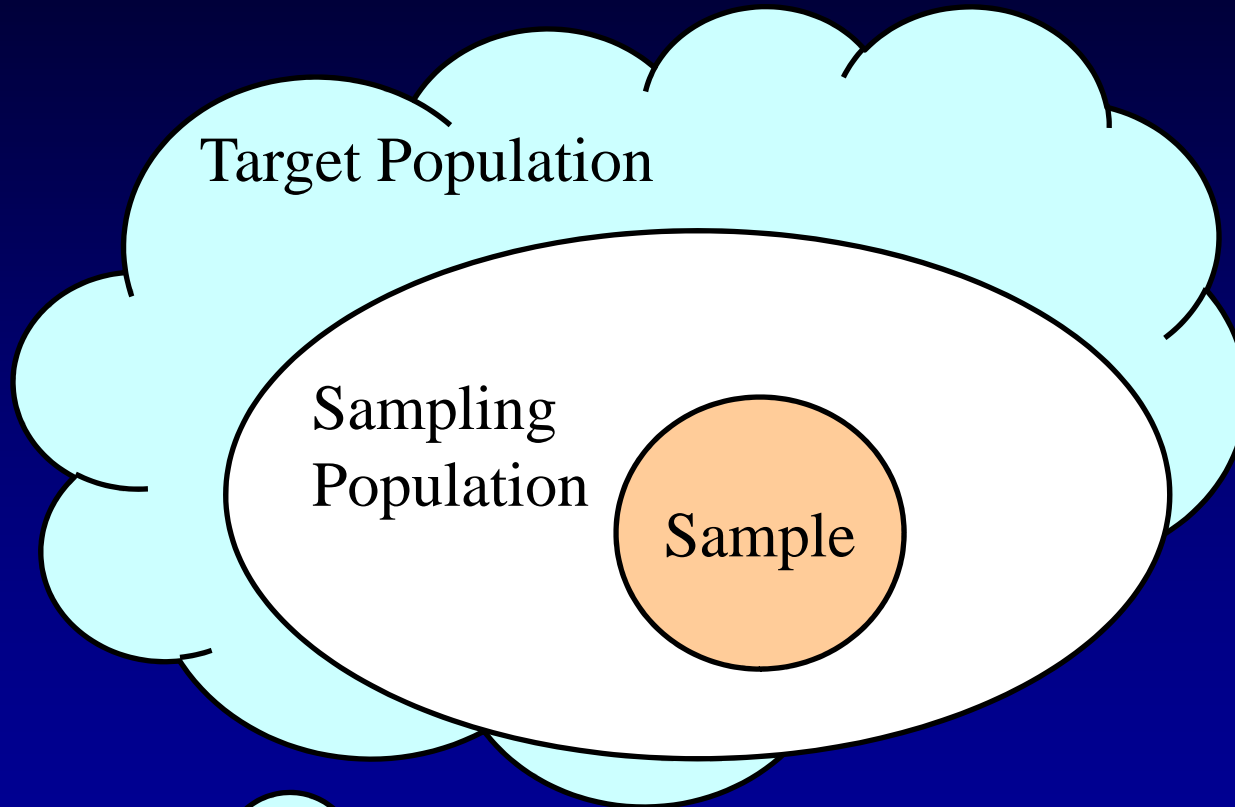
Population and samples

Except when a full census is taken, the data are for a *sample* from a larger group called the *population*.

The sample is of interest not in its own right, but for what it tells the investigators about the populations.

Because of chance, different samples give different results and this must be taken into account when using a sample to make inferences about the populations. This phenomenon, called sampling variation.

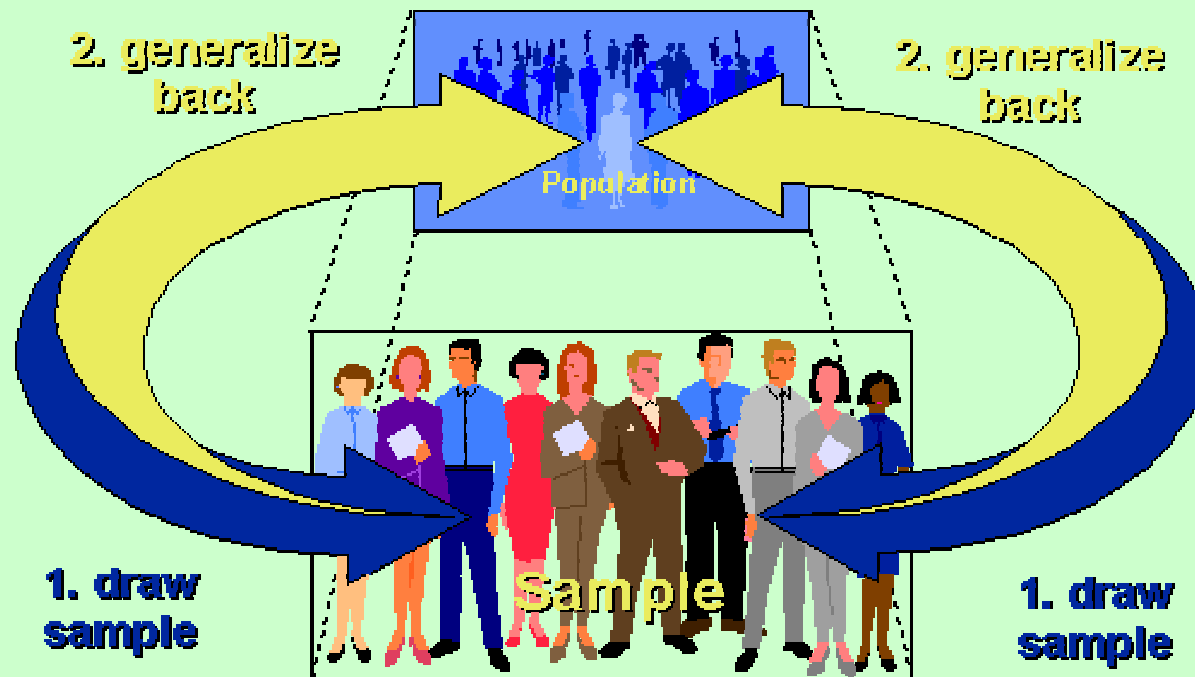
Sampling and representativity



Target Population → (Sampling Population) → Sample

Sampling

the process of selecting units from a population of interest



Sampling Model

Analysis

- By the time you get to the analysis of your data, most of the really difficult works have been done.
- It's much more difficult to :
define the research problem; develop and implement a sampling plan; conceptualize, operationalize and test your measurements; and develop a structure design.
- If you have done these works well, the analysis of the data is usually a fairly straightforward affair.

Data analysis involves three major steps

- Cleaning and organizing the data for analysis (Data Preparation)
- Describing the data (Descriptive Statistics)
- Testing Hypotheses and Models (Inferential Statistics)

Data Preparation

- logging the data
- checking the data for accuracy
- developing a database structure
- entering the data into the computer
- transforming the data

Logging the Data

In any research project data come from a number of different sources at different times:

- mail surveys returns
- coded interview
- laboratory
- etc

Checking the Data For Accuracy

- As soon as data are received you should screen it for accuracy.
- In some circumstances, doing this right away allows you to go back to the sample to clarify any problems or errors.
- There are several questions you should ask as part of this initial data screening:
 - ✓ Are the responses readable ?
 - ✓ Are all important questions answered?
 - ✓ Are the responses complete?
 - ✓ Is all relevant contextual information included (e.g. data, time, place, researcher)?

Developing a database structure

- **Defining variables**
- **Entering the data into the computer**
- **Data transformations**

Defining variables

- variable name
- variable description/label
- value labels
- missing values
- variable type (numeric, string, date etc)
- column format (width, alignment)
- measurement level (N O I R).

Entering the Data into the Computer

- There are a wide variety of ways to enter the data into the computer for analysis.
- In order to assure a high level of data accuracy, the analyst should use a procedure called **double entry**.

Data Transformations

- Recode
 - Compute
 - Select cases
 - Rank cases
- etc.

Modification of data files

- Opening an existing data file
- Defining new variables
- Entering new data
- Inserting and deleting cases and variables
- Saving data files

Analysis

- Descriptive
- Inferensial

Descriptive Statistics

- are used to describe the basic features of the data in a study.
- provide simple summaries about the sample and the measures.
- together with simple graphics analysis, they form the basis of virtually every quantitative analysis of data.
- with descriptive statistics you are simply describing what is, what the data shows.

Inferential Statistics

- investigate questions, models and hypotheses.
- we use inferential statistics to try to infer from the sample data what the population thinks.
- Thus, we use inferential statistics to make inferences from our data to more general conditions.

Softwares

- STATA
 - BMDP
 - EPIINFO
 - PEPI
 - **SPSS**
- etc