



Universitas Gadjah Mada  
Faculty of Engineering  
Department of Civil and Environmental Engineering

**Statistics and Probability**

# **Introduction**

# Statistics



## Merriam-Webster Dictionary

A branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data.

A collection of quantitative data.



## Simple notion

Statistics is the study of the collection, analysis, interpretation, presentation, and organization of **data**.

It is a mathematical discipline to collect, summarize data.

# Statistics

## Method to collect, summarize data

- The data are obtained from a repetitive operation
- Repetitive operation
  - observation
  - measurement
  - experiment



# Statistics

- Data
  - Quantitative data
  - Numerical data, numbers
  - Discreet
  - Continues
- Examples
  - You got 50; you are failed (in an exam)
  - Room occupancy is declining; it is now 40%
  - The budget leaks in 80% of regencies shows an alarming figure; its average is 35%, the highest is 55%, and the lowest is 15%

# Data

- Give examples of data in the following fields
  - Buildings
  - Geotechnics
  - Hydraulics
  - Meteorological data
  - etc.

# Data

- In the fields of engineering, and others too, data are used to define parameters; e.g. in modelling
- Data used to define model parameters shall be
  - homogeneous
  - representative

# Data

- Causes of data inhomogeneity
  - measurement station has been moved to another site
  - stream shifts to another location
  - construction of new structures (dam, levee, reservoir)
  - change of landuse
  - etc.

# Data

- Causes of unrepresentative data
  - data were obtained during particular seasons
  - insufficient number of data (short data series)



# Use of Statistics

- Descriptive statistics
- Statistical inference

# Descriptive Statistics & Statistical Inference

## ■ Example #1

- Result of an interview with **respondents randomly selected** from Parangtritis villages shows that 45% of them have sufficient awareness to tsunami hazard.
- A campaign has been recently launched to educate people in Parangtritis on tsunami hazard.
- Result of an interview, conducted after the campaign, with respondents from Parangtritis villages shows that 56% of them have sufficient awareness to tsunami hazard.

# Descriptive Statistics & Statistical Inference

## ■ Question

- Can we say that the increase in the number of people in Parangtritis who has sufficient awareness on tsunami hazard is due to the campaign?
- If we question on the claim that the campaign has increased the number of people who has sufficient awareness on tsuname hazard, then what shall we say on the data of both interviews.
- What about people in Parangtritis who were not inteviewed?

# Descriptive Statistics & Statistical Inference

- Descriptive statistics
  - Use of 45% to show part of people in the first respondents who has sufficient awareness on tsunami hazard.
  - In this case, “45%” was used to summarize the fact (data) that out of 140 people in the first respondents, there were 63 ones who have sufficient awareness on tsunami hazard.
  - Therefore, it can be said that
    - descriptive statistics is use of numbers to summarize known information on a particular state.

# Descriptive Statistics & Statistical Inference

- Statistical inference
  - If we use the result of the first interview to conclude that about 45% of the people in Parangtritis villages have sufficient awareness on tsunami hazard, this means that
    - we apply number to say something on a population, of which we do not have thorough information on that population.
  - Therefore, it can be said that
    - statistical inference is the use of number to conclude about information on a population based on information on sample taken from that population.

# Descriptive Statistics & Statistical Inference

- Example #2
  - Give examples of the use of statistics that you know.
  - Discuss on
    - descriptive statistics,
    - statistical inference.

# Descriptive Statistics & Statistical Inference

- Things commonly faced in statistics
  - How many number of respondents?
  - How to select respondent?

# Sample and Population

- What is sample
  - Sample is a compilation of objects (data) taken from a larger compilation of the objects.
  - The larger compilation of objects, if it contains all objects possible, then it is called population.

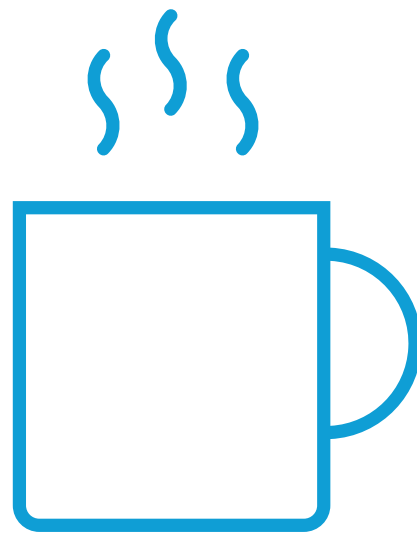


# Sample and Population

- What is sample?
  - Based on sample, one can draw conclusion on the population from which the sample was taken.
  - Statistics is a tool that enables one to draw valid conclusion on the population based on information contained in the samples (taken from that population)

# Random Sample

- Random sample
  - If every element in the population has the same probability to be included in the sample
  - Sample that is selected such that another sample taken from the same population will have similar characteristics
  - Nevertheless, random samples do not guarantee that they are representative of the population where they were taken from



# Statistics and Probability

Introduction