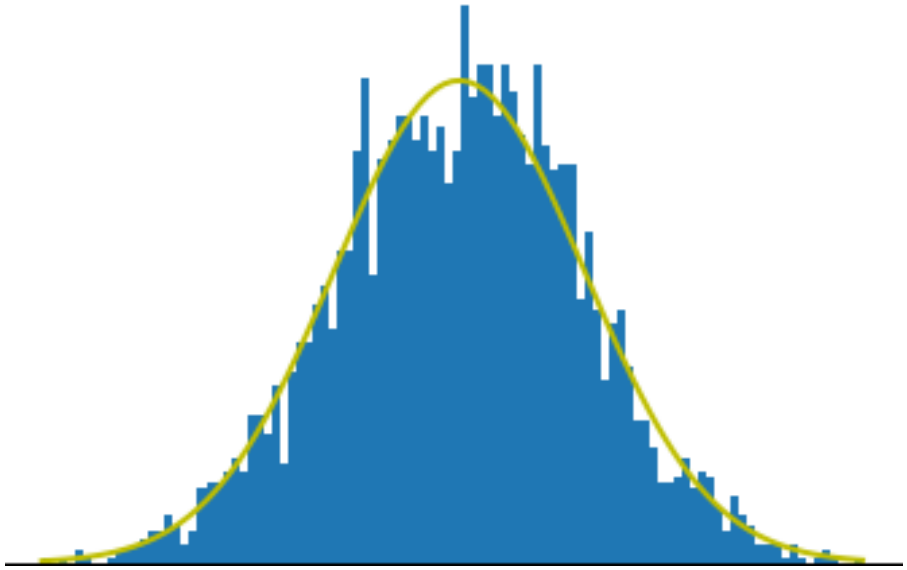
Probability and Statistics

ITS
Institut
Teknologi
Sepuluh Nopember

Introduction to Statistics



Why do we need statistics?

- What kind and how much data need to be collected?
- How should we organize and summarize the data?
- How can we analyse the data and draw conclusions from it?
- How can we assess the strength of the conclusions and evaluate their uncertainty?

Learning Outcome

By the end of this lecture you should be able to:

- Understand the difference between descriptive and inferential statistics
- Looking for a point estimate, interval estimate, margin of error, and level of confidence

Overview of Topics Covered

- Inferential Statistics: Estimation
- Inferential Statistics: Hypothesis Testing
- Correlation
- ANOVA

Useful Textbooks

- “Probability & Statistics for Engineers & Scientists”, by Ronald E. Walpole, Raymond Myers, Sharon Myers, Keying Ye
- “Probability and Statistics for Engineering and The Sciences” by Devore, Jay L. (ITS Library No 3100003052254)
- “Introduction to the Practice of Statistics”, Sixth Edition, by David S.Moore, George P. McCabe, and Bruce A. Craig

Statistics: Introduction

Statistics is a methods for collecting, analysing, organizing, and presenting data

Statistics provides methods for:

1. Design: Planning and carrying out research studies.
2. Description: Summarizing and exploring data.
3. Inference: Making predictions and generalizing about phenomena represented by the data.

Descriptive vs Inferential Statistics

Descriptive statistics uses the data to provide descriptions of the population, either through numerical calculations or graphs or tables.

Inferential statistics makes inferences and predictions about a population based on a sample of data taken from the population in question.

Can you tell the difference?

- The following line graph presents the number of students enrolled in Biology Class between 2015 and 2017

  (Descriptive)

- An engineer tested one out of every 20 units of LEDs manufactured to see if the LED meet the specification as stated

  (Inferential)

Can you tell the difference?

- The average age of a group of 10 student is 21

  (Descriptive)

- The Canadian government reports that the population of canada was 18,238,000 in 1961

  (Descriptive)

- A study found that children with light-coloured eyes are likely to have parents with light-coloured eyes

  (Inferential)

Inferential Statistics

Inferential Statistics draws conclusions / predictions about a **population** / process based on **sample** data.

Sample is randomly chosen from the population (**Sampling**)

There are some types of inferential statistics. In this lecture, we introduce two most common types:

1. Estimation
2. Testing

Estimation: an Introduction (1/2)

Statistical inference uses sample data to form two types of **estimators** of parameters.

A **point estimate** consists of a single number, calculated from the data, that is the best single guess for the unknown parameter.

An **interval estimate** consists of a range of numbers around the point estimate, within which the parameter is believed to fall.

Estimation: an Introduction (2/2)

A confidence interval estimate of a parameter consists of an interval of numbers obtained from a point estimate of the parameter together with a percentage that specifies how confident we are that the parameter lies in the interval.

The confidence percentage is called the confidence level.

i.e. 90%, 95%, 99%

Estimation: Point Estimate

The object of point estimation is to calculate, from the sample data, a single number that is likely to be close to the unknown value of the population parameter.

Usually, there are two obvious ways to get point estimate, by its sample mean or sample standard deviation

Confidence Interval

When population variable X is normally distributed and σ is known, a 100(1 − α)% confidence interval for μ is given by

$$\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

When α = 95%

$$\left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right)$$

Creating an Estimate

1. Decide on the level of confidence 90%, **95%**, 99%
2. Gather data to get a sample mean or sample standard deviation
3. Come up with the margin of error