1st Material Subject: Statistics Data

Undergraduate of Telecommunication Engineering

MUH1F3 - PROBABILITY AND STATISTICS

Telkom University
Center of eLearning & Open Education Telkom University
Jl. Telekomunikasi No.1, Bandung - Indonesia
http://www.telkomuniversity.ac.id

Lecturer: Nor Kumalasari Caecar Pratiwi, S.T., M.T. (caecarnkcp@telkomuniversity.ac.id)

السلام عليكم ورحمة الله وبركاته

WELCOME

TABLE OF CONTENTS:

LEARNING OBJECTIVES:

After careful study of this chapter, student should be able to do the following:

1. **Explain the definition of Statistics, Population, Sample and Data**
2. **Identify the classification of Data**
3. **Explain the difference between of Properties Numerical Data**

STATISTICS

Statistics is a science that helps us make decisions and draw conclusions in the presence of variability. The field of statistics deals with the collection, presentation, analysis, and use of data to make decisions, solve problems, and design products and processes. **In simple terms, statistics is the science of data**. Because many aspects of engineering practice involve working with data, obviously knowledge of statistics is just as important to an engineer as are the other engineering sciences. **Finally, Statistics is the Art of Drawing Conclusions from Imperfect Data**. Statistics are divided into two large clusters, they are:

- **EXPLORATORY DATA ANALYSIS (DESCRIPTIVE STATISTICS)**
- **CONFIRMATORY DATA ANALYSIS (STATISTICAL INFERENCE / INFERENTIAL STATISTICS)**

STATISTICS

- **EXPLORATORY DATA ANALYSIS (DESCRIPTIVE STATISTICS)**
  - Data Collecting
  - Data Presenting
  - Knowing the characteristics of Data
- **CONFIRMATORY DATA ANALYSIS (STATISTICAL INFERENCE / INFERENTIAL STATISTICS)**
  - Parameter Estimation
  - Hypothesis Testing
  - Withdrawal of Conclusions

THE POPULATION AND SAMPLE

In a statistical study, begin with statistics term **POPULATION**. The world ”population” comes from the early use of statistics in census counts. Census is a sample consisting of the entire population. The population can be defined as:

- The entire collection of objects or outcomes about which information is sought
- A set of data that characterizes some phenomenon
- The set of all items of interest in a statistical problem.
- The totality of all possible elements of interest in an investigation
- The entire groups of people, animals or things about which we want information
- The collection of all things under study

THE POPULATION AND SAMPLE

However, in the engineering environment, the data are almost always a sample that has been selected from the population. The **SAMPLE** can be defined as:

- A subset of data selected from population.
- A part of the population which we actually collect information, used to draw conclusions about the whole
- The set of all elements of a population that have been observed.

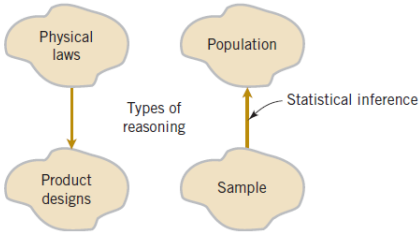

Figure 1: Statistical Inference between Population and Sample

DATA CLASSIFICATION

In common usage, the Descriptive Statistics deals with the tabulation of **DATA**, its representation in graphical and the calculation of numerical descriptions. The objective of statistics is to make inferences about unknown population parameters based on information contained in sample data. The **DATA** can be defined as:

- Numbers / measurements that are collected as a result of observations.
- Numbers, but they are not just numbers, Data are numbers with a context.

DATA can be measured on one of four scales:

- **NOMINAL**, the lowest level, because there is no order to the data.
- **ORDINAL**, considered to be a higher scale than the nominal scale.
- **INTERVAL**, having a meaning foulness in the separation between any two numbers on the scale.
- **RATIO**, The ratio scale represents the highest level of measurement. the ratio scale has a true zero point as its origin.

DATA CLASSIFICATION

Table 1: Characteristic of Four Class Data

Data Scale	Characteristics	Example
Nominal	Unordered Category	Sex of object: Male = 1 Female = 2 Religion of Subject
Ordinal	Ordered Category	Rank in Class (1st, 2nd, 3rd Rank in Personal Test (Low, High)
Interval	Order, equal intervals, and arbitrary zero point	Temperature Raw Score on Statistics Test
Ratio	Order, equal intervals, and real zero point	Height, Weight, Age

DATA CLASSIFICATION

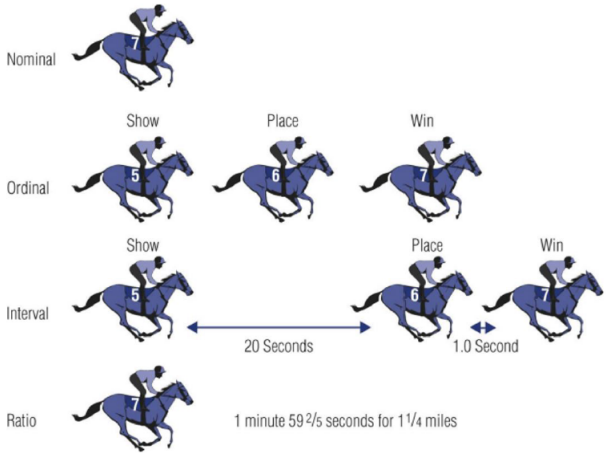

Figure 2: Four Class of Data Illustration

PROPERTIES OF NUMERICAL DATA

- **Central Tendency**
  - Mean
  - Median
  - Mode
  - Mid-range
  - Mid-hinge
- **Variation**
  - Range
  - Interquartile Range
  - Variance
  - Standard Deviation
  - Coefficient of Variance
- **Shape**
  - Skewness
  - Kurtosis
  - Interquartile Range
  - Variance
  - Coefficient of Variance
- **Location**

MEASURES OF CENTRAL TENDENCY

1. **MEAN**

   The mean acts as a balancing point of distribution. The arithmetic mean is the sum of the values divided by the number of values:

$$\bar{\mathbf{X}} = \frac{\sum_{\mathbf{i}=1}^{\mathbf{n}} \mathbf{x_i}}{\mathbf{n}} \qquad (1)$$

2. **MODE**

   Mode is the French word for fashion, the mode of a distribution is defined as the most frequent value.

3. **MEDIAN**

   The median is the middle value in an ordered sequence of data. Let $\mathbf{X_1}, \mathbf{X_2}, \cdots, \mathbf{X_n}$ denote sample arranged in increasing order of magnitude, then the median defines as:

$$\widetilde{\mathbf{X}} = \begin{cases} X_{\frac{n+1}{2}} & \text{, for n odd} \\ \frac{X_{\frac{n}{2}} + X_{\frac{n+1}{2}}}{2} & \text{, for n even} \end{cases} \qquad (2)$$

MEASURES OF CENTRAL TENDENCY

4. **MID-RANGE**

The mid-range is the average of the smallest and largest observations in a set of data.

$$\textbf{MIDRANGE} = \frac{\mathbf{X_{smallest}} + \mathbf{X_{largest}}}{\mathbf{2}} \quad (3)$$

5. **MID-HINGE**

The mid-hinge is the average of the first and third quartiles in a set of data.

$$\textbf{MIDHINGE} = \frac{\mathbf{Q_1} + \mathbf{Q_3}}{\mathbf{2}} \quad (4)$$

Where $\mathbf{Q_i}$ defines as:

$$\mathbf{Q_i} = \textbf{Data sequence to} \frac{\mathbf{i(n+1)}}{\mathbf{4}} \quad (5)$$

MEASURES OF VARIANCE

1. **RANGE**

   The difference between the largest and smallest observations data.

$$\text{RANGE} = X_{\text{largest}} - X_{\text{smallest}} \tag{6}$$

2. **INTERQUARTILE RANGE (IQR)**

$$\text{IQR} = Q_3 - Q_1 \tag{7}$$

Semi-interquartile range is:

$$\text{SIQR} = \frac{Q_3 - Q_1}{2} \tag{8}$$

3. **VARIANCE AND STANDARD DEVIATION**

Variance is the average of the squared differences between each observations in a set of data and the mean.

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1} \tag{9}$$

The sample standard deviation, given by the-symbol **S** is the square root of the sample variance, expressed as follows :

$$S = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}} \tag{10}$$

EXAMPLE

There are observational data as follows:

16 21 22 11 16 18 17 12 13

Determine the Mean, Mode, Median,The Quartile (Q_1, Q_2, Q_3), Mid-Range, Mid-Hinge, Range, IQR, SIQR, Variance and Standard Deviation!

Answer:

It appears that $\mathbf{n = 9}$, then sort the data from smallest to largest

11 12 13 16 16 17 18 21 22

- The Mean

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{11 + 12 + 13 + 16 + 16 + 17 + 18 + 21 + 22}{9} \frac{146}{9} = 16.22$$

- The Mode

  The most frequently out **Mode** $= 16$

- The Median equal to Q_2

$$\mathbf{Q_2 = Data\ sequence\ to \frac{2(n+1)}{4} = Data\ sequence\ to \frac{2(10)}{4} = 5}$$

So, the location of the median is in the fifth data, $\mathbf{Median = Q_2 = 16}$

- The Quartile (Q_1, Q_2, Q_3)

$$\mathbf{Q_1 = Data\ sequence\ to \frac{1(n+1)}{4} = Data\ sequence\ to \frac{1(10)}{4} = 2.5}$$

So, the location of the first quartile is in the 2.5 data, $\mathbf{Q_1 = 12.5}$

$$\mathbf{Q_3 = Data\ sequence\ to \frac{3(n+1)}{4} = Data\ sequence\ to \frac{3(10)}{4} = 7.5}$$

So, the location of the third quartile is in the 7.5 data, $\mathbf{Q_3 = 19.5}$

- The Mid-Range

$$\text{MIDRANGE} = \frac{X_{\text{smallest}} + X_{\text{largest}}}{2} = \frac{11 + 22}{2} = 16.5$$

- The Mid-Hinge

$$\text{MIDHINGE} = \frac{Q_1 + Q_3}{2} = \frac{12.5_1 9.5}{2} = 16$$

- The Range

$$\text{RANGE} = X_{\text{largest}} - X_{\text{smallest}} = 22 - 11 = 10$$

- The IQR

$$\text{IQR} = Q_3 - Q_1 = 19.5 - 12.5 = 7$$

- The SIQR

$$\text{IQR} = \frac{Q_3 - Q_1}{2} = \frac{7}{2} = 3.5$$

- The Variance

X_i	11	12	13	16	16	17	18	21	22	
$X_i - \bar{X}$	-5.22	-4.22	-3.22	-0.22	-0.22	0.78	1.78	4.78	5.78	
$(X_i - \bar{X})^2$	27.27	17.83	10.38	0.05	0.05	0.6	3.16	22.83	33.38	$\sum = \mathbf{115.55}$

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1} = \frac{155.55}{8} = 14.44$$

- The Standard Deviation

$$S = \sqrt{S^2} = \sqrt{14.44} = 3.8$$

Thank You