

DUM 2413 STATISTICS & PROBABILITY

# CHAPTER 1

## INTRODUCTION TO STATISTICS

PREPARED BY:

DR. CHUAN ZUN LIANG; DR. NORATIKAH ABU; DR. SITI ZANARIAH SATARI  
FACULTY OF INDUSTRIAL SCIENCES & TECHNOLOGY  
chuanzl@ump.edu.my; atikahabu@ump.edu.my; zanariah@ump.edu.my

# EXPECTED OUTCOMES

- Able to define the basic terminologies of statistics
- Able to differentiate the two branches of statistics, types of data and variables
- Able to apply of basic probability sampling techniques in solving application problems
- Able to differential between the observational and experimental studies

# CONTENT

1.1 OVERVIEW

1.2 DESCRIPTIVE AND INFERENCE STATISTICS

1.3 VARIABLES AND TYPES OF DATA

1.4 DATA COLLECTION AND SAMPLING TECHNIQUES

1.5 OBSERVATIONAL AND EXPERIMENTAL STUDIES

# 1.1 OVERVIEW

## 1.2

# DESCRIPTIVE AND INFERENTIAL STATISTICS

# WHAT IS “STATISTICS”?

## DEFINITION

**STATISTICS** represent *scientific procedures and methods for collecting, organising, summarising, presenting and analysing, as well as obtaining useful information, drawing valid conclusions and making effective decisions based on the analysis*

## TYPES

### DESCRIPTIVE STATISTICS

- \* *A procedures used to summarize, organize, and make sense of a set of scores, measurements or observations , which known as data*
- \* *Typically presented graphically*
- \* *Do not attempt to make prediction and draw conclusions about the population from which the sample are taken*

### INFERENCE STATISTICS

- \* *A procedures used that allow to infer, estimate, approximate, determination of association (relationship) between two variables, generalise observations make with sample to population from which they were selected.*
- \* *Concerned with making prediction/ inference and drawing conclusions about a population, based on sample*

# WHY WE NEED “STATISTICS”?

## REASON

Describing relationship between two variables

Aiding in decision making

# EXERCISE 1.1

**Descriptive statistics** is a branch of mathematics *dealing with summarisation and a description of the collections of data sets*. Conversely, **inferential statistics** is a branch of mathematics that involves *dealing conclusions about a population based on sample data drawn from it*. Given the statement below, identify types of statistics have been applied.

- i. The market analyst **claimed** that mortgage rates is hit bottom due to the economic crisis.

Inferential statistics

- ii. There is a **significant relationship** between smoking and lung cancer.

Inferential statistics

- iii. The **average** age of the students in a statistics & probability class is 19 year olds.

Descriptive Statistics

# EXERCISE 1.1-CONTINUE

- iv. The **chances** of won the grand prizes of a lucky draws are one chance in twenty-two million.

Inferential Statistics

- v. The **median** monthly income for male employees aged 25-34 is RM 2000.

Descriptive Statistics

- vi. Based on the previous reports, it was found that the national **average** annual medicine expenditure per person is RM 4208.

Descriptive Statistics

- vii. By **2070**, the unemployment rates of developing country will be reduced to 2.8%.

Inferential Statistics



# WHAT IS “SAMPLE AND POPULATION”?

## POPULATION

- *All elements, individuals, or units* that meet the selection criteria for a group to be studied

## SAMPLE

- *A set of data collected and/or selected from a statistical population* by a defined procedure.

### POPULATION

- A summary measure for entire population is referred to a “PARAMETER”

### SAMPLE

- A summary measure for a sample data is referred to a “STATISTIC”

### SAMPLING TECHNIQUE (SECTION 1.4)

### Types of population

#### Tangible population (Countable/Finite)

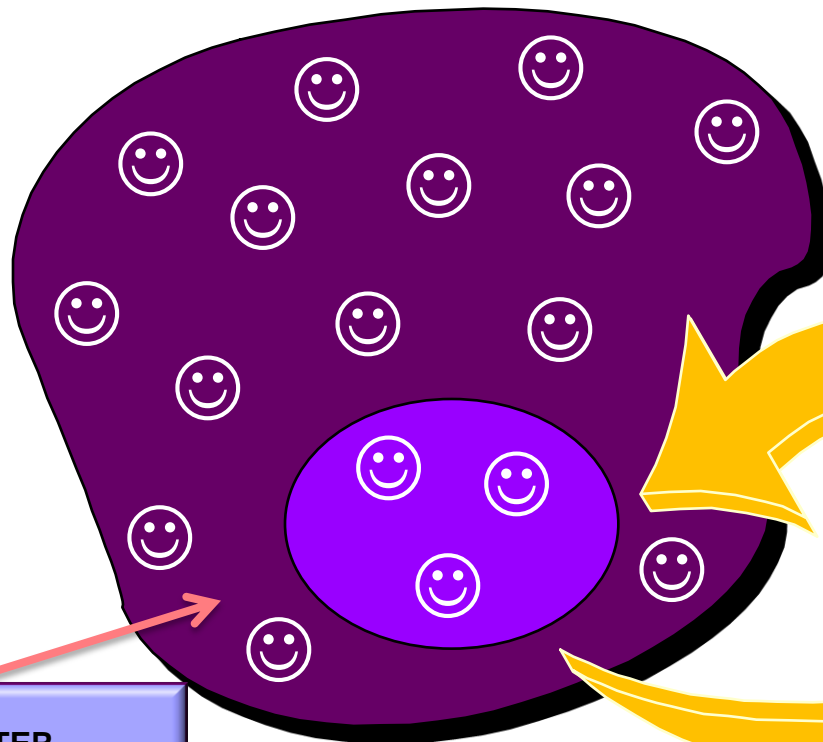
- It is *possible to count* the units contained in the population.
- E.g. : The number of vehicles crossing a bridge everyday, the number of births per years, the number of words in a book.

#### Conceptual population (Uncountable/Infinite)

- It is *impossible to count* the units contained in the population.
- E.g. : The number of germs in the body of a patient of malaria, the number of stars in a sky.

## POPULATION

## RANDOM SAMPLE



### STATISTIC

- Sample size :  $n$
- Mean :  $\bar{x}$
- Variance :  $s^2$
- Standard Deviation :  $s$
- Proportions :  $p$

Inference

Sampling

### PARAMETER

- Population size :  $N$
- Mean :  $\mu$
- Variance :  $\sigma^2$
- Standard Deviation :  $\sigma$
- Proportions :  $\pi$

# EXERCISE 1.2

In principle, the *numerical characteristic of the population* is known as a *parameter*, whereas the *numerical characteristic of the sample* is known as *statistic*. For each statement below, identify whether the parameter or statistic has been applied.

- i. 36 passengers on the bus had been killed during the road accidents occurred last week.

Parameter

- ii. According to the census data 2010, it was found that the median age of the Malaysian population 26.3 year olds.

Parameter

- iii. The average of the monthly electricity consumption in a community is found to be 901 kWh per month.

Statistic

# EXERCISE 1.2

- iv. In a state of Malaysia, there are 500 walk buttons that pedestrians can be pressed at traffic intersections. However, it found that 10% these buttons do not function.

Parameter

- v. A theatre owner randomly selected 40 movies and he found annual profit from ticket sales is RM 500.00 million.

Statistic

- vi. The lecturer of Statistics & Probability randomly selected 35 out of 60 students and measured their height. The analysis results show that the average height of the student is 165cm.

Statistics

- vii. According the report of Department of Statistics Malaysia, it was found that the average salary and wages of male employees is RM2345, while RM 2254 for female employees.

Parameter

# 1.3

## VARIABLES AND TYPES OF DATA

# DATA CLASSIFICATION

## VARIABLE

Any characteristic or property of the population of interest  
E.g. : The average monthly income per household in Malaysia, level of education and age of residents in a community, etc.

## QUANTITATIVE OR NUMERICAL

- \* Measures on the **numeric scale**
- \* Resulted a numeric response  
*E.g.: What is your age?  
The answer is a numeric.*

## QUALITATIVE OR ATTRIBUTE

- \* Measures on the **non-numeric scale**
- \* Resulted the categorical response  
*E.g.: What is your nationality?  
The answer is a non-numeric.*

## DISCRETE

- \* The numeric responses which arises from **counting process**  
*E.g.: How many balls in the basket?*

## CONTINUOUS

- \* The numeric response which arises from a **measuring process**  
*E.g.: What is your weight and height?*

# EXERCISE 1.3

The variables can be allocated to *quantitative* and *qualitative* variables. Determine the type of variable for the following variable as quantitative or qualitative variables. *If the variable is a quantitative, classify the variable as discrete data or continuous data.*

- i. The number of parking lots in Kuantan city.

Quantitative; Discrete

- ii. The numbers sewn on the shirts of the 5 football players are 5, 11, 4, 19 and 21 .

Qualitative

- iii. The total number of fresh milk can be produced in a dairy farm.

Quantitative; Discrete

# EXERCISE 1.3

- iv. A physician was randomly selected 15 women in a community. He found that average of systolic blood pressure among them is 61 mm Hg.

Quantitative; Continuous

- v. The achievement of students' for every subject will be evaluated by using grades of A, A-, B+, B, B-, C+, C, C-, D+, D, E or F.

Qualitative

- vi. The critic's ratings for a new released movie are must watch, recommended, not recommended and don't even think about going.

Qualitative

- vii. The volume of cola in a can of regular Coke is 325ml.

Quantitative; Continuous



# 4 Levels of Measurement of Data

## NOMINAL-LEVEL

- \* Categorised the data only regardless the arrangement in order.
- \* E.g.: race, gender, nationality, marital status, or other demographic or personal information

QUALITATIVE

## ORDINAL-LEVEL

- \* Categorised and order the data. However, differences among categories are meaningless
- \* E.g.: finishing order in a competition, education level, world universities ranking

QUALITATIVE

## INTERVAL SCALE

- \* The differences among categories are meaningful. However, there is no natural starting point and ratios among categories are meaningless
- \* E.g.: temperature (zero means no temperature), measurements of latitude and longitude

QUANTITATIVE

## RATIO SCALE

- \* Opposite for interval scale, namely there are natural zero stating point and the ratios among categories are meaning full.
- \* E.g.: length, height, weight and time.

QUANTITATIVE

# EXERCISE 1.4

There are four typical levels of measurements of variables, namely *nominal-level*, *ordinal-level*, *interval-level* and *ratio-level scale*. Determine the levels of measurements for the following variable.

- i. The *number of electronic components* manufactured in a production line.

Ratio-level scale

- ii. The *body temperature* inside a refrigerator.

Interval-level scale

- iii. The *level of satisfaction* of students to their statistics lecturer.

Nominal-level scale

- iv. The *ages* of the residents in a community.

Ratio-level scale

# EXERCISE 1.4

- v. The *age brackets (18-34, 35-50, 51-69 and 70-87)* for generational labels as millennial, Gen-X, Boomer and Silent.

Ordinal-level scale

- vi. The *types of dramas* such as Hong Kong drama, Korean drama, Thailand drama, etc.

Nominal-level scale

- vii. The years of Southwest Asian Sea Games held.

Interval-level scale

# 4 BASIC PROBABILITY SAMPLING TECHNIQUES

**SAMPLING TECHNIQUES**  
*A scientific methods of selecting samples from populations*

**Probability sampling**

**Non-probability sampling**

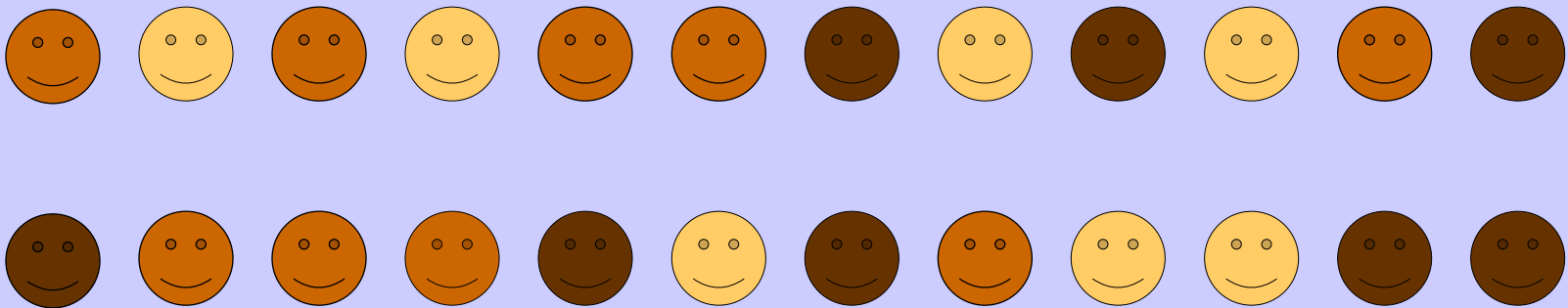
- \* Simple Random Sampling
- \* Systematic Sampling
- \* Stratified Random Sampling
- \* Cluster Sampling

The non-probability sampling technique did not discussed here

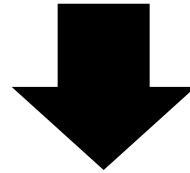
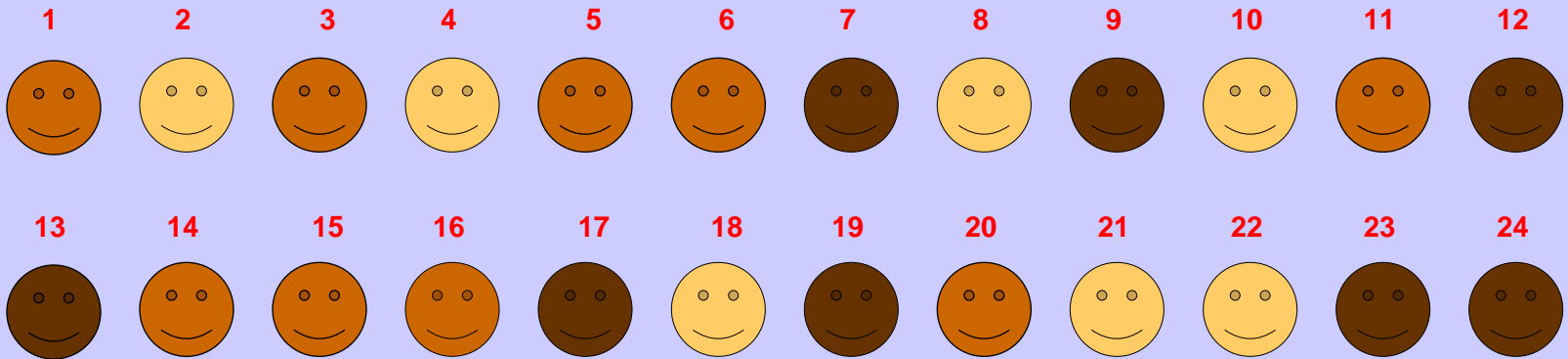
# EXERCISE 1.4

Suppose a population have 24 people ( $N = 24$ ), select a set of sample of size 12 people from the population by using ( $n = 12$ ),

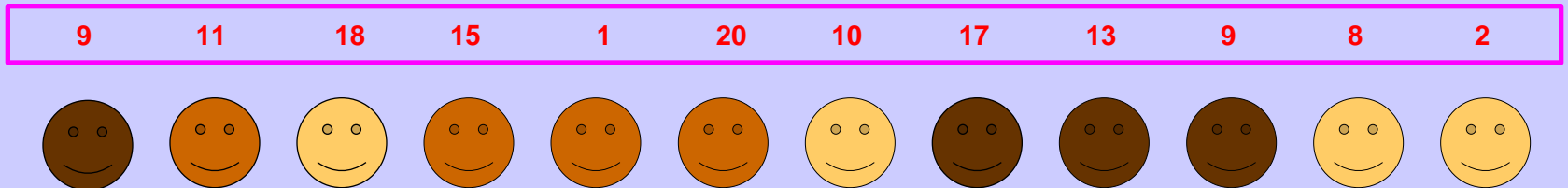
- ❖ **Simple random sampling technique**
- ❖ **Systematic sampling technique**
- ❖ **Stratified random sampling technique**
- ❖ **Cluster sampling technique**



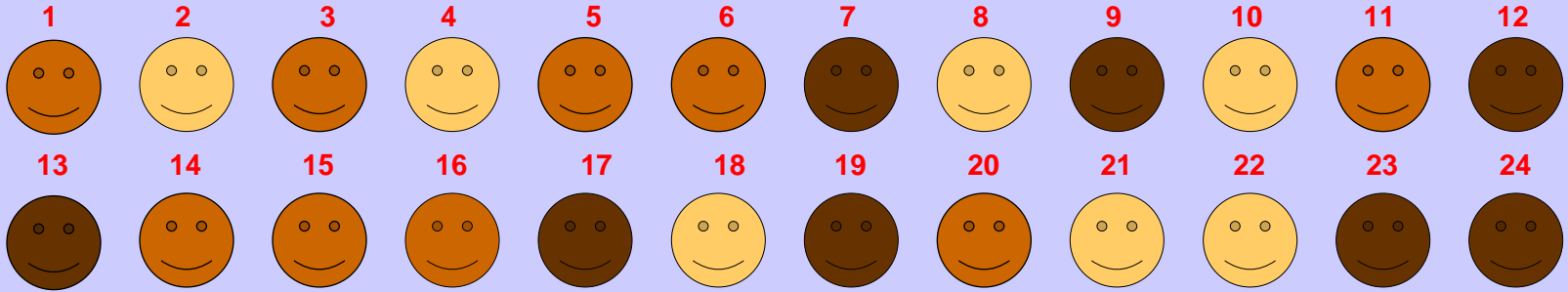
# SIMPLE RANDOM SAMPLING TECHNIQUE



## RANDOM NUMBER



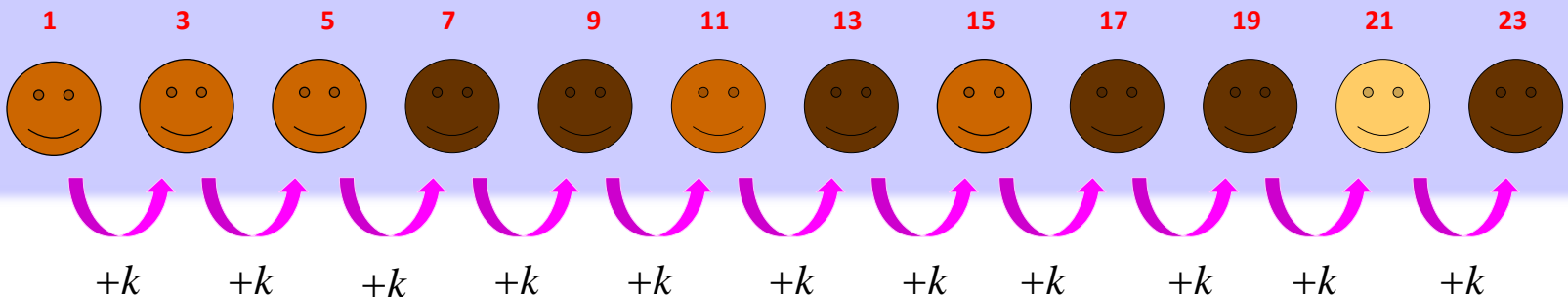
# SYSTEMATIC SAMPLING TECHNIQUE



POPULATION

First sample is randomly generate between 1 and 2 since  $k = 2$ .

$$\text{Calculate } k = \frac{\text{Population size}}{\text{Sample size}} = \frac{N}{n} = \frac{24}{12} = 2$$

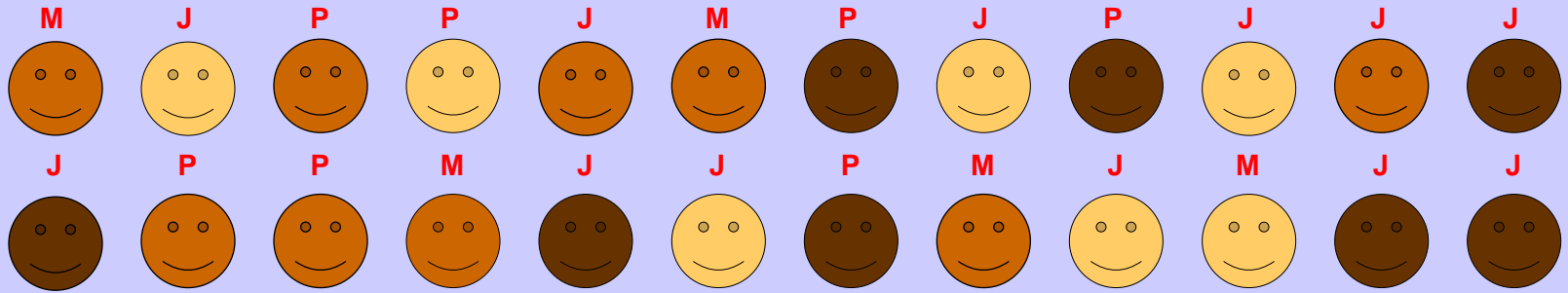


SAMPLE

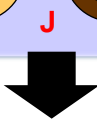
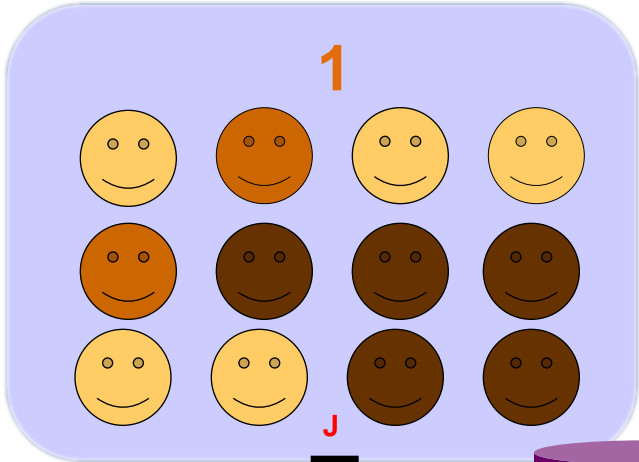




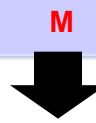
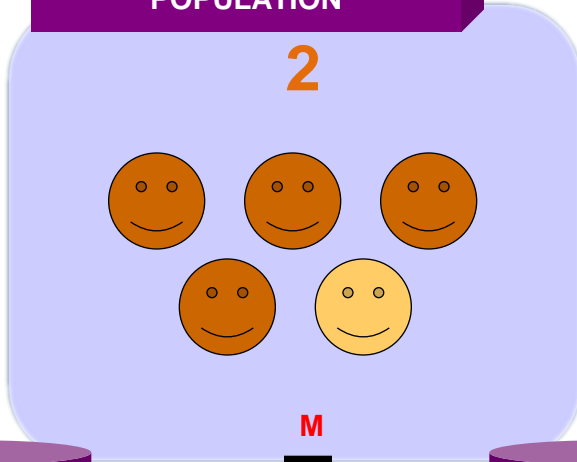
# CLUSTER SAMPLING TECHNIQUE



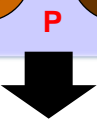
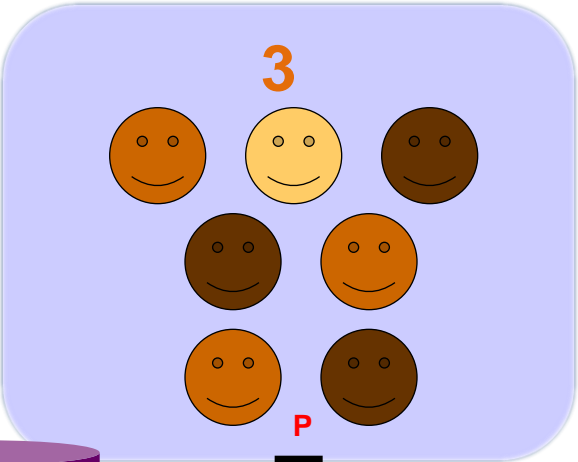
POPULATION



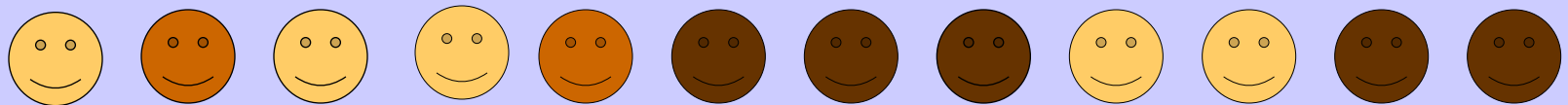
STATE



STATE



RANDOM NUMBER : 1



SAMPLE

# EXERCISE 1.6

In principle, there are four basic probability sampling techniques, constitute the *simple random sampling*, *systematic sampling*, *stratified random sampling* and *cluster sampling techniques*. Based on the given applications below, identify the sampling technique has been applied.

- i. On about day of the last presidential election, ABC News organized an exit poll in which specific polling stations were randomly selected and all voters were surveyed as they left the premises.

Cluster sampling technique

- ii. A team of researchers in local university conducted a study about the achievement of students according their major course. In order to pursue the objective, the researchers randomly selected 820 students majoring in mechanical engineering, 1463 students majoring in computer science, and 760 students majoring in social science.

Stratified random sampling technique

- iii. In order to sustain the quality of a product, a production engineer selects every 20<sup>th</sup> product produced in every production line.

Systematic sampling technique

# EXERCISE 1.5

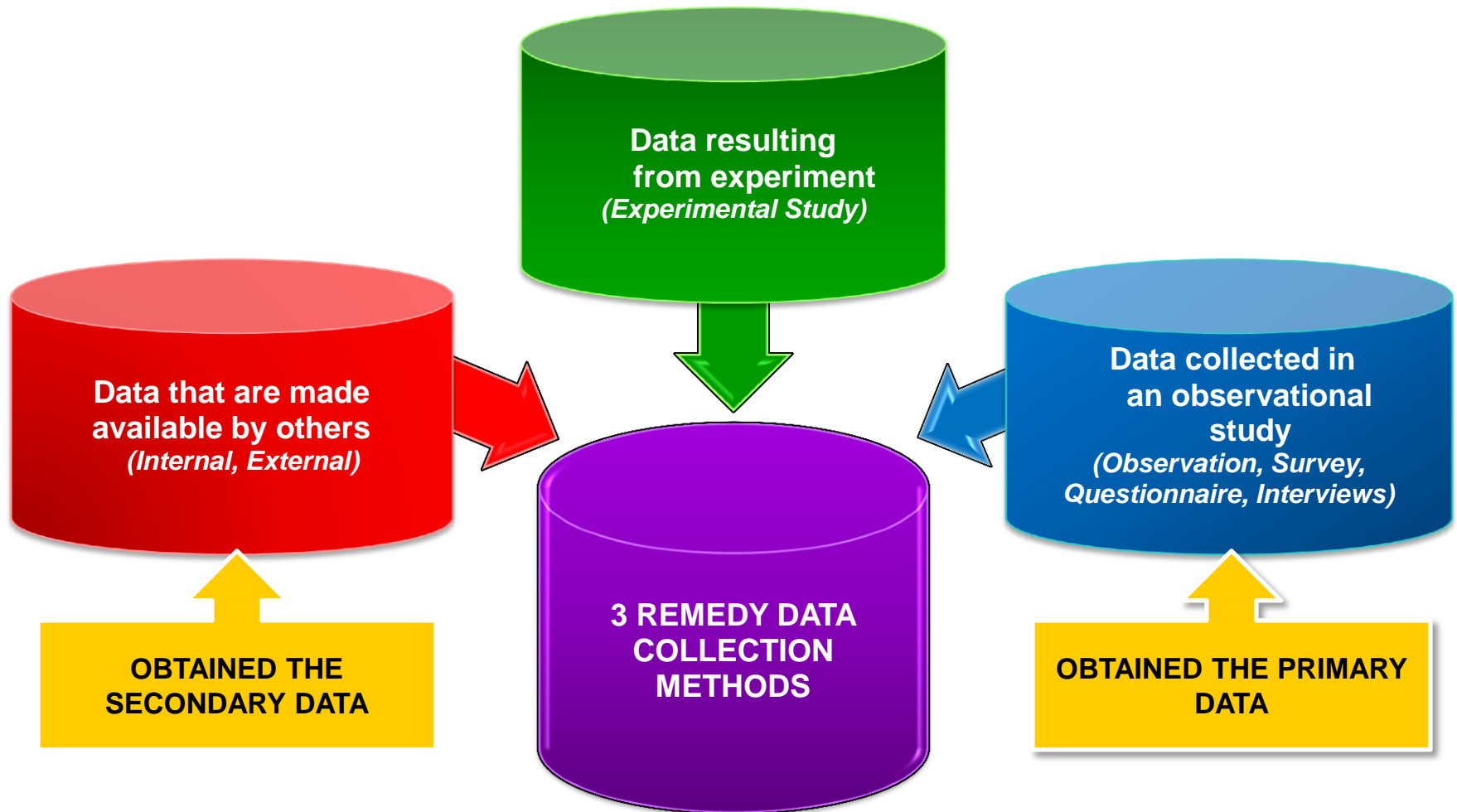
- iv. During the lecture of Statistics & Probability last Friday, the lecturer has been randomly selected eight out of sixty students to show that how they solve a statistical exercises based on the last number of their students ID, which generated randomly using a computer.

Simple random sampling technique

- v. A researcher wants to determine for a gender gap on the perception of 10000 residents in a small community about the development of the economy in the community. Therefore, she surveyed exactly 500 men and women randomly selected from the community.

Stratified sampling technique

# DATA COLLECTION METHODS



# EXERCISE 1.7

State whether each of the following statements is *true* or *false*.

- i. A study of statistics can be divided into two sections, namely descriptive and inferential statistics.

**TRUE.**

- ii. Inferential statistics consist of methods dealing with collection, tabulation, summarization and presentation of data.

**FALSE; The statement described descriptive statistics.**

- iii. The numerical descriptive measures of samples are called statistic.

**TRUE.**

- iv. The method of dividing the population elements into two groups based on their education levels and then selecting a simple random sample from each group is called cluster sampling.

**FALSE; Stratified Sampling.**

- v. The variable gender is an example of a quantitative variable.

**FALSE; Qualitative.**

- vi. Rating of new released smart phone as excellent, good, average or poor is a nominal-level of measurement.

**FALSE; Ordinal-Level.**

# EXERCISE 1.7

The production manager interested to determine the average time taken (in minutes) of an operator to produce 100 computer chip in a production line. Among 98 operators in the production line, he has randomly selected 8 operators based on their staff ID and recoded their time taken to complete the task.

- i. State the population involved in this study.

98 operators in a production line

- ii. State variable of interest and the type of variable involved in this study.

The average time taken of for an operator to produce 100 computer chip; Quantitative

- iii. What sampling method is used by the production manager?

Simple random sampling technique

- iv. What is the method of data collection used in this study?

Observational study.

# THANK YOU

## END OF CHAPTER 1