Chapter 2

Collection and Presentation of Data

Outline of Chapter 2

- **Preliminaries**
- **Methods of Data Collection**
  - **Common Methods of Data Collection**
  - **The Questionnaire**
- Sampling
  - Methods of Probability Sampling
  - Methods of Nonprobability Sampling
- Tabular and Graphical Presentation
- The Frequency Distribution
- The Stem-and-Leaf Display

Reference: Chapters 2-4,12 of Elementary Statistics by ACS

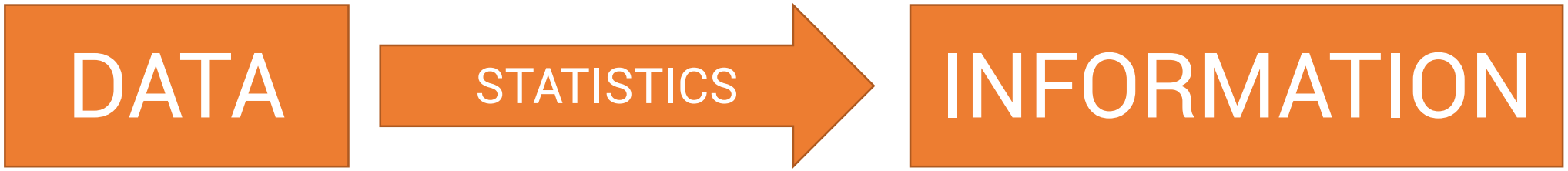
Part 1

Preliminaries and Methods of Data Collection

Preliminaries

Recall:

Information empowers!



Therefore, we need to collect data.

- accurate, complete, and up-to-date
- appropriate in solving/answering problems
- from reliable sources – used sound methodologies
- the information that we will get is only as good as the data on which it was based upon

Preliminaries: Measurement

Definition 2.1

Measurement is the process of determining the value or label of the variable based on what has been observed.

Example

1. In a hospital, a medical staff obtains one's blood pressure reading and blood sugar level.
2. In a preschool, a teacher grades the students with A – Very Good, B - Good, C - Fair, D – Needs Improvement.
3. In an exit poll, an enumerator asks a voter if s/he voted for a certain candidate.

Illustration

Based on our objectives, we can measure the variable of interest in various ways.

Variable of Interest: Educational level

Classification 1

0 – Pre-Primary

1 – Primary

2 – Lower Secondary

3 – Upper Secondary

4 – Post Secondary

5 – 1st Stage Tertiary

6 – 2nd Stage Tertiary

Classification 2

0 – Nursery/ Kinder/ Preparatory

1 – Grade 1-2

2 – Grade 3-4

3 – Grade 5-7

4 – 1st to 2nd Year High School

5 – 3rd to 4th Year High School

6 – 1st to 2nd Year College

7 – 3rd to 4th Year College

8 – 5th or upper College

9 – Master's Degree

10 – PhD

Classification 3

Number of years completed

Levels of Measurement

1. Nominal, 2. Ordinal, 3. Interval, and 4. Ratio

- It is important to know the level of measurement used to measure a variable to help us in the interpretation of the value that the variable takes on.
- Knowing the level of measurement helps us decide on the appropriate statistical tool to analyze the data.
- Our interpretation of the values in our data is dependent on the rule that we used to assign the values to the different categories of the variable.

Levels of Measurement: Ratio

Definition 2.2

The **ratio** level of measurement has ALL of the following properties:

1. The numbers in the measurement system are used to classify an element into distinct, nonoverlapping and exhaustive categories.
2. The system arranges the categories according to magnitude.
3. The system has a fixed unit of measurement representing a set size throughout the scale.
4. The system has an absolute zero.

Example

1. allowance of a student (in pesos)
2. distance traveled by an airplane (in kilometers)
3. speed of a car (in km/h)
4. height of an adult (in cm)
5. weight of a newborn baby (in kilograms)
6. number of children
7. exam score

Some Notes on the Properties

1. The numbers in the measurement system are used to classify an element into distinct, nonoverlapping and exhaustive categories.

The first condition requires that we use categories that would place the observations logically into one and only one category.

This means that two objects assigned the same value must belong in the same category and be placed in a different category if the characteristic of interest in really different.

Categories should be DISTINCT, NONOVERLAPPING, and EXHAUSTIVE.

Some Notes on the Properties

2. The system arranges the categories according to magnitude.

The second property requires that the measurement system must arrange the categories in descending or ascending order.

Some Notes on the Properties

3. The system has a fixed unit of measurement representing a set size throughout the scale.

The third property requires the scale to use a unit of measure that depicts a fixed and determinate quantity.

This means that a one-unit difference must have the same interpretation / meaning wherever it appears in the scale.

Some Notes on the Properties

4. The system has an absolute zero.

The fourth property requires the scale to have an absolute zero or the true zero point.

This means that the scale considers the value “0" as the complete absence of the characteristic itself.

Exercise

Variable of interest: Allowance of a student (in pesos)

Property 1: ?

Property 2: ?

Property 3: ?

Property 4: ?

Exercise

Which of the properties stated in the definition of the ratio level are not satisfied by the following scale?

Variable: Faculty rank

Measuring system:

1 – Instructor
2 – Assistant professor
3 – Associate professor
4 - Professor

Example

Variable of Interest: Intelligence

Scale: Stanford IQ scores

- Joe's IQ score equals 100 and John's IQ score equals 150.
- John has a higher IQ than Joe; that is, IQ scores can be arranged in order.
- John's IQ score is 50 points higher than Joe's IQ score; that is, differences can be calculated and interpreted.
- However, we cannot conclude that John is 1.5 times (150/100 = 1.5) more intelligent than Joe. An IQ score of zero does not indicate a complete lack of intelligence.

Levels of Measurement: Interval

Definition 2.3

The level of measurement is said to be **interval** if and only if it satisfies the following properties:

1. The numbers in the measurement system are used to classify an element into distinct, nonoverlapping and exhaustive categories
2. The system arranges the categories according to magnitude.
3. The system has a fixed unit of measurement representing a set size throughout the scale.
4. The system has NO true zero. (A zero in the scale does not mean the absence of the characteristic.)

Example

- Temperature readings in degrees centigrade/ Fahrenheit
- Intelligence quotient
- Calendar dates whether Gregorian, Hebrew, or Islamic

Note: Since there is no true zero in the system then the ratios of measurements taken using the scale are not interpretable.

Question: What is the level of measurement of thermodynamic temperature reading in Kelvin (which is proportional to the molecular activity or thermal energy of matter)?

Levels of Measurement: Ordinal

Definition 2.4

The level of measurement is said to be **ordinal** if and only if it satisfies the following properties:

1. The numbers in the measurement system are used to classify an element into distinct, nonoverlapping and exhaustive categories.
2. The system arranges the categories according to magnitude.
3. The system has NO fixed unit of measurement representing a set size throughout the scale.
4. The system has NO true zero.

Example

1. Product rating (poor, good, excellent)
2. Socioeconomic class (lower, middle, upper)
3. Pain level (none, low, moderate, severe)
4. Customer satisfaction (Not satisfied, Fairly satisfied, Satisfied, Very satisfied)
5. Automobile size description (Subcompact, compact, intermediate, full-size)

Note: differences and ratios of measurements taken using the ordinal scale are not interpretable

Exercise

Question: Which of the following systems' level of measurement is ordinal?

Variable: size of shirt

System 1	System 2
1 – Small	1 – Extra Large
2 – Extra Large	2 – Large
3 – Medium	3 – Medium
4 – Large	4 – Small

Levels of Measurement: Nominal

Definition 2.5

The level of measurement is said to be **nominal** if and only if it satisfies the following properties:

1. The numbers in the measurement system are used to classify an element into distinct, nonoverlapping and exhaustive categories
2. The categories in the measurement system are NOT arranged according to magnitude.
3. The system has NO fixed unit of measurement representing a set size throughout the scale.
4. The system has NO true zero.

Example

Religious affiliation:

1-Catholic 2-Protestant 3-Muslim 4-Iglesia ni Kristo 5-Others

Type of movie:

1-romance 2-adventure 3-horror 4-others

Note: The nominal level is the weakest level of measurement since we use symbols or numbers for the sole purpose of classifying an individual/object into two or more categories. We can only count the number of observations per category and compute for proportions and percentages.

Categorical Data

Observations measured using the nominal or the ordinal levels of measurement are usually referred to as **categorical data**.

Special type of analysis is used for categorical data because of the restrictions in the system.

In the BS Statistics curriculum, we study these types of analyses and techniques in *Stat 132: Nonparametric Inferential Statistics* and *Stat 149: Introduction to Categorical Data Analysis*.

Categorical Data: Example

- Smoking Habits (Smoking, Nonsmoking)
- Drinking Habits (Drinking, Nondrinking)
- Sex (Male, Female)
- Opinion on a statement (Agree, Neutral, Disagree)
- STS Bracket (A, B, C, D, E1, E2)
- Scholarship Status (Good Standing, Suspended, Terminated)
- Grade in Stat 101 (if the scale has no fixed unit)

Exercise

State the level of measurement used to measure the following variables:

a) Postal zip code
b) Performance rating of an employee as excellent, very good, good, fair, and bad
c) Student number
d) Ranking of a student in class
e) Body temperature of a child measured in Celsius
f) Annual salary of employee (in pesos)
g) Years in which LAKAS won presidential election
h) Letter grade (A, B, C, and D) of a pre-school student

Exercise

State the level of measurement used to measure the following variables:

i) Tax identification numbers of an employee

j) Classification of heavy equipment as bulldozer, excavator, loader, and roller

Exercise: UP Grading System

State the level of measurement used.

Teacher 1		Teacher 2	
96 – 100	1	99 – 100	1
91 – 95	1.25	97 – 98	1.25
86 – 90	1.5	95 – 96	1.5
81 – 85	1.75	93 – 94	1.75
76 – 80	2	91 – 92	2
71 – 75	2.25	86 – 90	2.25
66 – 70	2.5	81 – 85	2.5
61 – 65	2.75	71 – 80	2.75
56 – 60	3	61 – 70	3

Levels of Measurement: Summary

Nominal – “nom” – name

Ordinal – order

Interval – interval/“width” – difference is meaningful

Ratio – true 0 – getting a “ratio” of quantities is meaningful

N-O-I-R / R-I-O-N

We DO NOT use the levels of measurement to classify the variable. We use it to classify the system or scale used to measure the variable.

Always note how the scale/rule/system was used.

Methods of Data Collection

- Use of Documented Data
- Survey method
- Experiment
- Observation method
- Other methods
  - ❑Use of internal data
  - ❑Registration
  - ❑Computer simulation

Use of Documented Data

- The researcher can obtain documented data from previous studies of individuals or private, government, non-government agencies.

- The researcher may find documented data in published or written reports, unpublished documents, periodicals, and others.

- In the Philippines, the Philippine Statistical System (PSS) is the governmentwide system of collecting and disseminating national statistics.

Use of Documented Data

We can classify documented data by source:

Definition 2.6.

1. Primary Data

Primary data are data documented by the primary source. The data collectors themselves documented this data.

2. Secondary Data

Secondary data are data documented by a secondary source. An individual/agency, other than the data collectors, documented this data.

Philippine Statistics Authority

- Created by the virtue of RA 10625.
- The PSA is the central statistical authority of the Philippine government on primary data collection.

- **Reading Assignment:**

Visit psa.gov.ph and read about the PSA and the Philippine Statistical System.

Examples: Primary Data

1. **Bangko Sentral ng Pilipinas (BSP)** is a primary source of data on banking and finance.
2. **Philippine Statistics Authority (PSA)** is a primary source of data on population, housing, and establishments.
3. **Pulse Asia** is a primary source of data on opinions or sentiments of the people on current issues.

Examples: Secondary Data

a) The United Nations' compiled data for its yearbook, which were originally gathered by government statistical agencies of different countries;

b) A medical researcher's documented data for his research paper, which were originally collected by the Department of Health;

c) The documented data of the research team of a congressman for its report, which were originally collected by the Department of Education and Commission on Higher Education; and

d) The documented data of a student for his thesis, which were originally collected by the Department of Labor and Employment.

Comparison

Advantages of primary over secondary data include:

1. The primary source frequently provides vital information that are crucial in assessing the applicability and accuracy of the collected data. This includes the definitions of terms and statistical units used in the survey, methodology including a copy of the questionnaire and a discussion of the sampling design or experimental design.
2. The primary data are usually more comprehensive. Secondary data have already been filtered by secondary source to address their purpose.
3. The secondary data may contain mistakes due to errors in transcription made when the figures were copied from the primary source.

Survey Method

Definition 2.7

The **survey method** is a method of collecting data by asking people questions.

When the data came from asking all the people in the population, then the study is called a **census**. On the other hand, when the data came from asking a sample of people selected from a well-defined population, then the study is called a **sample survey**.

Survey Method

The people who answer the questions in a survey are called **respondents**.

The questionnaire contains all the questions asked in a survey.

Notes:

- The success of the survey as a data collection method relies on the honesty of the respondents and their capability to give truthful answers.
- It is crucial for the success of a survey to have a well-defined questionnaire.

Survey Method

3 Classifications of Respondents

1. **Good Respondent** – gives answers that will make you happy
2. **Bad Respondent** – gives answers that will make you annoyed
3. **Honest Respondent** – gives honest answers to your questions

Example

Some actual surveys in the Philippines are as follows:

1. Pulse Asia conducted a sample survey on voter response to political ads in the May 2004 election. Its respondents were selected registered voters who intend to vote in the 2004 election.

2. The Department of Energy regularly conducts the Household Energy Consumption Survey (HECS) to measure the level and pattern of energy consumption at the national and regional levels.

Communicating to Respondents

- **Personal interview** : interviewers personally ask the respondents and record their answers on the questionnaire
- **Telephone interview** : interviewers ask the respondents through the telephone
- **Using self-administered questionnaires** : respondents fill up the questionnaires themselves without any assistance from an interviewer
- **On-line surveys** : respondent reads the questions and sends his responses via the internet or e-mail
- **Focus group discussions (FGD)** : a moderator follows a focus group discussion guide to direct a freewheeling discussion among a small group of people

Considerations

In choosing the method of communicating the questions in a survey, the following considerations may be useful:

- Ability to secure the type of data needed
- Cost
- Speed
- Accuracy of data obtained/Quality of response

- Response rate
- Geographic flexibility
- Availability of good interviewers and field supervisors
- Population coverage

Experiment

The **experiment** is a method of collecting data where there is direct human intervention on the conditions that may affect the values of the variable of interest.

- The **explanatory variable/s** are the variables in the study whose values are believed to have an effect on the value of the response variable/s.
- The **treatments** or **factor levels** are the values or categories of the explanatory variable that are being considered in the study.
- The **extraneous variable/s** are those variables that may have an effect on the response variable but their effects are not of interest in the study.

Example: Experiment

What was the first experiment you did when you were still a child?

MONGO EXPERIMENT!

Objective:

Response Variable:

Explanatory Variable:

Treatments:

Extraneous Variables:

Distinct Feature of Experiments

The researcher intervenes by controlling the conditions that may affect the response variable by:

i. using a randomization mechanism in assigning the treatments

ii. controlling the identified extraneous variables.

- researcher can isolate the effects of the EV on the RV
- researcher can clarify the direction and strength of their relationship

Through **(i)**, the effects of all other extraneous factors that the experimenter failed to control are expected to cancel each other out.

This is why, unlike surveys, the experiment is a more effective method of data collection in establishing cause-and-effect.

Advantages and Disadvantages

Advantage:

- most effective method in establishing cause-and-effect

Disadvantages:

- not always feasible to randomize the assignment of treatments
- difficult to assess the reliability of inferences about a well-defined population if experimental units were not selected through a randomization mechanism
- results may be different when applied to the natural setting (since we perform experiments in a controlled environment)

Observation Method

Definition 2.9

The **observation method** is a method of collecting data on the phenomenon of interest by recording the observations made about the phenomenon as it actually happens.

Examples

1. Behavior patterns in panic situations a big fire, Mt. Pinatubo eruption, and the collapse of the World Trade Center in New York
2. Behavior of animals in their natural habitat
3. Behavior of newborn babies in the nursery
4. Growth and other characteristics of plants in farms that grow organic produce and those that do not (Note: There should have been no human interference on assignment of treatments.)
5. The adjustment of college freshmen to campus life by observing the behavior of freshmen in various settings.

Some Notes

- The observation method requires the observation of the element in its natural setting, recording as observations what the researchers see and hear. Sometimes the researchers would have to wait for a long time for the phenomenon of interest to occur.
- It is practical to use when elements cannot verbalize their answers because they cannot speak (e.g., studies on animal behavior since we cannot administer a questionnaire to animals).

Some Notes

- The data collected are usually the subjective perceptions and interpretations of the researcher on the event under study. Usually, the researchers give a qualitative narrative of what they have observed about the phenomenon of interest. This type of data is difficult to analyze using formal statistical techniques.
- However, it is still possible to collect objective quantitative data using the observation method.

Specially-Designed Measuring Instruments

- For **traffic studies**: distance-measuring devices, traffic-counting devices, speed-measuring devices

- For **meteorological studies**: special devices to measure temperature, humidity, and precipitation

- For **media research**: attach special devices on tv sets to monitor tv viewing habits of individual household members

Reading Assignment

- Methods of Collecting Quantitative Data using the Observation Method
  - Page 40 in Elementary Statistics, by Almeda, Capistrano and Sarte
  - Topics
    - Duration Recording
    - Frequency-count Recording
    - Interval Recording
    - Latency Recording
    - Time Sampling
    - Participant vs Non-Participant Observation

Comparison of Methods

Aspect	Survey	Experiment	Observation
Assessing the reliability of generalizations about a well-defined population	Generally possible	Sometimes difficult	Oftentimes difficult
Ability to establish cause-and-effect	Poor	Superior	Poor
Realism of data	Realistic	Least Realistic	Most Realistic

Exercise

Classify the method of data collection (survey, experiment, or observation)

a) A local TV network asked voters to indicate whom they voted as they exited the polling booth.

b) A private hospital divides terminally ill patients into two groups, with one group receiving medication A and the other group receiving medication B. After a month, they measured each subject's improvement.

c) A researcher investigates the level of pollution in key points in Metro Manila by setting up pollution measuring devices at selected intersections.

d) The school administration asked students whether they are willing to have an increase in laboratory fees if there is an upgrade of computers.

Exercise

What method of data collection is most appropriate for the following cases? Give a brief explanation for your choice.

a) Studying two groups of patients and determining if exercise lowers the blood pressure.

b) The Department of Health monitors and evaluates the benefits of the family planning methods given to a certain community.

Other Methods of Data Collection

- Use of **internal data**. These are data generated from the operation and administration of the researcher's company. These data are possible byproducts of the administrative and management functions of the company.
  - Example: personnel records, financial statements of the company, inventory reports, production and sales reports, purchasing reports, and payroll
- Use of **registration data**. These are data generated by other agencies/ organizations through the process of registration, as required by some law, regulation or usual custom. Registration data will not always be complete such as the data on eligible voters in the Philippines from COMELEC.

Other Methods of Data Collection

- Use of **Registration Data**
  - Example: data on the birth, death, and marriage of all individuals in the country from civil registry documents of PSA, data on motor vehicles registered at the Land Transportation Office, data on students registered at the University Registrar

- **Computer simulation**. This makes use of a special kind of a mathematical formula called a statistical model that computes for values of the variable of interest by incorporating the use of a randomization mechanism.
  - In translating the reality into mathematical equations, we use statistical models.

Reading Assignment

Please read

- Section 2.3 The Questionnaire. Be able to identify a FAULTY questionnaire
- Chapter 4 Presentation of Data. Be able to identify a FAULTY graph or table.