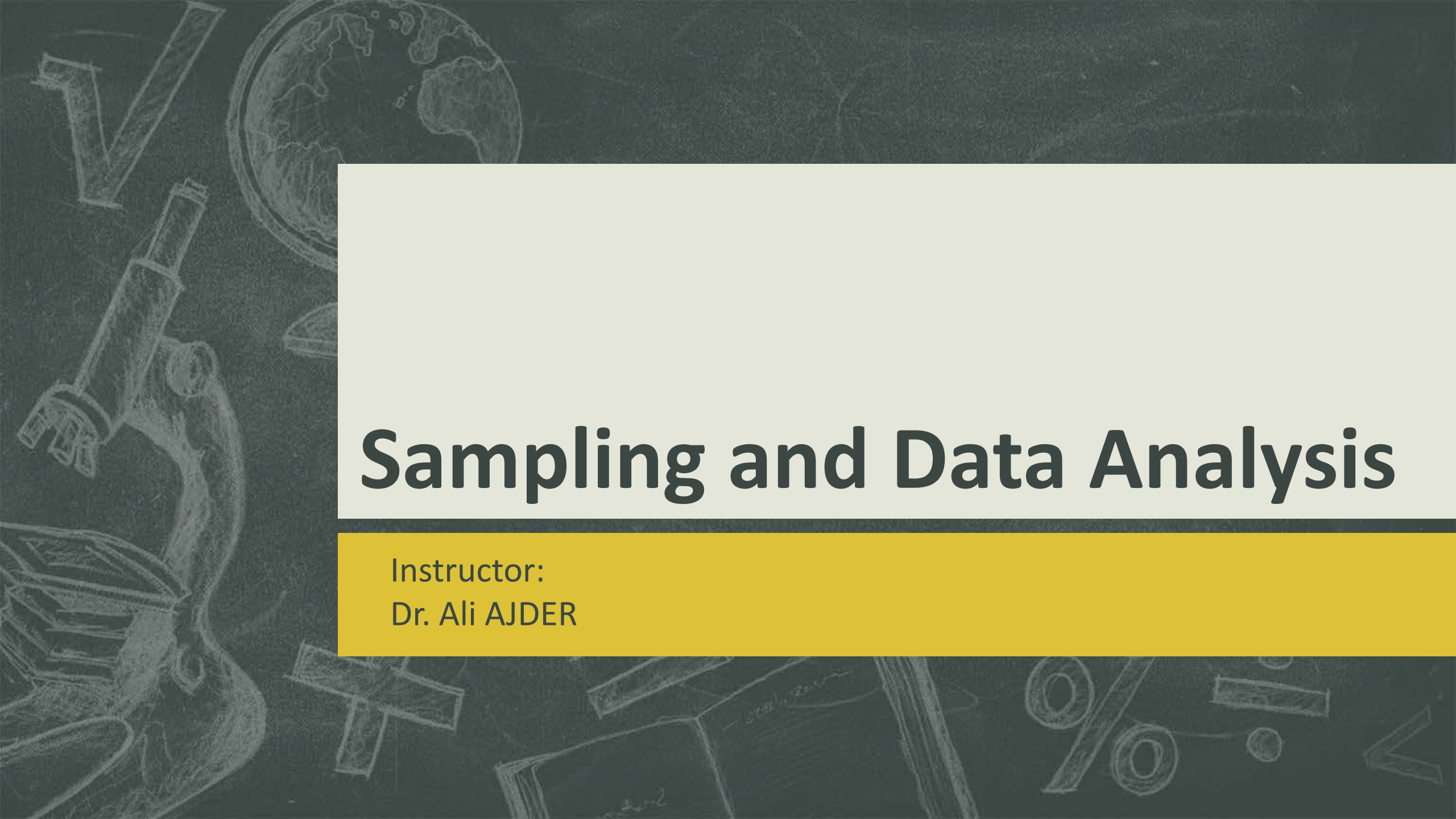


Sampling and Data Analysis

Instructor:
Dr. Ali AJDER

Week-2

- Introduction to Statistic
- Variables
- Levels of Variables/ Scales
- Grouped/ Categorical Data

Introduction to Statistics

Data/ Information Analysis

"Workers must be equipped not simply with technical know-how, but also with the ability to create, analyze, and transform information and to interact effectively with others. That is, separate the facts from opinions, and then organize these facts in an appropriate manner and analyze the information." **Dr. Alan Greenspan**

Examples

Researcher,

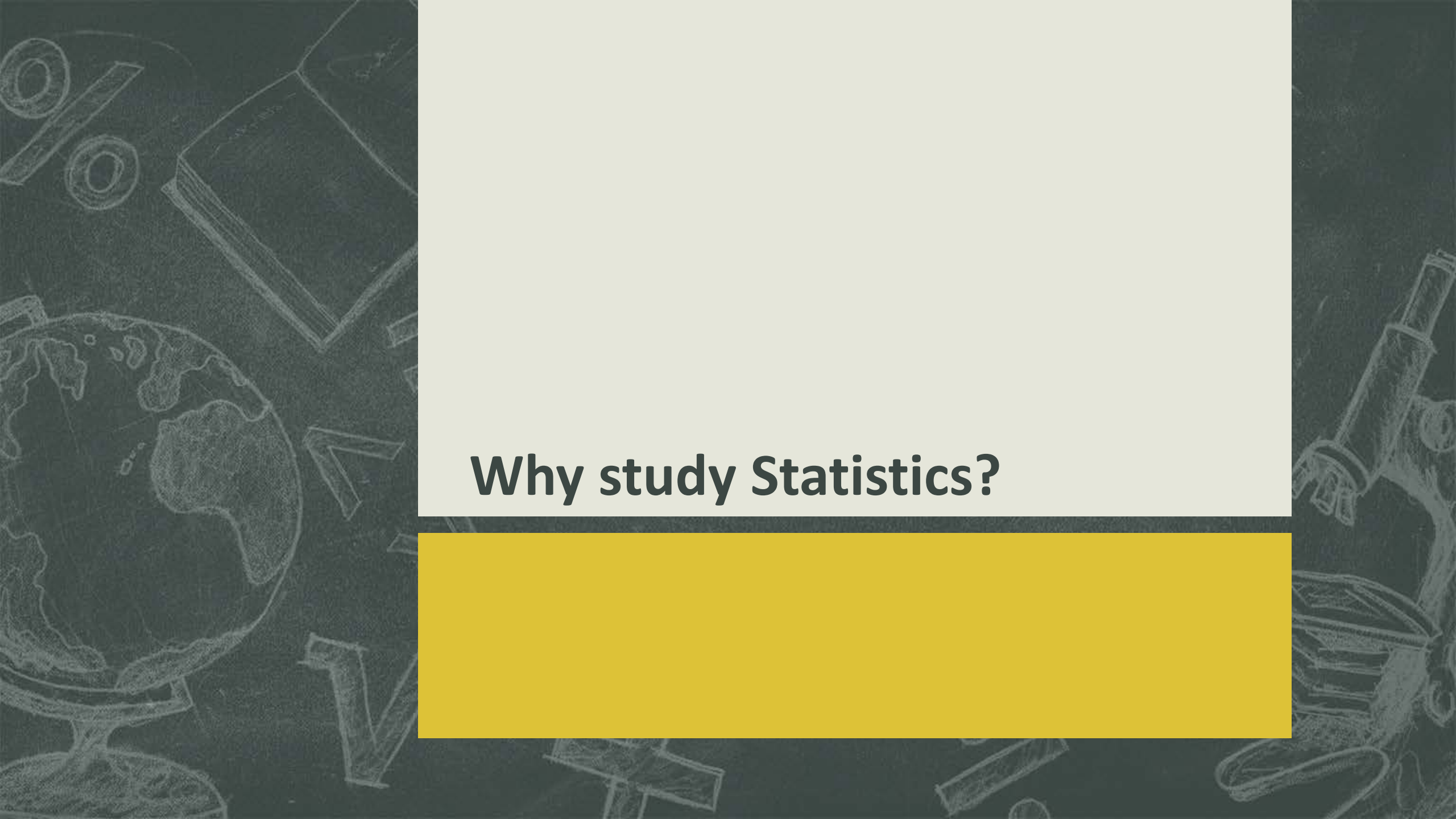
Whether a medical intervention helps in reducing the burden of a disease,
How personality relates to decision-making,
Whether a new fertilizer increases the yield of crops,
How a political system affects trade policy,
Who is going to vote for a political party in the next election,
What are the long-term changes in the population of a fish species…

Governments and organizations,

Life expectancy of a population,
The risk factors for infant mortality,
Geographical differences in energy usage, migration patterns, or reasons for unemployment….

In business,

Identifying people who may be interested in a certain product,
Optimizing prices, and
Evaluating the satisfaction of customers are possible areas of interest…

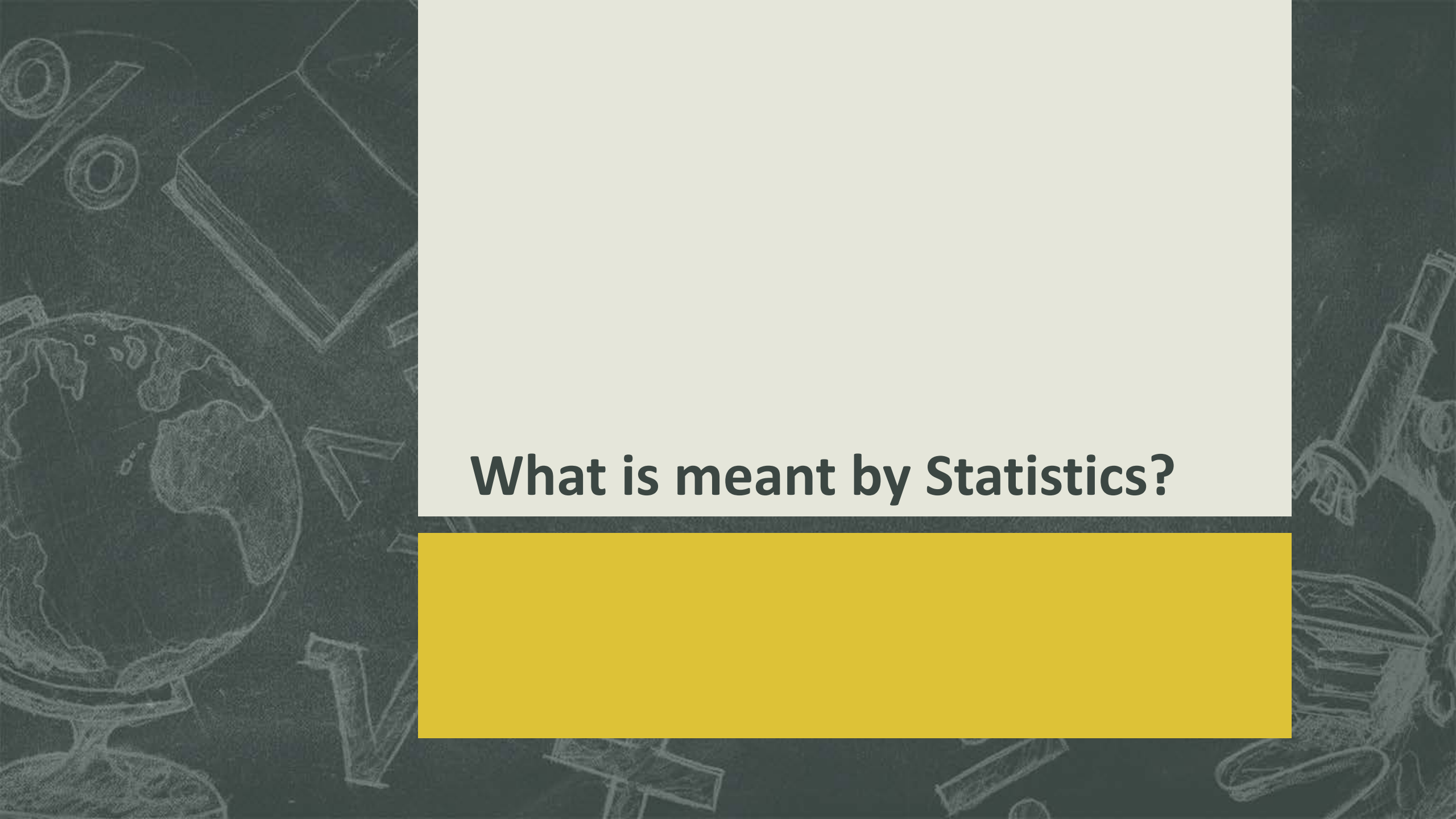


Why study Statistics?

There are at least three reasons for studying statistics…

1) Data are everywhere,
2) Statistical techniques are used to make many decisions that affect our lives, and
3) No matter what your career, you will make professional decisions that involve data.

An understanding of statistical methods will help you make these decisions more effectively.

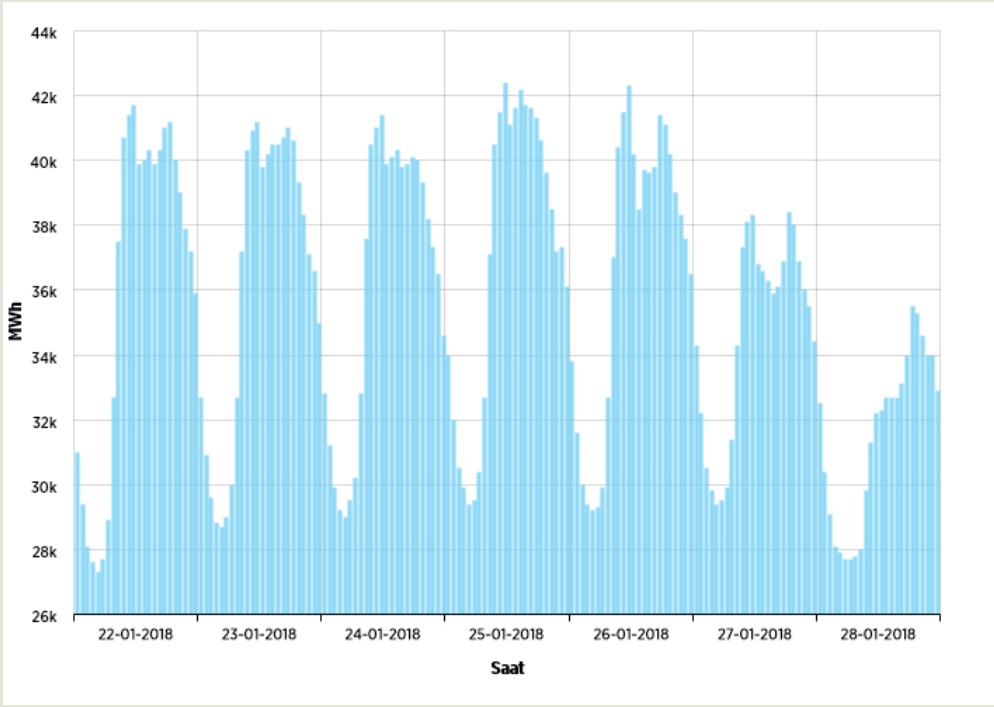


What is meant by Statistics?

It really has two meanings.

1. In the more common usage, statistics refers to **numerical information.**
2. In this lecture, it has a much broader meaning than just collecting and publishing numerical information;

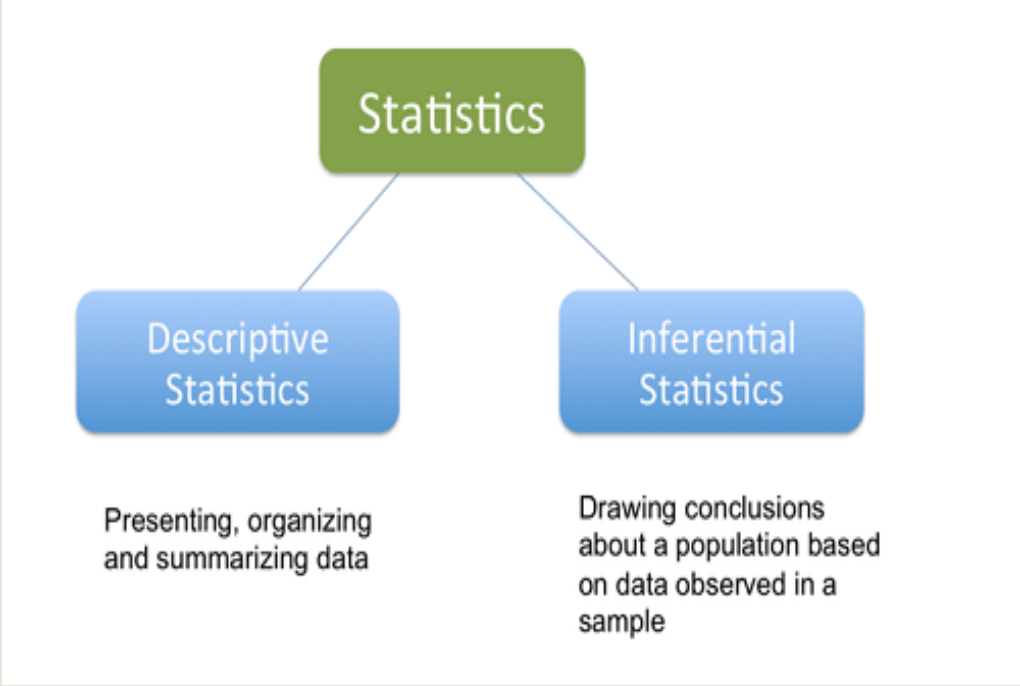
STATISTICS The science of collecting, organizing, presenting, analyzing, and interpreting data to assist in making more effective decisions.



Types of Statistics

DESCRIPTIVE STATISTICS Methods of organizing, summarizing, and presenting data in an informative way.

INFERENTIAL STATISTICS The methods used to determine something about population on the basis of a sample.



Basic Definitions

- **Observation-** ω

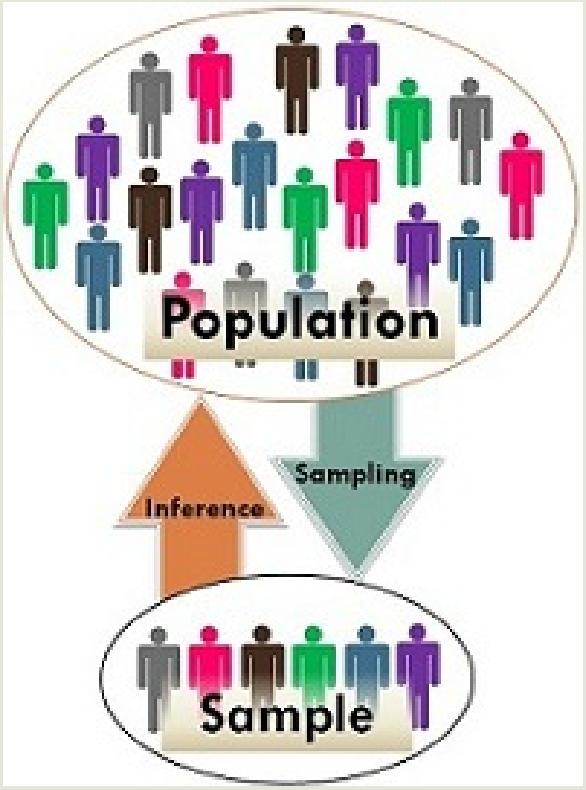
Data/ information

- **Sample-** $\omega_1, \omega_2, \dots \omega_n$

Why take a *sample* instead of studying every member of the *population*?

- **Population-** $\omega_1, \omega_2, \dots \omega_n \subseteq \Omega$

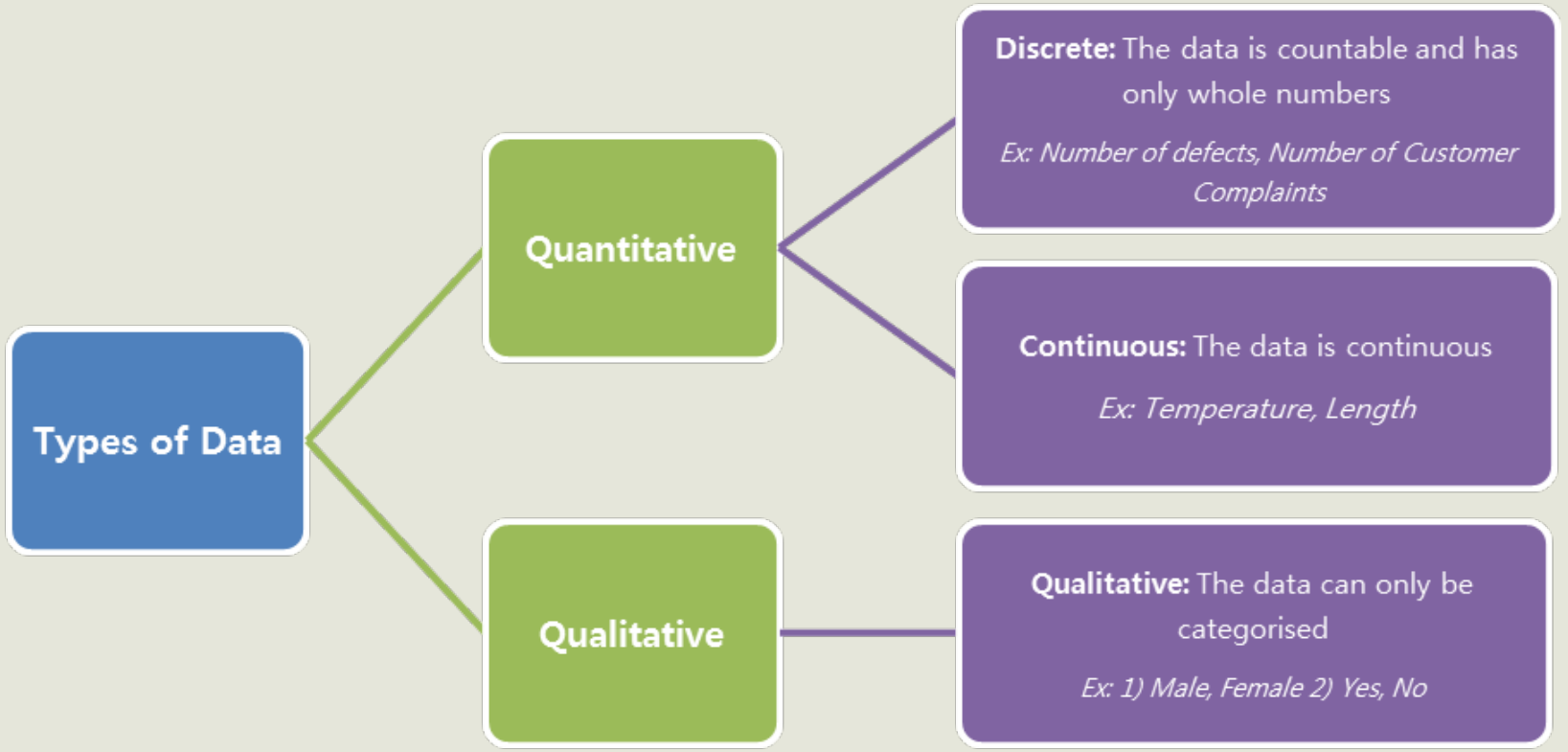
Population of Turkey?



Data Collection

- Survey
- Experiment
- Observational Data
- Primary and Secondary Data

Types of Variables



Levels of Variables

1. Nominal

- can only be classified
- can **not** be ordered
- (Colours of M&M chocolate)

2. Ordinal

- can be ordered
- (Rate of instructor in this class/ Good- Average- Poor)

3. Interval

- all the characteristics of the ordinal level
- between values is a constant
- (temperature)

4. Ratio

- all the characteristics of the interval level
- 0 point and ratio of 2 numbers is meaningful
- (price)

What is the level of measurement reflected by the following data?

- The age of each person in a sample of 50 adults who listen to one of the 1,230 talk radio stations in the United States is:

35	29	41	34	44	46	42	42	37	47
30	36	41	39	44	39	43	43	44	40
47	37	41	27	33	33	39	38	43	22
44	39	35	35	41	42	37	42	38	43
35	37	38	43	40	48	42	31	51	34

- In a survey of 100 business, 50 were from Istanbul, 25 from Ankara, 15 from Izmir, and 10 from Eskisehir.

Summary-1

Statistics is the science of collecting, organizing, presenting, analyzing, and interpreting data to assist in making more effective decisions.

There are two types of statistics.
A. Descriptive statistics are procedures used to organize and summarize data.
B. Inferential statistics involve taking a sample from a population and making estimates about a population based on the sample results.
1. A population is an entire set of individuals or objects of interest or the measurements obtained from all individuals or objects of interest.
2. A sample is a part of the population.

There are two types of variables.
A. A qualitative variable is nonnumeric.
1. Usually we are interested in the number or percent of the observations in each category.
2. Qualitative data are usually summarized in graphs and bar charts.
B. There are two types of quantitative variables and they are usually reported numerically.
1. Discrete variables can assume only certain values, and there are usually gaps between values.
2. A continuous variable can assume any value within a specified range.

Summary-2

There are four levels of measurement.

A. With the nominal level, the data are sorted into categories with no particular order to the categories.

1. The categories are mutually exclusive. An individual or object appears in only one category.

2. The categories are exhaustive. An individual or object appears in at least one of the categories.

B. The ordinal level of measurement presumes that one classification is ranked higher than another.

C. The interval level of measurement has the ranking characteristic of the ordinal level of measurement plus the characteristic that the distance between values is a constant size.

D. The ratio level of measurement has all the characteristics of the interval level, plus there is a zero point and the ratio of two values is meaningful.

Questions?