

# Variables in Research

Prof. Khaymar Mya

MBBS, MMedSc(PH), PhD(PH), DipMedEd

Former Rector

University of Public Health, Ygn

Acknowledgement; Dr. Win Maw Tun, DMR

# What is variable

- **Variables** are **properties or characteristics** of people or things that **vary**  
(in quality or magnitude from person to person or object to object)
- A variable is a characteristic of a **person, object, or phenomenon** (that can take on different values) that is studied/ measured in research to achieve “objectives”
- A Variable may be situation specific
  - Gender is variable but not for the class of all girls

→ These may be in the form of

- numbers (age, height, weight) or
- groups (sex, occupation)

→ What **information** are we going to collect in our study to meet the objectives

→ What we need **to measure** in the study

→ In observational research, **to describe** the problem more precisely

- Eg. To know the prevalence of **malnutrition** in a group
- Eg. To determine **Calcium level** of post menopausal women
- Eg. To know the prevalence of **dental caries** among primary school children

- Whether there is a relationship between two variables → to describe the **relationship**
- Whether certain factors influence the problem to what **extent**
  - Eg. Relationship between type of food and drink consumption and dental caries
  - Eg. consumption of sugary (snack food, and beverages) with high blood sugar
- In Experimental Research, **manipulate** some variables and then measure the **effects** of this manipulation on other variables


# Types of Variables

- From the viewpoint of **the unit of measurement**

-  **Numerical Variables** (can be measured)

- ❖ Discrete variables

- ❖ Continuous variables

-  **Categorical Variables** (cannot be measured but count)

- ❖ Nominal variables

- ❖ Ordinal variables

- From the viewpoint of **causal relationship**

-  **Independent variables**

-  **Dependent variables**

# Numerical Variables

❖ which can be measured and always numeric

## 1. Discrete variables

 Can be counted

 numbers can only have full values

e.g. number of visits to a clinic (0, 1, 2, 3, 4)

number of sexual partners (0, 1, 2)

## 2. Continuous variables

 Measurements made by the measuring instrument

e.g. height in centimetres (2.5 cm or 2.546 cm)

temperature in degrees Celsius (37.2°C or 37.19°C)

# Categorical Variables

## 1. Nominal variables

The groups in these variables do not have an order or ranking – just name

e.g. **Blood Group**: A, B, O, AB

**Sex**: male, female

## 2. Ordinal variables

These are grouped variables - ordered or ranked in increasing or decreasing order

e.g. **High income, Middle income and Low income**

Categorical variable with only 2 levels – **Dichotomous** (or) **binary variable**

e.g. **age**: under 65 group /65 and over group

**gender**: male/ female)

➤ Notice!!

❑ Numerical variables can be made as -  
categorical variable

e.g. Age:

25.3, 26.0, 28.4, 33.5, 40.0, 45.7, 44.5, 37.8,  
43.7, 29.4, 25.0, 45.0, 39.2, 36.5, 27.6

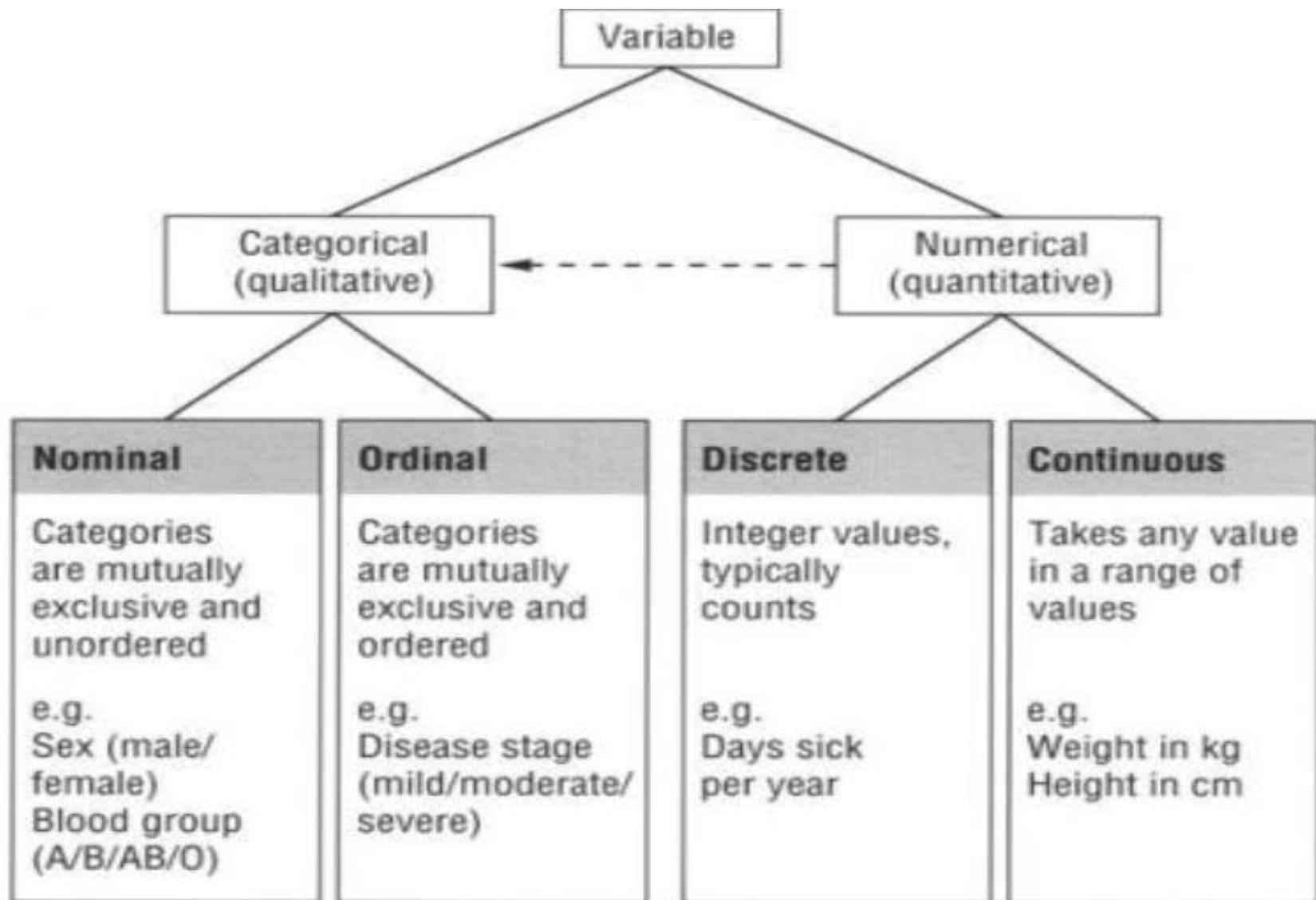
Age group:

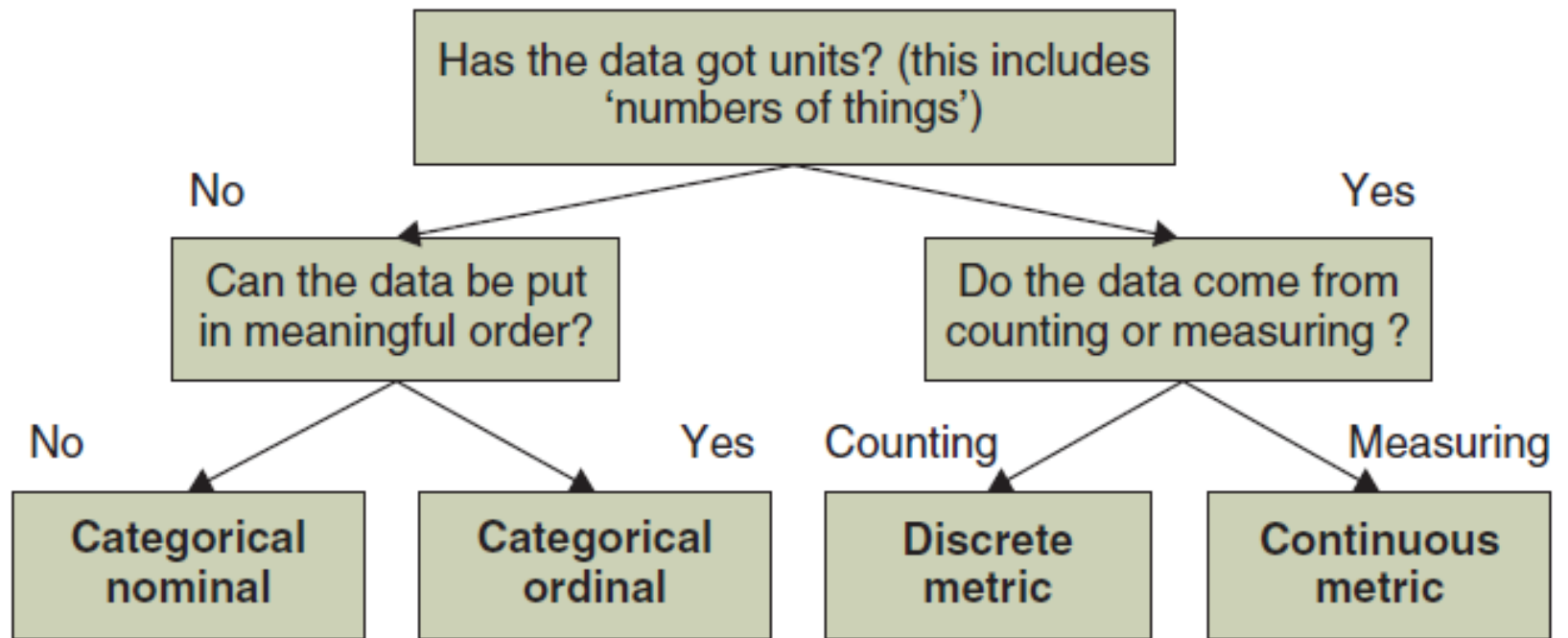
30.0-34.9

35.0-39.9

40.0-44.9







Depending on **relationship** between variables,



variables can be identified into:

 Independent variables




 Dependent variables

 Confounding variables

## Independent variables

-  The independent variable is the **predictor** variable that is hypothesized to **influence** our dependent variable.
-  variable that can be **manipulated** or controlled by the researcher; usually affecting the dependent variable

## Dependent variables

-  the **response** variable
-  What is measured as an **outcome** in a study
-  the variable a researcher is interested in

Eg. The higher the temperature of water, the faster an egg will cook

- Independent variable → temperature of water
- Dependent variable → time to cook an egg

### Independent

- Cause
- Action
- Intervention

### Dependent

- – Effect, outcome
- – Reaction
- – Response

# Dependent/Independent?

- **Determined by the statement of the problem and objective of the study**

Eg. Effect of income on nutritional status

Nutritional status -- Dependent

Income -- Independent

Eg. Effect of educational attainment on income

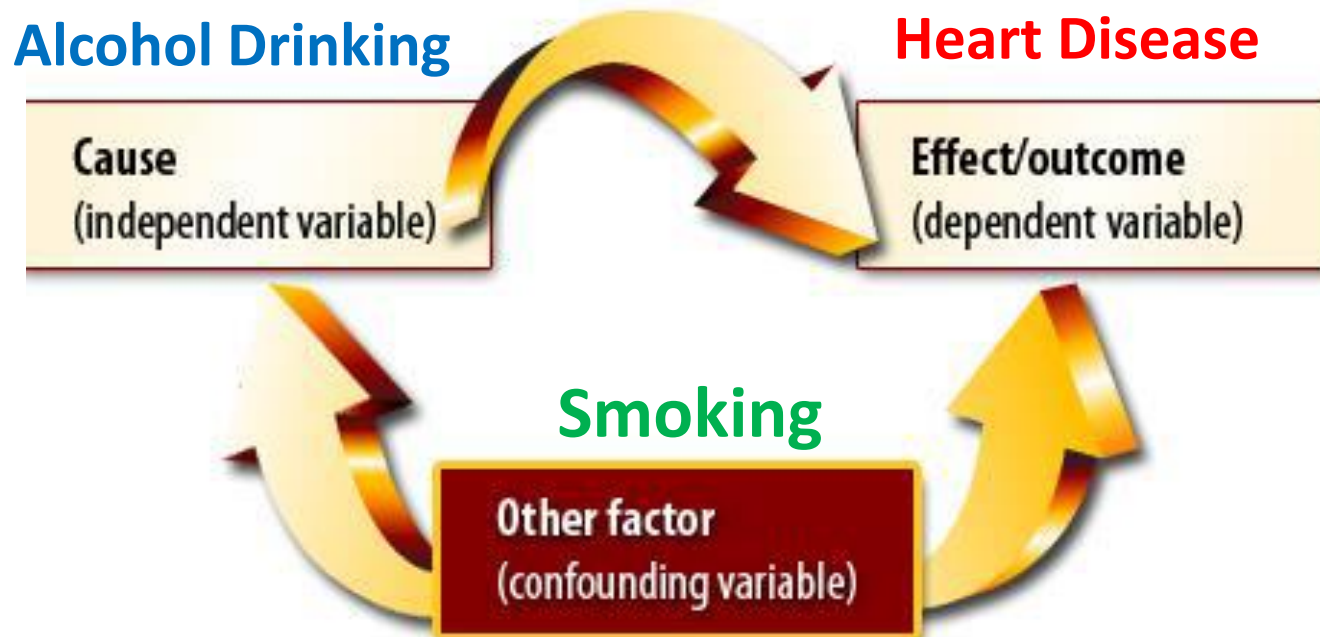
Income -- Dependent

Educational attainment -- Independent

- A **hypothesis** states a presumed **relationship** between two **variables**
- It may take the form of a cause-effect statement, or an "if x,...then y" statement
  - **Eg.** Age at marriage and level of education  
(the higher the education-the higher the age at marriage)
  - Education Level      &      Age at marriage

# Confounding variables

- ✎ Also known as intervening variables
- ✎ Variable which have associated both with disease and exposure being studied





## **BACKGROUND VARIABLES**

**Such as age, sex, educational level, marital status, religion and socio-economic status etc.**

- Often relate and to independent variable
- Influence the problem indirectly
- Should be measured if they are important/take comparable samples

# Research Question and Variables

**Research question:** What are the factors that contribute to prevalence of anaemia in primary school children?

- **Independent Variable:** age, sex, ethnic group, urban/rural, diet, snack consumption pattern
- **Dependent Variable:** haemoglobin level

# OPERATIONALIZING THE VARIABLES AND CHOOSING INDICATORS

- ✏ Some variables can easily be determined or measured
- ✏ Some are difficult to determine
- ✏ **operationalize** with **Precise Indicators** →→  
make **measurable** and avoid **ambiguity**
  - Once appropriate indicators have been identified – can know exactly **what information to look** for
  - This makes the **data collection** as well as the **analysis** more focused and efficient

# How Is It Measured?

- **Conceptual definition:** the abstract meaning of a variable that is usually based on theory; dictionary meaning
  - A conceptual definition tells you what the concept means
  - Eg. **Adolescence** is a transitional stage of physical and psychological development that generally occurs during the period from puberty to legal adulthood
  - Adolescence is usually associated with the teenage years, but its physical, psychological or cultural expressions may begin earlier and end later
- **Operational definition:** An operational definition is how (the researcher) decide to measure the variables in the study - defining the variable to make it measurable (working definition)
  - operational definition tells you how to measure it
  - made before the collection of data begins
  - Eg. **Ages 13 - 19 years**. This may vary in research ...

# Operationalizing variables means that → make them measurable





## Conceptual Definition

- Age
- Haemoglobin level
- Immunization coverage
- Exclusive breast feeding

## Operational Definition

- Completed age
- Hb conc. in capillary blood  
measured by hemoglobinometer
- % of children immunized in a  
particular (age) group
- No other food or drink, not even  
water, except breast milk  
(including milk expressed or  
from a wet nurse) for 6 months  
of life, but allows the infant to  
receive ORS, drops and syrups  
(vitamins, minerals, medicines)

E.g. **Nutritional status of under-5 year olds** - Indicators used include:

-  Weight in relation to age (W/A)
-  Weight in relation to height (W/H)
-  Height in relation to age (H/A)
-  Mid-upper-arm circumference (MUAC)

E.g. **Ways to define Obesity**

Body Mass Index (BMI) = Weight (Kg)/Height (m)

- Under weight    \* < 18.5                      <18.5
- Normal                \* 18.5-22.9                      18.5-24.9
- Over weight        \* 23.0-27.5                      23.0-29.9
- Obesity                \* >=27.5                      >=30

\*For Asians

## E.g. **Classification of nutritional status**

- internationally accepted categories already exist, which are based on standard growth curves
  - For the indicator weight/age, children are:
    - **Well nourished** if they are above 80% of the standard
    - **Moderately malnourished** if they are between 60% and 80%
    - **Severely malnourished** if they are below 60%
- need to review the literature to find out what definitions have been used by other researchers, so to **standardize** the **definitions** and thus be able later to easily **compare** the **findings** with other studies

## **Factors rephrased as variables**

- Most of the variables usually take on negative values (e.g. lack of personal knowledge)
- formulate them in a neutral way, so that they can take on positive as well as negative values

### **Factors as presented**

#### **in the Analysis Diagram**

#### **Variables**

- |                       |   |                                   |
|-----------------------|---|-----------------------------------|
| • Long waiting time   | → | Waiting time                      |
| • Absence of drugs    | → | Availability of drugs             |
| • Lack of supervision | → | Frequency of<br>supervisory visit |
| • Low dietary intake  | → | Dietary intake                    |



## Operational Definitions For...

- Fatigue?
- Psychological adaptation?
- Pain?
- Happiness?
- Urinary tract infection?
- Wound healing?

- variables to be studied are **selected** on the basis of their relevance to the **objectives of the study**
- 1. **Clear definition** of variables in terms of **objectively measurable facts** (i.e., **operational definition**)
- 2. The **scale of measurement** to be used in data collection

# Measuring Variables

- ✎ To establish relationship between variables →  
observe the variables and record their observations
- ✎ This requires that the variables be measured
- ✎ The quantitative data - have different levels of  
measurement
- ✎ four **levels of measurement**:  
**nominal, ordinal, interval, and ratio**

# Scale of Measurements

 **Nominal Scale**

 **Ordinal Scale**

 **Interval Scale**

 **Ratio Scale**

# 1. Nominal Scale

- ✎ The attribute (**unique quality or feature**) regarded as a characteristic of the variable  
e.g. the variable of **gender** has the unique attributes of **male** and **female**
- ✎ There is **no ordering or ranking** of the attributes
- ✎ A nominal scale simply describes differences between things (by assigning them to categories)
- ✎ Nominal data are, thus, counted data

# 1. Nominal Scale

 classifies data into **mutually exclusive** (non-overlapping)

E.g.

- ☐ Gender (male, female)
- ☐ Eye color (blue, brown, green, hazel)
- ☐ Surgical outcome (dead, alive)
- ☐ Blood group (A, B, AB, O)
- ☐ Religious status

## 2. Ordinal Scale

- ✎ similar to nominal because the measurements involve categories (names)
- ✎ can be rank-ordered, but the distances between the attributes are not the same (have no meaning)
- ✎ the use of an ordinal scale implies a statement of 'greater than' or 'less than' (an equality statement is also acceptable) without being able to state how much greater or less

E.g. ☐ Likert-type scales, Net Promoter Score (NPS)

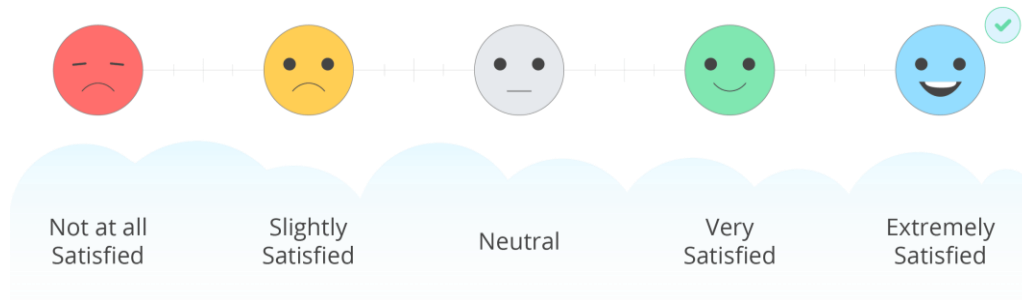
☐ Stage of cancer (stage I, II, III, IV)

☐ Education level (elementary, secondary, college)

☐ Satisfaction level (very dissatisfied, dissatisfied, neutral, satisfied, very satisfied)

# 2. Ordinal Scale

## Likert scale



**Customer satisfaction is key to organizational growth.**







- ☐ 1- Completely agree
- ☐ 2- Somewhat agree
- ☐ 3- Neutral
- ☐ 4- Somewhat disagree
- ☐ 5- Completely disagree



### 3. Interval scale

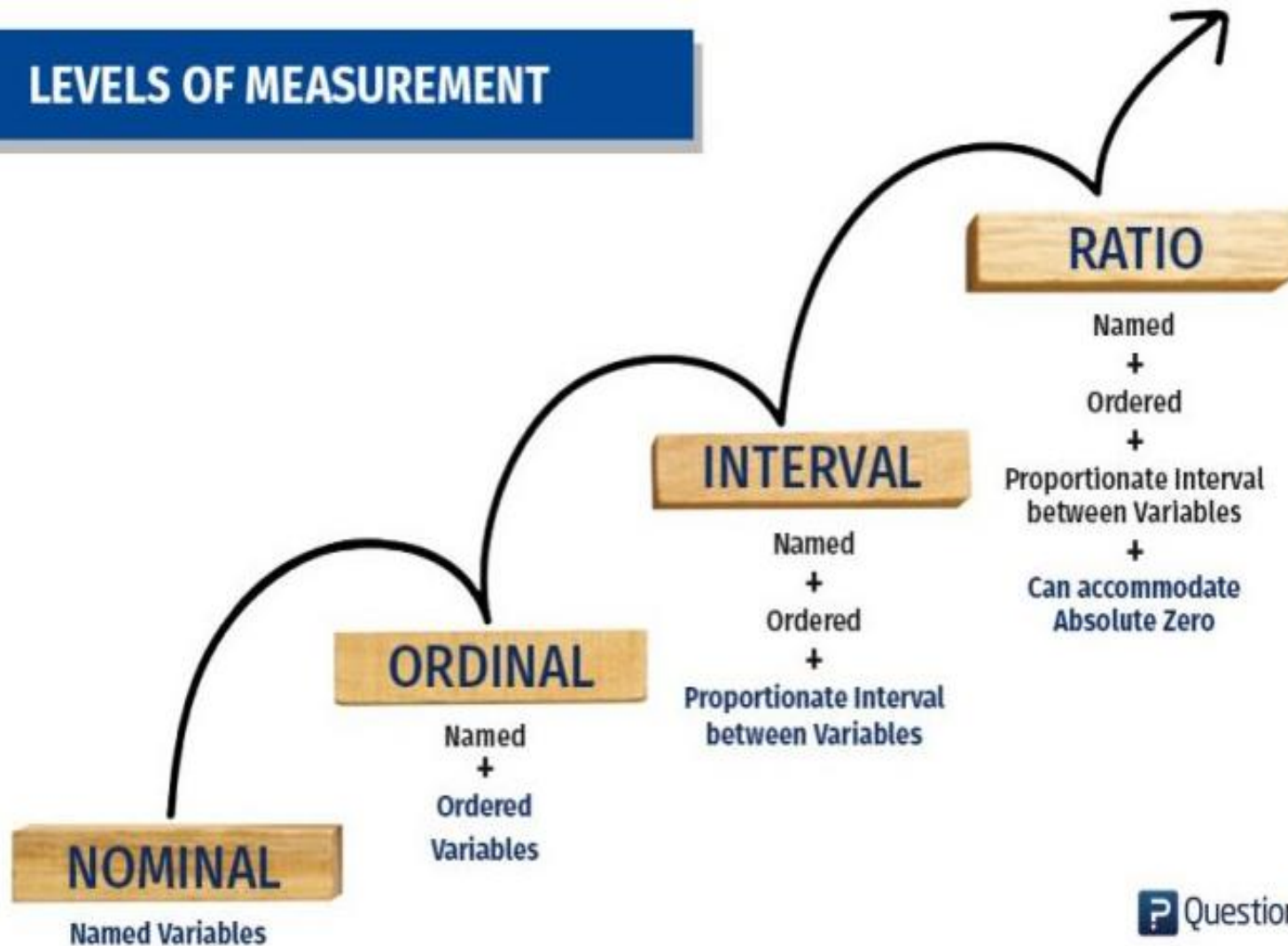
- ✎ Always numeric with **equal intervals**
- ✎ The **distance between the attributes are the same** so they can be interpreted (the difference)
- ✎ Does not have a true zero (zero does not mean for absence/ nothing in this scale) - does not have the capacity to measure the complete absence
  - E.g. The Fahrenheit scale, where  $0^{\circ}$  does not correspond to an absence of heat
- ✎ provide more powerful measurement than ordinal scales - incorporates the **concept of equality of interval** - more powerful statistical measures can be used

## 4. Ratio scale

-  Also numeric
-  Measurements have equal intervals
-  There is a true zero (zero stands for negative existent)  
Eg . distance, height, weight, and time
-  Ratio is the most advanced level of measurement
-  have significant **comparison** which is not possible in case of an interval scale (how many times more of less...)
-  all statistical techniques are usable with ratio scales and all manipulations that can carry out with real numbers can also be carried out with ratio scale values

Eg. Multiplication and division can be used with this scale but not with other scales

## LEVELS OF MEASUREMENT



- proceeding from the nominal scale (the least precise type of scale) to ratio scale (the most precise), relevant information is obtained increasingly

# Example of Defining Variables

## Conceptual Definition

## Operational Definition

## Scale of Measurement

**Age**

Completed Age

Continuous  
(Months)

**Haemoglobin level**

Hb conc. In  
capillary blood  
measured by  
hemoglobinometer




Continuous  
(g/dl)

**Nutritional Status**

Wt in relation to age  
compared to std  
growth curve

Ordinal  
(Low, Medium, High)

# Variables and Data

-  Variables are the characteristics you're studying
-  Data are the values of those characteristics that you measure and record
-  Some of the variables may have produced numerical data, while other variables produced categorical data

## Variables

## Data

Hb level

9.9, 10.5, 10.9, 11.5,

Household member

3, 4, 5, 7, 8

Blood Group






A, B, O, AB

Nutritional status

Low, Medium, High

- ✎ for **analyzing data**, it is important to determine the **type of variable**
- ✎ This is crucial because the type of variable largely determines the **type of statistical techniques**
- ✎ The methods used to display, summarize, and analyze data (data analysis and presentation) depend on whether the variables are numerical or categorical

# Factors to be considered in selection of variables

-  Choose appropriate variables to meet objectives of research
-  List all variables to be studied to solve the research problems
-  Definition of variable must be clear
-  Define independent, dependent, confounding variables
-  Measure variables by appropriate measurement

- ✎ Which variables need to be **operationalized** (by choosing **indicators** to measure them), and
- ✎ What **definitions** are needed for the variables and indicators that have been selected
- ✎ In developing questionnaire, potential **confounder variables** (that might influence the relationship between exposure and outcome) should be included
- ✎ Choose appropriate type of variables for targeted/ desired statistical analysis

E.g. To do correlation between two variables - these variables must be continuous variables

(if you collect data with categorical variable you cannot transform these categorical into continuous variable – age group ~ individual age)



# Reference

Corlien M. Varkevisser, Indra Pathmanathan, Ann Brownlee (2000) *Designing and Conducting Health Systems Research Projects* ,vol.1,no.8,pp.41-51

**THANK YOU!**