Variables and Data

Gbenga Ogunfowokan
Lead, Nigerian Regional Faculty
The Global Health Network
19th May 2017

Objectives

At the end of this presentation you should be able to

1) Define a variable
2) Classify variables
3) Know how to measure variables
4) Have an idea about how a variable can interact with or confound other variables
5) Practice how to code variables

Pre-test

- The ages of participants in a statistic workshop was grouped into 25-30 years, >30-35years, >35-40years, >40-45years, >45-50years. Answer true or false
- 1) >30-35years is a variable
- 2) The measurement of the ages of the participants is a continuous variable
- 3) Age is a common confounding variable
- 4) Age of the participants is a nominal variable
- 5) Age of the participants is a polychotomous categorical variable

Definition

Class	Yes	No	Row total
Nursery	42	58	100
Primary	55	45	100
Junior Secondary	76	24	100
Senior Secondary	81	19	100
Column Total	254	146	400

A variable is not only something that we measure,
but also something that we can describe or manipulate and
something we can control for.

Definition

A variable is a quantity or measurement which can take any of a specified set of values e.g. Age, Gender, BP etc.

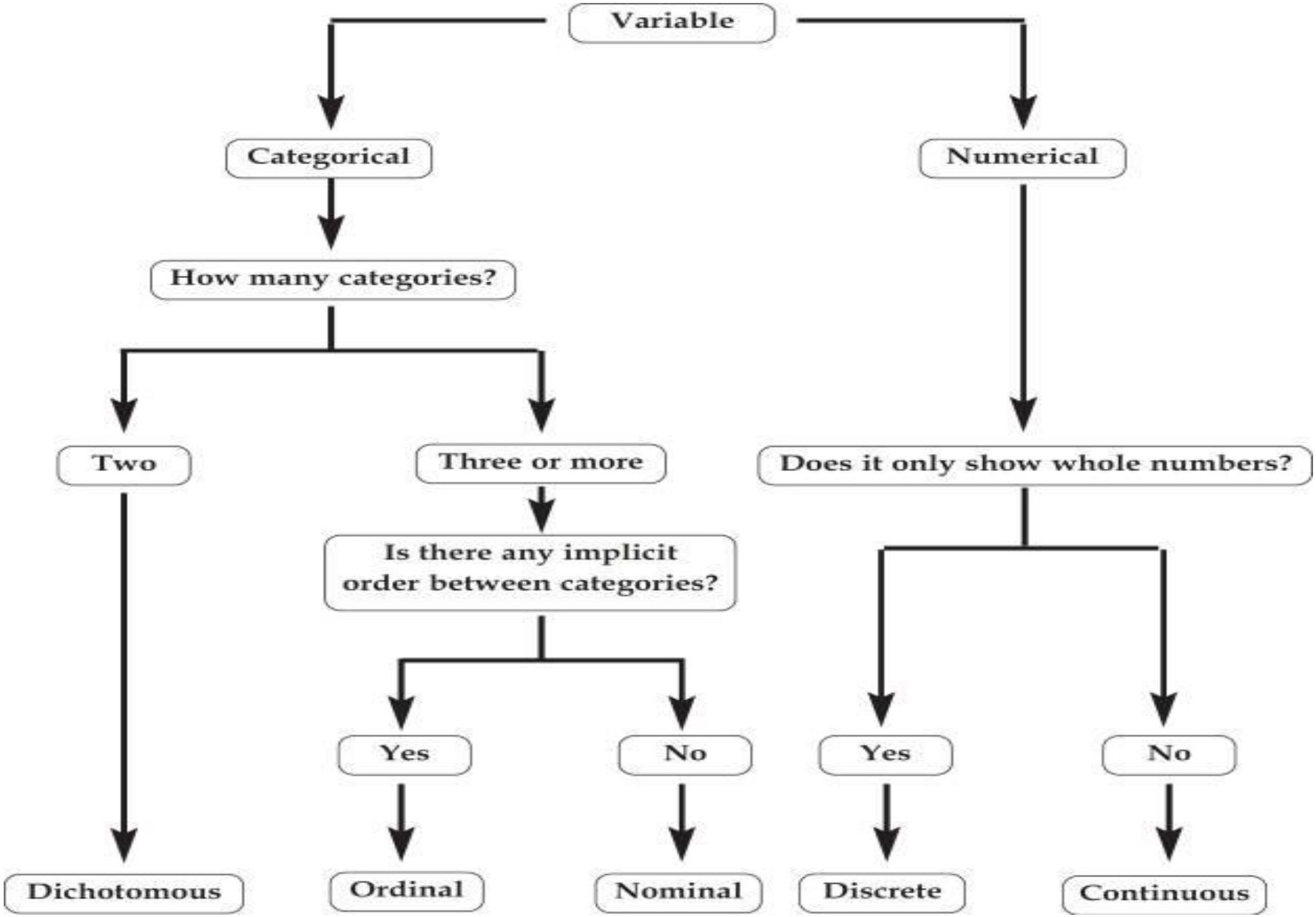
Data

- A set of values of quantitative or qualitative variables.

Classification of Variables

- 1) Based on how they are measured.
- 2) Based on the roles they play.

TYPES OF VARIABLES BASED ON HOW THEY ARE MEASURED



PRACTICE: GIVE 2 EXAMPLES OF EACH

Types of variables

- 1) Qualitative or Categorical
- 2) Quantitative

- NB- Do not confuse qualitative and quantitative variables with qualitative and quantitative study. What is the difference?

Qualitative Variable

- They are categorical or non-numerical.
- A) nominal or not in any particular order and mutually exclusive.
- 1) Binary or dichotomous i.e. two categories e.g. Answer to an objective question (T or F), Other examples ???
- 2) Polychotomous e.g. colours: blue, green, yellow, black or Marital status: single, married, separated, divorced, Tribe. Other examples???

N.B Gives the least information. Only measure of central tendency is mode.

Categorical variables(Nominal)

Gender	Frequency
Male	24
Female	36

Research team	Frequency
Doctors/investigators	30
Research Pharmacists	8
Research Nurses/ study coordinator	10
Research Laboratory Scientists	6
Data Managers	4
Biostatisticians	2

Qualitative variable

- B) Ordinal or ordered or ranked e.g. Age group in years 1-5, 6-10, 11-15,16-20 or Level of Education: primary, secondary, tertiary,

Other Examples: ??????????

Categorical Variables(ordinal)

CODE	Doctors	Frequency
A	Consultants	5
B	SR	20
C	JR	50
D	SHO	80
E	HO	120

Allow for ranking to tell A is higher than B, but does not tell size of difference.

Median & mode can be computed

Quantitative variables

- A) Discrete
- B) Continuous

Quantitative variables

- A) Discrete - A discrete variable is a quantitative variable with a finite number of values. For example, imagine you rolled a six-sided die four times and measured how many times you rolled an even number. What are your possible outcomes? {0, 1, 2, 3, 4}, Number of consulting rooms, Number of participants in this course.
- Other examples ????????????

Quantitative variables

- B) Continuous- A continuous variable is a quantitative variable with an infinite number of values. Take temperature for example. Temperature can take on an infinite number of values, such as 80 degrees, or 80.01 degrees, or 80.0050592359 degrees.
- Other examples ???????

Continuous variables

- 1) Interval variables

OR

- 2) Ratio Variables

Continuous variables(interval)

- for example, temperature measured in degrees Celsius or Fahrenheit but not Kelvin).
- Not only allows for ranking of values but comparing & quantifying the size of the difference between ranks. A temperature of $40^oC > 30^oC$. An increase of 20^oC to 40^oC is 2x that from 30^o to 40^o
- Mean can be computed.
- Zero value does not represent complete absence of the trait. 0^oC is part of the continuum of temperature in Celsius scale.
- Other examples ???????

Continuous variable(Ratio)

- **Ratio** variables are interval variables, but with the added condition that 0 (zero) of the measurement indicates that there is none of that variable.
- So, temperature measured in degrees Celsius or Fahrenheit is not a ratio variable because 0^0C does not mean there is no temperature.
- However, temperature measured in Kelvin is a ratio variable as 0 Kelvin (often called absolute zero) indicates that there is no temperature whatsoever.

Continuous variable

- Other examples of ratio variables include height, mass, distance and many more.
- The name "ratio" reflects the fact that you can use the ratio of measurements. So, for example, a distance of ten metres is twice the distance of 5 metres.
- NB. It provides the most information.

TYPES OF VARIABLES BASED ON THE ROLE THE VARIABLE PLAYS IN A STUDY

VARIABLES

- 1)DEPENDENT, OUTCOME, EFFECT, OUTPUT, CRITERION, ENDOGENOUS, TEST, RESPONSE

VARIABLES

- 2) INDEPENDENT, INPUT, COVARIATE, EXPLORATORY, ORGANISMIC, PREDICTOR, EXOGENOUS, MANIPULATED, TREATMENT, EXPLANATORY

OTHER VARIABLES

- 3) CONFOUNDING or INTERVENING
- 4) INTERACTION
- 5) CONTROL OR CONSTANT
- 6)DUMMY OR INDICATOR

Dependent vs. Independent Variables

- Imagine that a tutor asks 100 students to complete a maths test. The tutor wants to know why some students perform better than others. (**Research Question**)
- Whilst the tutor does not know the answer to this, she thinks that it might be because of two reasons: (1) some students spend more time revising for their test; and (2) some students are naturally more intelligent than others. (**hypothesis**)

Dependent vs Independent variables

- As such, the tutor decides to investigate the effect of revision time and intelligence on the test performance of the 100 students.

- 1) What is/are the dependent variable and why?

- 2) What is/are the independent variable and why?

Answer

- **Dependent Variable**: **Test Mark** (measured from 0 to 100)
- **Independent Variables**: **Revision time** (measured in hours) **Intelligence** (measured using IQ score)

Independent variable

- input, predictor, cause, risk factor

- A variable whose intentional manipulation causes alterations in the measurable value(s) of another variable of interest

Dependent variable

- outcome, output, result, effect

- A variable whose value changes as the independent variable changes.

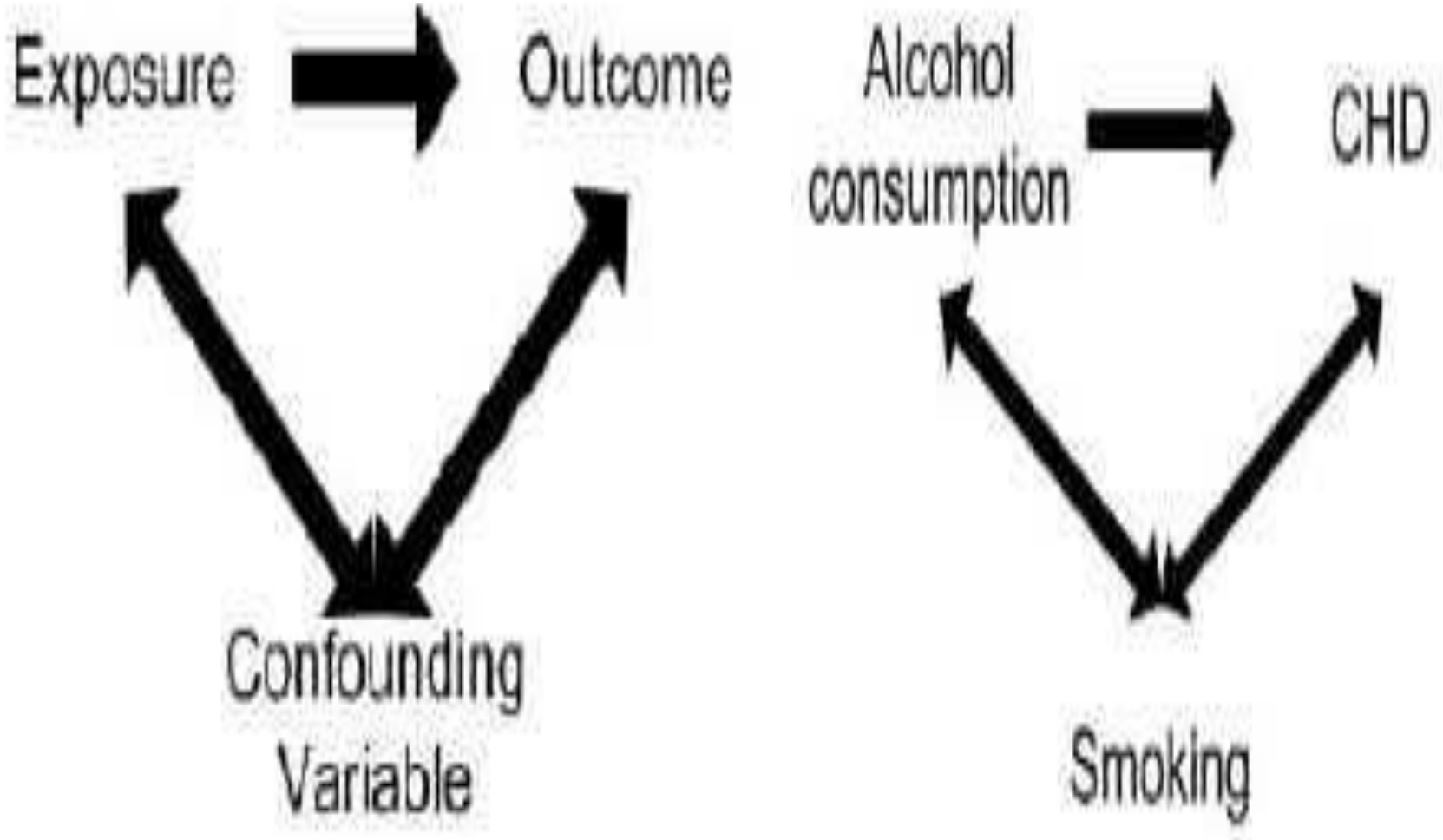
Confounding

- also known as ***extraneous*** variables or ***intervening*** variables
- confounding variables "muddy the waters"
- alternate causal factors or contributory factors which unintentionally influence the results of an experiment, but aren't the subject of the study.
- .

Confounding variable

- Two variables are confounding if their separate effects can not be distinguished.
- Other variables which influence the effect of the independent variable on the dependent variable.
- Demographics: Age, sex, race, occupation, marital status, are common confounders

Confounding Variable



Confounding variable

In order for a variable to be considered as a confounder:

- 1) The variable must be independently associated with the outcome (i.e. be a risk factor).
- 2) The variable must be also associated with the exposure under study in the source population.
- 3) It should not lie on the causal pathway between exposure and disease.

Confounding variables

- For example, a study may find that the risk of lung cancer is more in manual workers. A good investigator will not assume that manual work predisposes to lung cancer, before looking for other possible explanations. The result may, for example, be due to a fact that manual workers are more likely to smoke and it is smoking, not manual work, which is associated with lung cancer.

Confounding variable

- There are several ways to deal with confounding:
- 1) to think of it in planning and designing the study;
- 2) to measure/record the presence of the confounder during implementation of the study;
- 3) and to allow for it in the analysis.
- 4) To randomize or match the participants if possible.
- 5) To avoid contamination after randomization

Confounding variable

- The case mix or patient mix, which refers to baseline differences among research subjects, can be a confounding factor.
- Matching is an important technique for creating a control group by pairing subjects, based on one or more confounding factors.
- An alternative is to use the control-table method, in which stratification is done afterwards.
- Rather than arranging subjects by groups as the study is designed, results are calculated within specified subdivisions.
- When more than a few confounding variables are present, the statistical technique of multivariate analysis is used.

Confounding variable

- As an example of analysis for confounding factors we may look at a study of the relationship between the working status of mothers and the duration of breast feeding. The study may show that women who are employed full-time are less likely to breastfeed for a long duration than women who are employed part-time and women who are not employed.

Confounding variable

- However, the level of education of the mother may be a confounding variable, since it can affect the outcome (duration of breastfeeding) and it may correlate with the working status. Before blaming work for the shorter duration of breastfeeding, there is a need to consider the confounding factor of education.

Confounding variable

- Stratification may be used. A cross-tabulation table may be constructed for mothers at different educational levels, for example those who had no schooling, less than 5 years of schooling, 5–9 years and 10 years or more. For each table, we look at duration of breastfeeding in mothers who are employed full-time, employed part-time and not employed.

Confounding variable

- An alternative way of considering this confounding factor is matching at the design and implementation phase. For each employed mother with less than 5 years of schooling, we would choose a non employed mother with a similar educational level.

Confounding variable

- Crude rates are the terms used when results have not been adjusted for confounding factors.
- Adjusted rates are the terms used when results have undergone statistical transformation to permit fair comparison between groups differing in some characteristic that may affect risk of disease.

Interaction (effect modification)

- Interaction occurs when the direction or magnitude of an association between two variables differs due to the effect of a third variable. It may reflect a cumulative effect of multiple risk factors which are not acting independently and produce a greater or lesser effect than the sum of the effects of each factor acting on its own.
- A(TOBACCO)B(OBESITY) =C(LIVER CANCER) NOT EQUALS TO A(TOBACCO) = C OR B = C

Dummy variables

- **Dummy coding** refers to the process of coding a polychotomous nominal or ordinal categorical variable into dichotomous variables. For example, we may have data about participants' religion, with each participant coded as follows:

A polychotomous nominal variable with three categories

Religion	Code
Christian	1
Muslim	2
Atheist	3

Dummy variable

Full dummy coding for a categorical variable with three categories

Religion	Christian	Muslim	Atheist
Christian	1	0	0
Muslim	0	1	0
Atheist	0	0	1

Dummy variable

Dummy coding for a categorical variable with three categories, using Atheist as the reference category

Religion	Christian	Muslim
Christian	1	0
Muslim	0	1
Atheist	0	0

Dummy variable

A categorical or nominal variable with three categories

Religiosity	Code
Atheism	0
Religious	1

Conclusion

- Variables are the building blocks in research and statistics. A good understanding of variables lay the strong foundation for systematic collection, collation, analysis, interpretation, presentation and publication of clinical research data.

Thank you for your attention