Statistics for Public Health Research

Center for Data Research and Analytics

Session 4: Measurement Scales & Analysis Plan

Table of content

Recapitulation

Quantification of a research question

- Measuring the diseases and exposures
- Variables
- Research database

Dynamics of Population

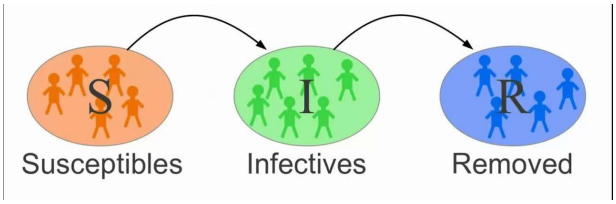
- Ideally populations are dynamic. We often assume a populations as static.
- The static assumption tells, for a certain time interval and a geographic region, the population characteristics will remain reasonably the same
- Suppose we want to know the proportion of children in Bangladesh.
- During the study time, lot of children will born, die, migrate, and immigrate. Overall, the study population will not change that much within a short time

Water in the bucket is indicative to the population dynamics

Study finding will not remain valid if you make the wrong assumption

Example: Dynamics of a population that affects

- Suppose, we are interested in measuring the disease burden of COVID-19
- Counting the number of infections is not suitable as the number depends on how many people are susceptible and How infectious the disease is.
- At the second day of an epidemic, no new infection might observe, but after two months it could reach to the millions per day. For example, COVID-19
- Rate of new infection (incidence rate) for this situation would be better that prevalence



Population is divided into three groups and each person will belong to either of the groups

To study the dynamic population, appropriate statistical method is needed to capture the dynamics.

Measurement Scale

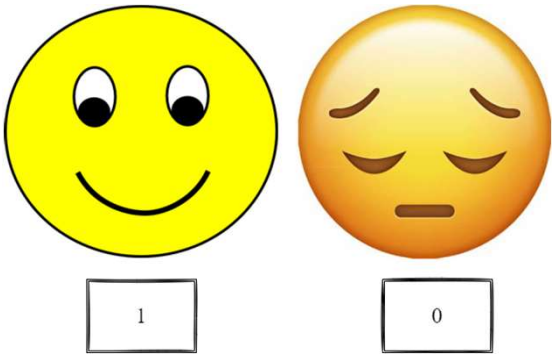
- Measure the content of a material (mass)
- The standards (wight of 1 Kg)
- The mechanism to compare (The scale)
- Get numbers for materials (1 Kg)
  - The scale can precisely measure the mass (Precision)
  - Get the same number repeatedly for repeating weight process (Consistency)
  - We can compare weight with the other object of the mass (Comparability)

A measurement replace a real object with a number.

Does the measurement of an object will provide the same value in the moon?

Quantitative research and measurement scales

- Quantification involves heavily in quantitative research methods. We convert all observations to numerical data.
- There are multiple scales we commonly use such as Nominal, ordinal, interval and ratio scale.
- After measuring all subjects, we store the numbers under a specific name/column which we call variable
- A research might have several variables to capture all necessary information of study subjects.



The statistical methods varies depending on the measurement scale of variables

Nominal Variable

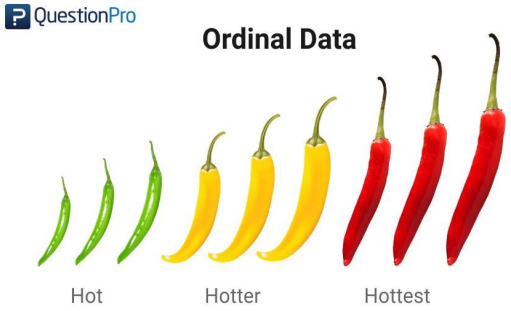
- A number is used to describe a characteristic
  - Eye color: 1 = black, 2 = brown, and 3 = others
  - Patients' division: 1 = Dhaka, 2 = Chittagong, 3 = Rajshahi, 4 = Khulna, 5 = Rangpur, 6 = Sylhet
- There is no concept of order in this scale. Variables measured in the nominal scale are qualitative/categorical variables
- Tables or charts commonly used to present data
- Common statistics such as proportion, percentage are used to analyze categorical variables measured in the nominal scale
- Binary variable is a special case of nominal however, 0, 1 is used for presence of absence of something.

Give few examples of nominal variable that you have ever used in your research

Ordinal Variable

- An ordinal variable is an ordered categorical variable. The distance between values might not be the same
- Example: Educational level might be categorized as
  - 1: No formal education
  - 2: Primary
  - 3: Secondary
  - 4: higher Secondary

Average education 3.5 may not make sense



Careful attention is required for any statistical or numerical procedure.

Interval variable

- The interval variable takes values from a scale where each point placed at an equal distance from one another.
- There is no concept of zero (absent)
- Limited arithmetic operations such as addition and subtraction make sense. So, mean, median, and the summary statics makes sense.
- **Examples of Interval Variable**
  - **Temperature:** Temperature, when measured in Celsius or Fahrenheit is considered as an interval variable.
  - **IQ Test:** A subject IQ cannot be zero, therefore satisfying the no zero property of an interval variable. The level of IQ determines, which interval the score she/he falls in.
  - Likert scale is an example of interval.

Ratio Variable

- The ratio variable has an absolute zero that indicates the absence of something
- Mathematical operations such as addition, subtraction, multiplication and division is possible
- It has an intrinsic order with an equidistant scale.
- All statistical analysis including mean, mode, median, etc. can be calculated on the ratio scale.
- Weight, height can be measure in this scale

It seems difficult to differentiate between interval and ratio scale, while using the scale for the measurement, keep in might how you will deal with the zero.

Measuring an outcome/disease

An outcome can be measured in many ways. For example, breast cancer can be evaluated as follows:

- Perform a diagnostic test. Positives will receive 1 otherwise 0. (Binary/nominal)
- Positives could be further detailed (ordinal scale)
  - The tumor is graded, and cancer is staged for a cancer patient. Cancer has 4 stages (I, II, III, and IV) and 4 different tumor grades (1, 2, 3, and 4)
- Positives even further detailed (ratio scale)
  - Size/volume of the tumor [0, infinity)
- Level of pain due to cancer
  - on the scale of -5(no pain) to 5 (extremely painful)
  - [-5, -4, -3, -2, -1, 0, 1, 2, 3, 4, and 5]

Try it on you on outcome of interest

How to find a measurement

- Choice of a scale depends on how we want to answer our research questions.
- Variance is dependent on scale of measurement. That affects most common statistical tests.
- Some measurement method are biased to measure some variable. For example, measuring the smoking behavior by asking Y/N question might provide the poor result.
- Some measurements are time dependent. For example, a follow-up evaluation of a cancer patient. Time scale definition is important.
- For your research consult with a statistician to finalized the measurements of your variables.

Data Collection Tools

Data collection tools are a setup or system which is ready to measure the variables of study subjects

Data collection tools should be pre-defined to avoid bias

Example: Common data collection tools

- Questionnaire
- Laboratory data collection sheet
- Diagnostic outcomes
- Secondary resources

Recent modified data tools

- Cell phone tracking
- Facebook, twitter, …
- Global Positioning System (GPS) data
- Credit card transactions
- Real time monitoring device (temperature, smoke detector)
- Video and images
- Recorded sound

Mention a few data collection tool you want to use to address your research question

Data Collection Tools

Questionnaire

- Data collected by asking question to the participants or to someone related to the participants
- Potential source of error with questionnaire are response and non-response bias.
- A well-trained person is needed to use a questionnaire to collect information unbiasedly

Appendix 2a : Baseline Questionnaire TB patient
Title of the Research Project: Assessing methods of early detection of active tuberculosis.

A) **Identification**
1. TB Patient ID Number _____
2. Name _____
3. Spouse of TB patient's Name
_____
4. Spouse of TB Patient ID Number _____
5. Date of interview _____
6. Address _____
7. How much time does it take you to travel from your home to the Shyamoli treatment center?
_____ hours
8. Please provide a phone number we can use to reach you:

B) **History**
1. Age _____
2. Sex: 1) Male 2) Female
3. Have you ever been treated for tuberculosis before?
1) yes 2) no 3) not sure
If B3 = No or Not sure skip to B5.
4. For how many months did you take medication for TB? __
5. Have you ever been told you have diabetes?
1) yes 2) no 3) not sure
6. Do you earn income from work outside the home?
1) yes 2) no 3) not sure
If B6= No or not sure, skip to C1

7. During a typical week, how many different persons do you speak with on the job? _____

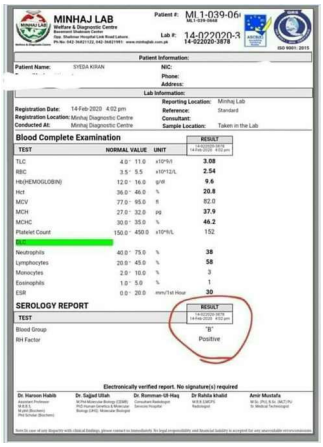
C) **Symptoms**
1. As part of your current illness have you had increased coughing?
1) yes 2) no 3) not sure
If he answer to C1 no or not sure skip to C3
2. How long have you had increased coughing? ____ days
3. As part of your current illness have you had fever?
1) yes 2) no 3) not sure
If he answer to C3 no or not sure skip to C5
4. How long have you had fever?
_____ days
5. As part of your current illness have you had weight loss?
1) yes 2) no 3) not sure
If he answer to C5 no or not sure skip to C7
6. For how long have you been losing weight?
_____ days

D) **Smoking**
1. In your whole lifetime have you smoked > 100 cigarettes?
1) yes 2) no 3) not sure
If D1 = no, then skip to end
2. Do you smoke cigarettes now?
1) yes 2) no 3) not sure
3. For how many years have you smoked cigarettes? _____
4. During all of the time that you have been smoking, how many cigarettes do you smoke on a normal day?
_____

Extensive attention is required to develop a questionnaire to minimize the response bias and inconsistency

Data Collection Tools

- Laboratory diagnostic reports
- Reliable for outcome measurement
- Caution is required to the identification of the sample
- Often require special biosafety labs
- Often difficult to interpret the outcome values



MINHAJ LAB

Patient Information:

Patient Name: SYEDA KIRAN

Lab Information:

Registration Date: 14-Feb-2020 4:02 pm
Registration Location: Minhaj Diagnostic Centre
Conducted At: Minhaj Diagnostic Centre
Reporting Location: Minhaj Lab
Reference: Standard
Consultant:
Sample Location: Taken in the Lab

Blood Complete Examination

TEST	NORMAL VALUE	UNIT	RESULT
TLC	4.0 - 11.0	x10^9/l	3.08
RBC	3.5 - 5.5	x10^12/L	2.54
Hb(HEMOGLOBIN)	12.0 - 16.0	g/dl	9.6
Hct	36.0 - 46.0	%	20.8
MCV	77.0 - 95.0	fl	82.0
MCH	27.0 - 32.0	pg	37.9
MCHC	30.0 - 35.0	%	46.2
Platelet Count	150.0 - 450.0	x10^9/L	152
Neutrophils	40.0 - 75.0	%	38
Lymphocytes	20.0 - 45.0	%	58
Monocytes	2.0 - 10.0	%	3
Eosinophils	1.0 - 5.0	%	1
ESR	0.0 - 20.0	mm/1st Hour	30

SEROLOGY REPORT

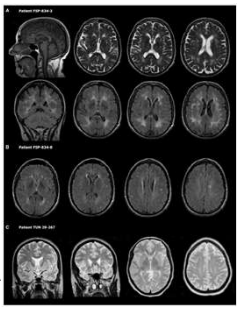
TEST	RESULT
Blood Group	'B'
RH Factor	Positive

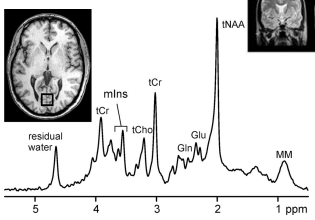
Electronically verified report. No signature(s) required

A laboratory report could provide multiple variable measurement.

Data Collection Tools

- Data can be more complex to interpret
- For example, MRI and MR Spectrometry
- Measuring size of the tumor from the image
- Metabolites in brain tissue from MRS
- Metabolites in brain tissue from tissue sample
- A single data stored in data matrix





Analyzing this kind of data requires special knowledge and expertise. We will not discuss this kind of difficult outcomes in the training

Research Database

- Once data are collected, the data stored in some spreadsheet. The sheet included all subjects' information selected in the sample. We perform statistical analysis on this dataset
- Each row represents data of a single subject.
- Each column represents a variable
- Variables related to the outcome's stores data about the parameter of the population

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	ID	Last Name	First Name	City	State	Gender	Student Status	Major	Country	Age	SAT	Average score (grade)	Height (in)	Newspaper readership (times/wk)
2	1	DOE01	JANE01	Los Angeles	California	Female	Graduate	Politics	US	30	2263	67	61	5
3	2	DOE02	JANE02	Sedona	Arizona	Female	Undergraduate	Math	US	19	2006	63	64	7
4	3	DOE01	JOE01	Elmira	New York	Male	Graduate	Math	US	26	2221	78	73	6
5	4	DOE02	JOE02	Lackawana	New York	Male	Graduate	Econ	US	33	1716	78	68	3
6	5	DOE03	JOE03	Defiance	Ohio	Male	Graduate	Econ	US	37	1701	65	71	6

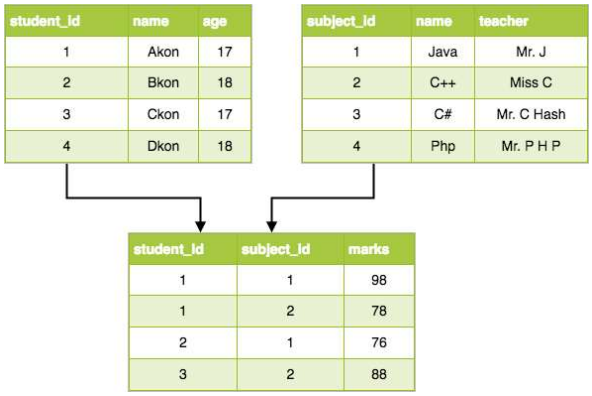
An artificial example of a research database spreadsheet

Jaynal will introduce the datasets with more detail for statistical analysis

Relational Database

Database could be difficult

- For example, a study might maintain two different data table. One for the questionnaire response and the other for the lab report. Two data table may be maintained in two different institutions. A column remains common in both data table to match the data. These columns are called key variables.
- Sometimes multiple variables are used to define a key variable.



student_id	name	age
1	Akon	17
2	Bkon	18
3	Ckon	17
4	Dkon	18

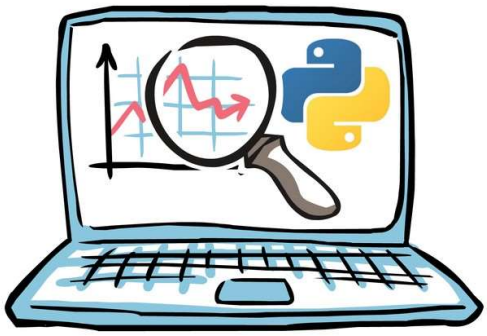
subject_id	name	teacher
1	Java	Mr. J
2	C++	Miss C
3	C#	Mr. C Hash
4	Php	Mr. P H P

student_id	subject_id	marks
1	1	98
1	2	78
2	1	76
3	2	88

Take support from an expert to setup, manage and use a research database

Preparing for Data Analysis

- Data analysis involves finding evidence in response to the research questions.
- Statistical analysis plan should be predefined in the research protocol.
- For exploratory or descriptive research, additional analysis seems useful to develop the research hypothesis.
- Explore data for unusual observations
- Don’t run after p-values. P-value is not an important finding



Prepare your software you will be using. Import data from the reliable source. Then proceed with summary statistics first. See what summary statistics tells