Statistics for Public Health Research

Center for Data Research and Analytics
Session 2:

Fundamentals of Statistical Methods in Public Health Research

Session-2 Foundation of Statistical Methods



- Population and parameters
- Sample and estimators
- Research question
- Random vs. biased sample
- Data collection
- Sources of bias

Center for Data Research and Analytics

17 December 2021

Population



- The population is the complete collection of similar objects, items, or events
 - Example: Males of Bangladesh, all trees in Bangladesh.
- Researchers *define* a population. Population varies by research objectives
- Each object included in the population is a study unit. We call it a subject or study subject.
 - For the male population in Bangladesh, males are subjects.
- A population usually varies by time and space. So, to define a population, a researcher should explicitly specify the time and space.
 - For example, female income of Bangladesh in 2021 may be different from 2000.

Center for Data Research and Analytics

17 December 202:

3

What is your population and subject of interest?

Parameter



- The parameter is a measurable characteristic of a population
 - For example, the average monthly income of females in Bangladesh, 2021.
- Research aims to know those measurable characteristics (parameters). A known parameter helps us to make decisions about the population.
 - For example, the number of cancer patient in Bangladesh could help us to plan for adequate cancer care setup in Bangladesh.
 - A kid garment company needs information about the weight distribution of children <5 in Bangladesh.
- Parameters are unknown constants that **specifies** the population.

enter for Data Research and Analytics

17 December 2021

ļ

Research Question



- The very first step of a scientific research
- Researcher ask questions about the parameter of a population
 - For example, what is the **prevalence** of cancer among people in Bangladesh
 - What is the yearly **trend** in cancer reporting in government facilities in Bangladesh
- It is researcher's primary role to pose a research question. A statistician/biostatistician can help you define the study population, quantify the parameter, and restate your research question to proceed forward.
- Next couple of slides we have introduced a few research questions and their relevant study populations and parameters

enter for Data Research and Analytics

17 December 2021



Research Questions

- What is the level of CO₂ concentration in the atmosphere in 2021?
- What causes CO₂ concentration increase over the last 10 years in the atmosphere?



The world atmosphere is the population and level of CO2 is the parameter of the for these research questions.

Center for Data Research and Analytic

17 December 2021



Research Questions

- What is the prevalence of breast cancer in Bangladesh in 2021
- What are the main causes of breast cancer among women over age 40 in Bangladesh



The population consists of all females in Bangladesh who are at risk of breast cancer until 2021. The prevalence, underlying causes are the parameter.

Center for Data Research and Analytics

17 December 2021



Research Questions

- What is the prevalence of prostate cancer in Bangladesh?
- What are the risk factors for the prostate cancer?
- Which age male are at most risk of prostate cancer in Bangladesh?



The population consists of all males of Bangladesh who are at risk prostate cancer. The average age at which males get prostate cancer is a parameter

Center for Data Research and Analytics

17 December 202:



Research Questions

- What is the COVID-19 incidence rate in Bangladesh in 2021
- What is the risk factors for COVID-19 infection in Bangladesh
- Which age group is most likely to get COVID-19 in Bangladesh
- How many people are need to be vaccinated to control the spread of COVID-19 in Bangladesh



The population consists of all people in Bangladesh who are at risk of COVID-19

Center for Data Research and Analytic

17 December 2021

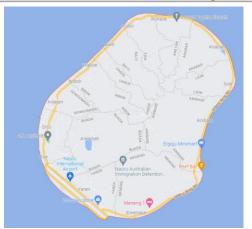
9

Give a few examples of research question? Write your response in github

Research Problem



- Nauru is an independent island country
- Area: 8 Sq. miles
- Population: 11,000
- Female population approx. 6,000
- Female >25 years of age is about 3,000
- RQ: what is the number of breast cancer patients in Nauru?
- What will be your population?
- What is the parameter?



Center for Data Research and Analytic

17 December 2021

Research Problem



- Female >25 years of age is population.
 This age group female were at risk of breast cancer
- There exists an X number of breast cancer patients in Nauru. X is the parameter.
- We want to know the X
- Reach 3,000 females, test them for cancer, and count the number of positives.



Answering research question is easy if the population is known.

Center for Data Research and Analytic

17 December 2021

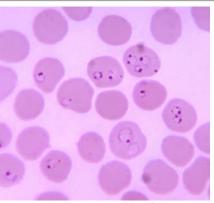
11

What is your parameter of interest? (should be consistent with your research questions)

Research Problem



- A patient is suspected as malaria infected
- What percentage of patients red blood cell (RBC) that are infected with malaria parasites
- The percentage is the parameter that would help the doctor to decide the dose of the medication
- The whole blood cells that the patient have in his/her body is the population
- We cannot get the true parameter



Malaria parasites within the RBCs

Population can be of infinite size. We often don't have access to the whole population. Still, we need the parameter for decision making

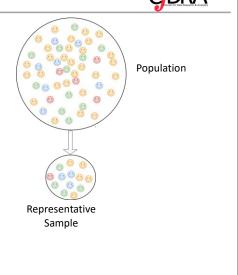
Center for Data Research and Analytics

17 December 2021

Sample and Estimator



- In Statistics, a sample is a representative part of a population
 - For example, blood sample drawn from a malaria patient
- We estimate (not determine) the parameter from the sample and generalize the estimate to the population. The estimate of the parameter is called estimator
- From a representative sample of Bangladesh, we found, life expectancy is 72.6 years. We say that the life expectancy of Bangladesh is 72.6 years.



Center for Data Research and Analytic

17 December 2021

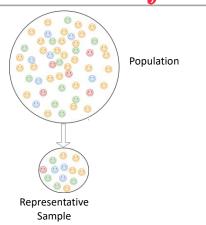
13

Give an example of a biased sample based on your research question and population.

Sample and Estimator

CDRA

- The sample size *n* determines the consistency of the estimate. As *n* grows until the population size, the estimate becomes precisely the parameter.
- We try to reach the population through representative samples generated from the population.



Center for Data Research and Analytic

17 December 2021

Collect Data about the Parameter



- There exist the true parameter in the population
- When we collect data from the population/sample, (philosophically) that parameter generates the data
- Data carries the information about the parameter with uncontrollable error
- Error are expected to be random
- Subjects' data stored systematically under a variable say (Y).

Example

- Bangladesh girls true age at marriage is $\mu = 18$
- A representative sample of married girl from Bangladesh was drawn
- Looking for the age at marriage for the selected subjects
- We will observe $Y = 18 \pm \varepsilon$ where ε is an uncontrolled random error
- The observe data Y = 20, 15, 25, 16.5, ...

Data includes measurement of parameter in a subject. The parameter of the population is the center to generate data.

Center for Data Research and Analytics

17 December 2021

Sources of Bias in Research



- Sampling bias
 - Subjects are sampled in such a way which does not represents the total population
- Measurement bias
 - Using a measurement of the outcome that is biased
 - For example: Use COVID-19 screening kit to measure the incidence of COVID-19
 - Asking about the smoking habit to the smoker and non-smokers
- Due to inappropriate statistical method
 - Wrong statistical method leads to the wrong interpretation
- Confounding
 - Causal association between exposure and outcome due to the confounding factor

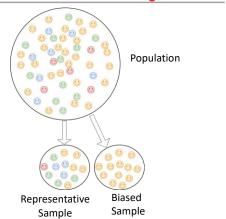
enter for Data Research and Analytics

17 December 2021

Example of Biased sampling



- A Physician in the United Hospital, Bangladesh decided to conduct a research to see the diet pattern among coronary heart disease patients in Bangladesh
- He/she enrolled 100 adult patients who visited the outpatient department for coronary heart disease problem. The sample include patients from both gender, and different age groups



The person who are rich and live in and around the rich community of Dhaka are more likely to be selected in the sample. The sample is biased

Center for Data Research and Analytics

17 December 2021

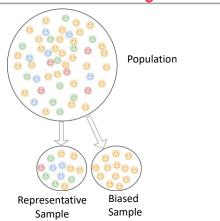
17

Give an example of a biased sample based on your research questions and population.

Example of a Random Sample



- A researcher wants to determine the average height of class V students in Dhaka city.
- The researcher find 10 primary school randomly from the government list
- He/she select 10 class V students randomly from each school
- The researcher has a sample of 100 students



Chance of bias is less in the random sample. Estimate from biased sample is not generalizable to the population

Center for Data Research and Analytics

17 December 2021

18

Can you propose a random sampling plan for your study population?

Measurement Bias



- Using histidine-rich protein 2-based malaria rapid diagnostic test for detecting malaria
- The kit has sensitivity 97.9%, specificity 53% in detecting malaria
- If the researcher intend to use this kit for estimating the disease burden, the prevalence would be unexpectedly higher than the actual



A test with high sensitivity and specificity such as PCR would reduce the bias.

Center for Data Research and Analytics

17 December 2021

Inappropriate Statistical Method Property Value Mean of x 9 Sample variance of $x : s_x^2$ 11 Mean of y 7.50 Sample variance of y: s²_v 4.125 Correlation (x , y) 0.816 **Linear regression line** y = 3.00 + 0.500x0.67 All four data set provides the same statistics, only data that provided figure 11 is reliable. Source: Anscombe's quartet - Wikipedia

Center for Data Research and Analytics

17 December 2021

Summary of Session 2



- In research questions we ask about the parameter of the population
- The population is not accessible, so we use a random sample from the populations
- We estimate the parameter based on the sample and conclude about the population with the estimate
- Bias in research limits the generalizability of the finding from sample to the population
- Bias in research could be due to the sampling, measurement scale and inappropriate statistical method
- Know as much as possible the sources of bias in research
- Discuss with a statistician from the inception of the research and plan the statistical methods before data comes

Center for Data Research and Analytics

17 December 2021

Session 3 Outline



- Practical demonstration
 - Research question
 - Study population
 - Sample

Center for Data Research and Analytics

17 December 2021

Questions OR Comments

Center for Data Research and Analytics

17 December 2021