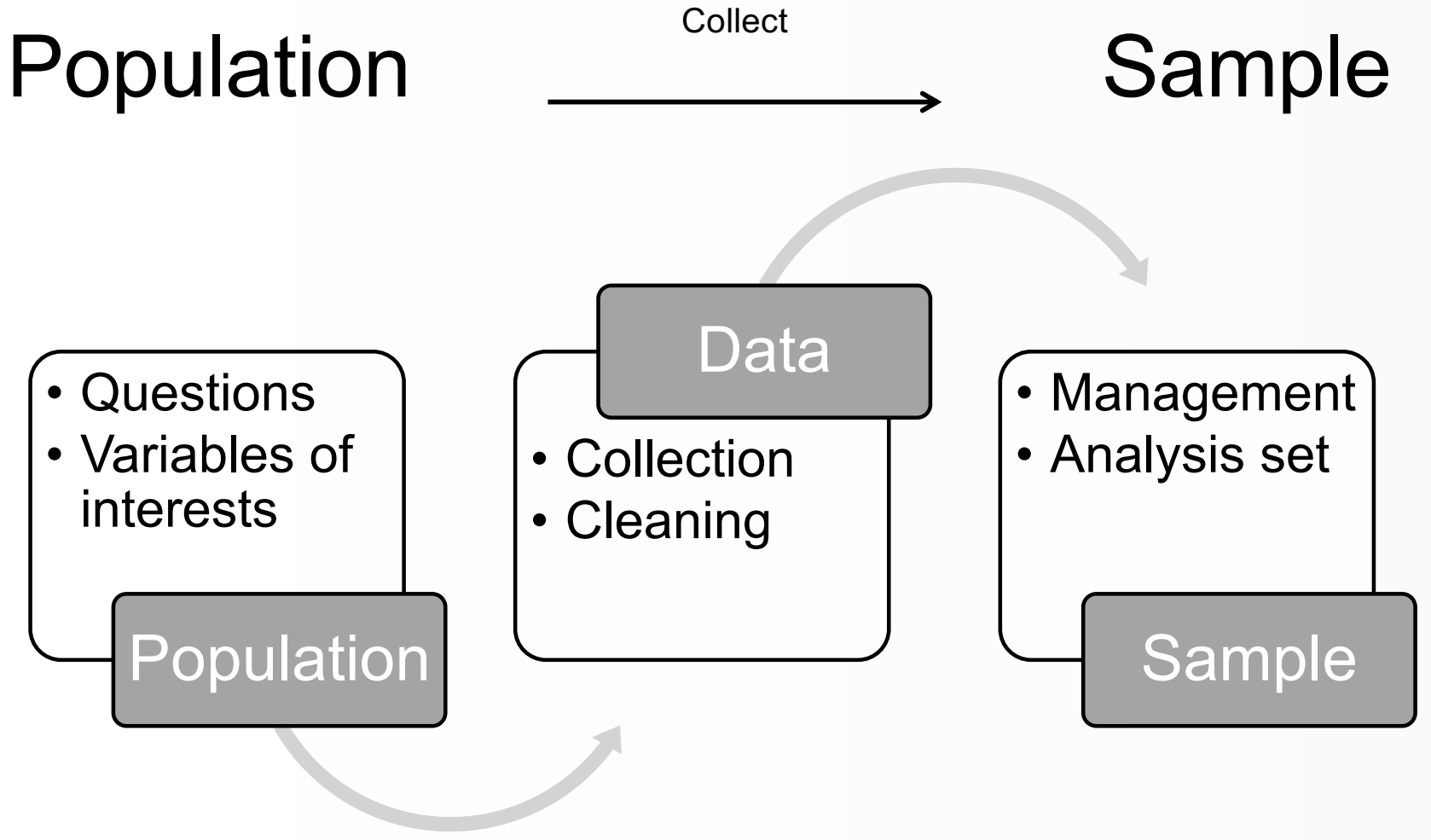
Choosing a Proper Statistical Method

Yuda Chongpison, PhD, MS, MBA
Research Affairs, Faculty of Medicine
Chulalongkorn University

Have you ever done a scientific research?


Population
Collect
Sample
• Questions
• Variables of interests
Population
Data
• Collection
• Cleaning
• Management
• Analysis set
Sample

How to choose a statistical method

- Research Question
- Type of variables or type of measurements
- Designs: Randomized controlled trial (experimental) vs observational design
- Single measure vs repeated measures
  - independent vs dependent
- How many dependent variables (outcome variables)?
  - Are the outcome variables related to each other?

Descriptive vs Inferential Statistics

- What are the differences?
- For which situations, you would use descriptive statistics?
- When would we use inferential statistics?

Descriptive vs Inferential

Descriptive Statistic

- Concerned with the *presentation*, *organization*, and *summarization* of data
- Describe and summarize what we found in tables or graphs
- Describe or compare outcome by gender, age group, or educational levels

Inferential Statistics

- Allow us to *generalize* from our sample of data to a larger group of subjects
- Test a hypothesis

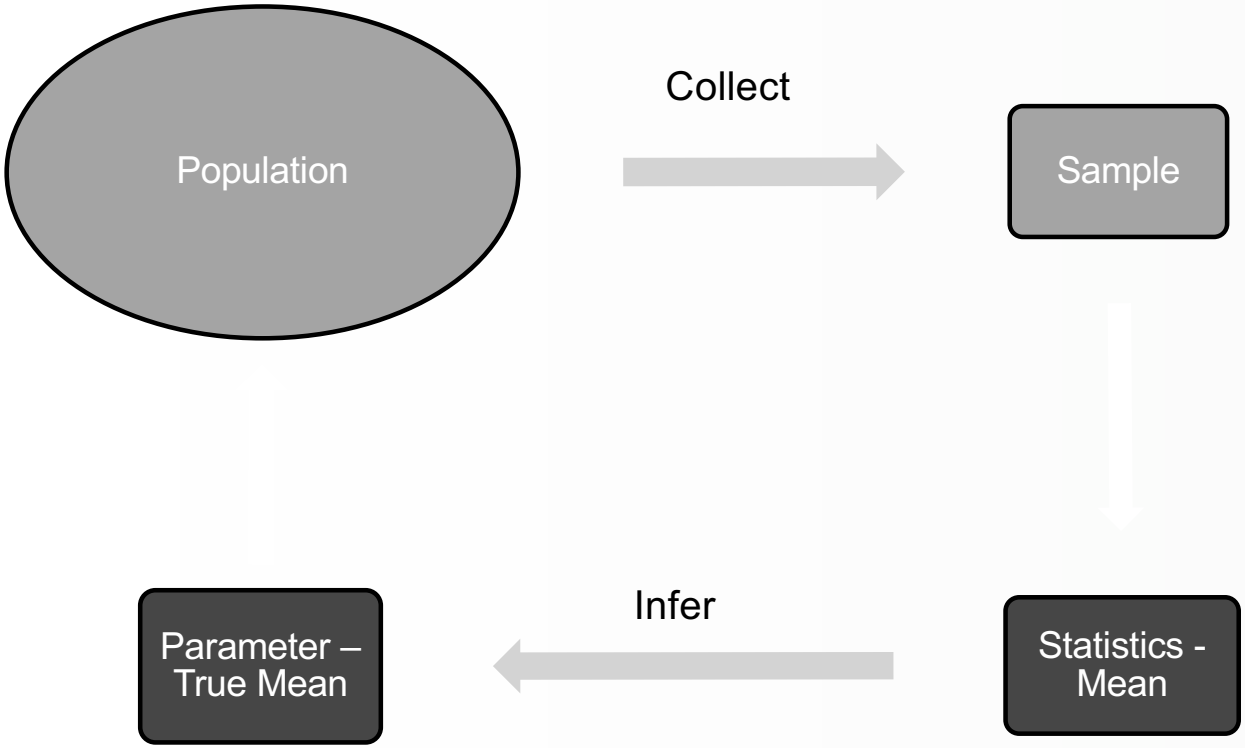
Descriptive Statistics

- Graphs: box plot, histogram, or bar chart
- Calculate measures of center (e.g., means/medians), or variability (e.g., standard deviation/range) from our sample to give the readers an idea of what is going on in the sample

Inferential Statistics

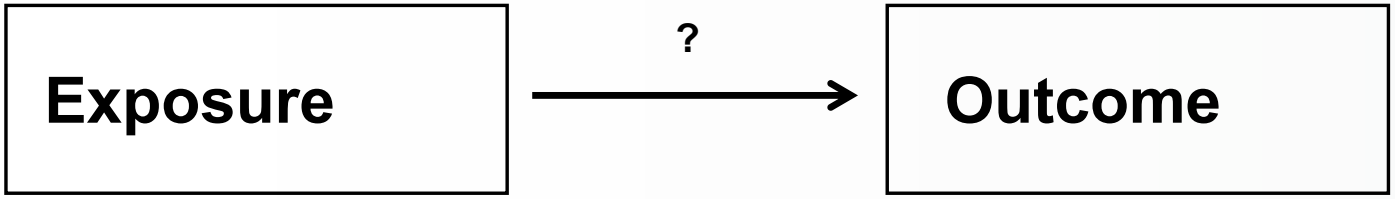
- Allow us to *generalize* from our sample of data to a larger group of subjects
- Test a hypothesis

Map of Statistical Inference



RESEARCH QUESTIONS

Scientific Hypothesis



- An exposure causes an outcome
- Examples
  - Whether cigarette smoking impact the risk of having lung cancer
  - Whether using recreational drug influence the risk of congenital anomalies

Dependent variables: Outcome or response variables
Independent variables: Predictors, explanatory variables, exposure variables, or risk factors

PICO(T) Framework

- P: Patient population
  - Who is your patient?
- I: Intervention, Issue of interest, or investigated condition
  - What do you plan on doing for the patient?

PICO(T) Framework (cont.)

- C: Comparison, control, comparator, or comparison condition
  - What alternative are you considering?
- O: Outcome(s)
  - What do you wish to accomplish?
- T: Time frame
  - How long will it take to reach the desired outcome

Statistical Hypothesis

- Null Hypothesis (H_o)
  - No difference between two groups
  - No difference in mean between two groups
  - No difference in sensitivity between two screening tests
- Alternative Hypothesis (H_a)
  - There is a difference between two groups
  - Mean of anxiety scores in treated group is lower than that in control group
  - One test performs better than the other one

Exercise I

- “The study has been designed specifically to investigate whether there is heterogeneity in the response to exercise training in severe chronic obstructive pulmonary disease (COPD) patients.”

Wedzicha et al.; *Eur Respir J* 1998: 12: 363-369

Exercise II

- “We assessed the association between consumption of ultra-processed foods and the risk of cardiovascular disease (CVD)”

Srour et al. *BMJ* 2019;365:l1451

Exercise III

- “This trial aimed to evaluate the potential benefits of eRAPID for patients and clinicians when added to usual care (UC) during chemotherapy in a population of predominantly early-stage cancer treated with curative intent”

Absolom et al.; *Journal of Clinical Oncology* 39, no. 7 (March 01, 2021) 734-747

Unit of analysis

- From Exercise II, what is the unit of analysis?
- Sometimes called study unit
- The entity that frames what is being analyzed in a study
  - Person
  - Eye
  - Cheek
  - Image: e.g., CT scan, MRI scan

We translate research questions to

VARIABLES

Variables

- An element, feature, or factor that can vary or change
- Types of variable (in Epidemiology)
  - Independent variable
    - Synonyms: potential risk factor, exposure, predictor
  - Dependent variable
    - Synonyms: outcome, effect, end-point

Variables (cont.)

- From Exercise I, what would be variables in the study? Can you guess?
  - "*to compare cardiorespiratory and cerebrovascular functions between Karen women wearing brass neck coils and those without neck coils.*"
- Let's look at various types of measurement

Types of Measurement

- What is being measured?
  - A construct or a concept
  - Examples: speed, muscle strength, ability to clearly see, ability of a tool to detect changes in physiology, complications, satisfactions in one's life

Types of Measurement (cont.)

- How is it measured?
  - Speed: against how fast your walk vs. distance per a unit of time
  - Muscle strength: depending on the muscle? Hang grip strength test
  - Pulmonary Function: forced expiratory volume during the first second (FEV1) vs. self-reported difficulty breathing
  - Complications: number of complications vs. focus on one important complication vs. patient-reported complication
  - Life satisfaction: Yes/No vs. Life Satisfaction Questionnaire

Examples of Type of Measurements

Constructs	Proxy Measures
Health	Self-rated or physician-rated health status
Effectiveness of a Therapy	Number of patients cured
Obesity	Change in cloth size, Body Mass Index

Discrete vs Continuous Data

- Discrete variables have values that can assume only whole number.
  - Number of siblings a person has
  - Number of lesions on a right hand
- Continuous variables have any values within a defined range.
  - Weights and temperature

Binary, Nominal, Ordinal, Interval, and Ratio Data

- Binary - Only take two values
- Nominal (or Categorical) - Only categorize
- Ordinal/Ranked - Categorize and an order of values are meaningful
- Interval - Categorize, put in order, and have equal distances between values
- Ratio Data - Same as interval and has a meaningful zero

Examples

- Binary: Gender - female and male
- Nominal:
  - Marital Status - Single/married/separated/widowed/divorced
  - Regions of Thailand - North, South, Northeast, and Central
- Ordinal/Ranked
  - Smoking status - Never, Former, Current
  - Cancer stages - Stage I, II, III, IV

Examples

- Interval
  - Number of physicians working at KCMH
  - Number of children a woman has before 45 years of age
- Ratio - length or duration

Exercise – Type of Measurement

Variables	Type of Measurement
COPD status (Yes/No)	
Cancer Stages (Stage I, II, III, IV)	
Age (<45, 45-60, >60)	
Underlying disease (DM, HT, CKD)	
Number of acne per cheek	
Cardiac Output (liter per minute)	

Linking research question and variable

- PICO Exercise I
  - Exercise status
- PICO Exercise II
  - CVD status
- PICO Exercise III
  - Consumption of ultra-processed foods

Consultation Information

- Biostatistics Consultation:
  http://rsapp.md.chula.ac.th/login

- Tuesday 9-12 AM;
  Wednesday 9-5 PM

Research Affairs, Faculty of
Medicine, Chulalongkorn University