SESSIONS 4

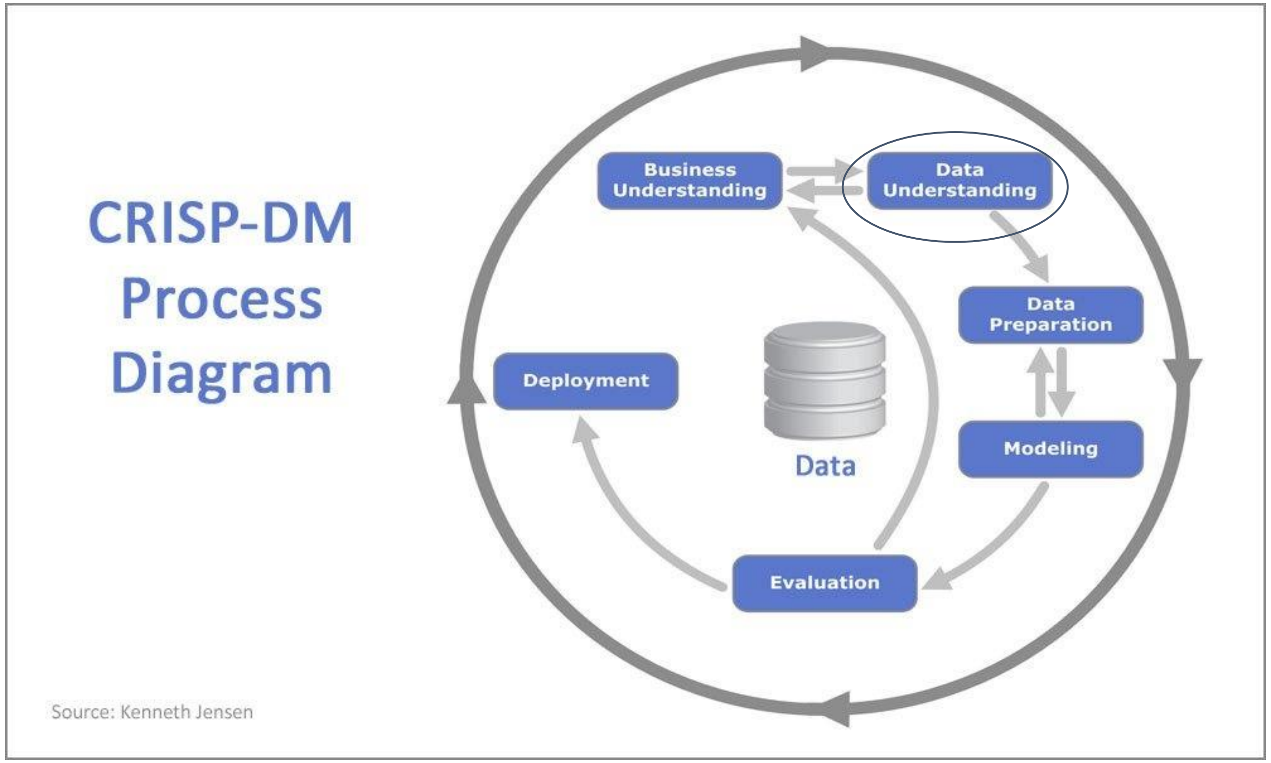
Design Thinking of Statistics

Data Science Program

Purwadhika
Startup and Coding School

Outline

- Experimental and Observational Study
- Population and Sample
- Sampling
- Randomness
- Experimenting



Source: Kenneth Jensen

Design Thinking

What aspect that we design in statistics?

- ❑ Type of Study
- ❑ Population and sample
- ❑ Randomness
- ❑ Sampling
- ❑ Experimental

Type of Study

Experimental Study

- A researcher conducts an experimental study, or more simply, an experiment, by assigning subjects to certain experimental conditions and then observing outcomes on the response variable (or variables).
- The experimental conditions, which correspond to assigned values of the explanatory variable, are called treatments.
- Example: **A/B Testing** of new web design to increase the conversion rate.

Observational Study

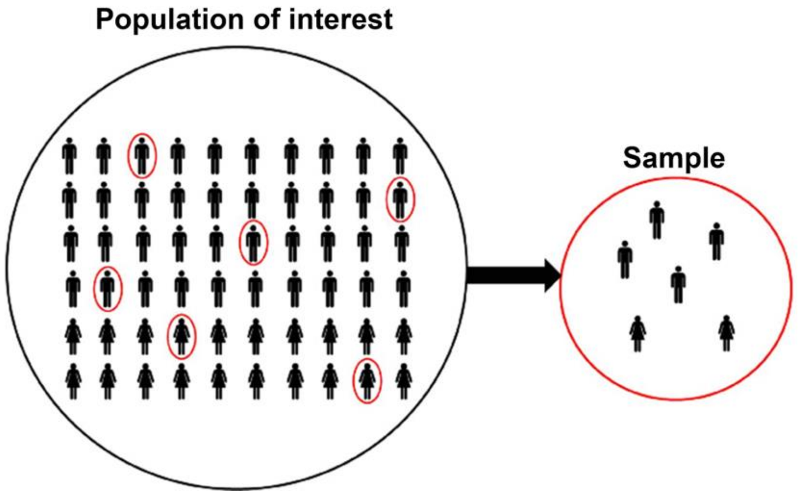
In an observational study, the researcher observes values of the response variable and explanatory variables for the sampled subjects, without anything being done to the subjects (such as imposing a treatment).

Example: **Sample survey** for quick count.

Population & Sample

Population and Sample

- Population can be characterized as the set of individual persons or objects in which an investigator is primarily interested during the research.
- Set of individual or object observed as representation of the population is called sample
- Sample should represent the population.



Population and Sample

- Population always represent the target of an investigation. We learn about the population from the samples
- **Finite Population** is a population that could be physically listed. Ex:
    - Student at Purwadhika
    - Chair at the classroom
- **Hypothetical Population** is a population that was abstract and arise from the phenomenon under consideration. Ex:
    - Factory producing light bulb, if the factory keep the same equipment, using the same produce method, and raw materials. The bulb produced could be consider as hypothetical population

Sampling

Purwadhika
Startup and Coding School

Sampling

- **Sample** should **represent** the **population** of interest
- If possible, define the sampling frame (population that could be physically listed).
- A method to choose appropriate sample is called **sampling**
- Sample obtained randomly
- Be cautious of Sampling bias

Randomness

Purwadhika
Startup and Coding School

Randomness

- Notice that young people have higher proportion suffer from heart disease 88.8%.
- Is that make sense that younger people have tend to have higher risk of suffering from heart disease ?

Let’s analyze data gathered from a hospital.

	Heart Disease	No Heart Disease	Total
Age > 50	57.8 %	42.2 %	100%
Age < 50	88.8 %	11.2 %	100%

Randomness

- Data from table are **not collected randomly**. The reason why younger people have higher proportion of heart disease may be because their awareness to check their health is lower than older people.
- **Younger people** only **check** only **if** they start to **feel** something **wrong** with their body while **older people** because of their **awareness** regardless there is something wrong with their body or not.
- **Data** is still part of population but **not representative**.

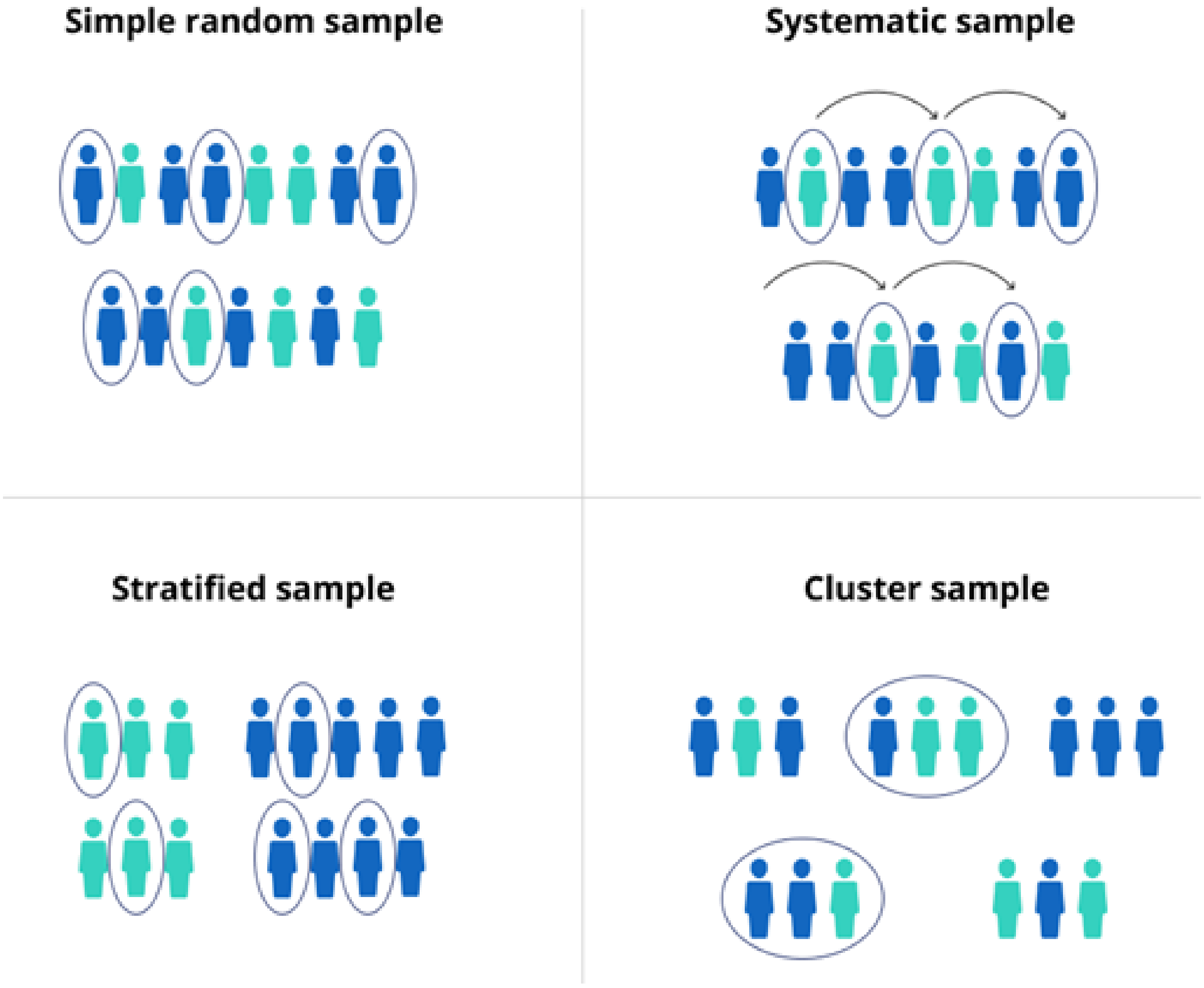
Why do we need sample ?

Things to Consider

- Resource
- Time
- Cost

- When a doctor want to test your blood, Should he/she take all of your blood ?
- When your Mom is cooking, should she eat all the cook or eat only small part of it ?

How to take sample ?



Simple random sample

Systematic sample

Stratified sample

Cluster sample

Purwadhika
Startup and Coding School

Experimenting

Purwadhika
Startup and Coding School

Experimenting

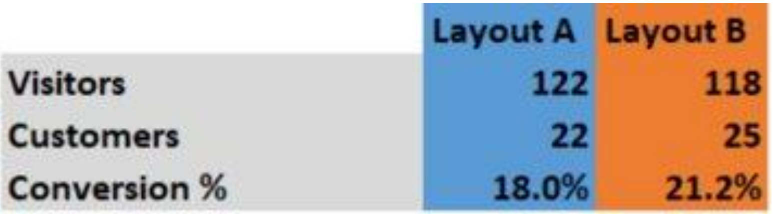
Key Parts of a Good Experiment

- A good experiment has a control comparison group, randomization in assigning experimental units to treatments, and blinding.
- The experimental units are the subjects—the people, animals, or other objects to which the treatments are applied.
- The treatments are the experimental conditions imposed on the experimental units. One of these may be a control (for instance, either a placebo or an existing treatment) that provides a basis for determining whether a particular treatment is effective. The treatments correspond to values of an outcome.
- Randomly assign the experimental units to the treatments. This tends to balance the comparison groups with respect to lurking variables (covariate).
- Replication of studies (sample size) increases confidence in the conclusions.

Example of An Experimentation

Amazon want to test whether their new app design help the company to increase the conversion rate.

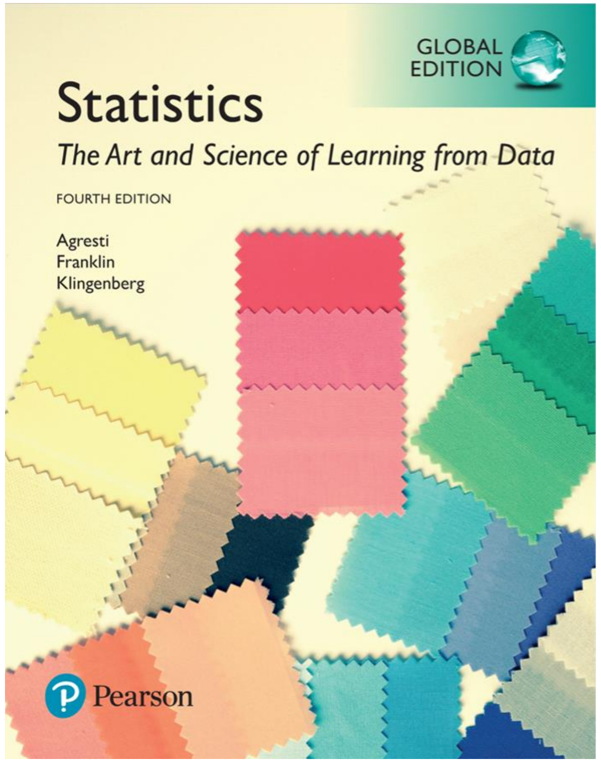
1. Six months ago, the company randomly selected 240 newly signed up users and assigned 122 of them to the control and each 118 to the new designs.
2. Control group went to current design (layout A)
3. The 240 users represent just a small portion of the app’s total users (sample).
4. Subject on this experiment is people.
5. By Assigning randomly, we minimize effect of covariate. (ex gender, age, job, etc) for each group.
6. In the end we want to compare the conversion rate.

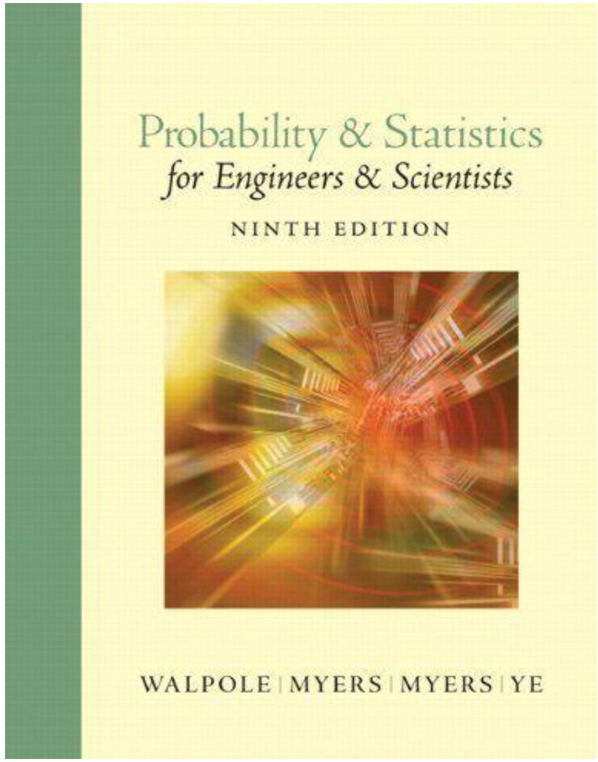


	Layout A	Layout B
Visitors	122	118
Customers	22	25
Conversion %	18.0%	21.2%

Purwadhika
Startup and Coding School

Reference





Purwadhika
Startup and Coding School

Reference

https://towardsdatascience.com/data-science-you-need-to-know-a-b-testing-f2f12aff619a

https://towardsdatascience.com/data-science-fundamentals-a-b-testing-cb371ceecc27

https://www.niagahoster.co.id/blog/ab-testing-adalah/

https://vwo.com/blog/ab-testing-examples/

https://www.scribbr.com/methodology/sampling-methods/