Unit 4: Sampling

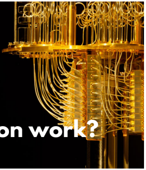
2. Designed experiments and observational studies

Stat 140 - 02

Mount Holyoke College

Dr. Shan Shan

Slides posted at *http://sshanshans.github.io/stat140*



What is Quantum Computing and How does Quantum Error Correction work?

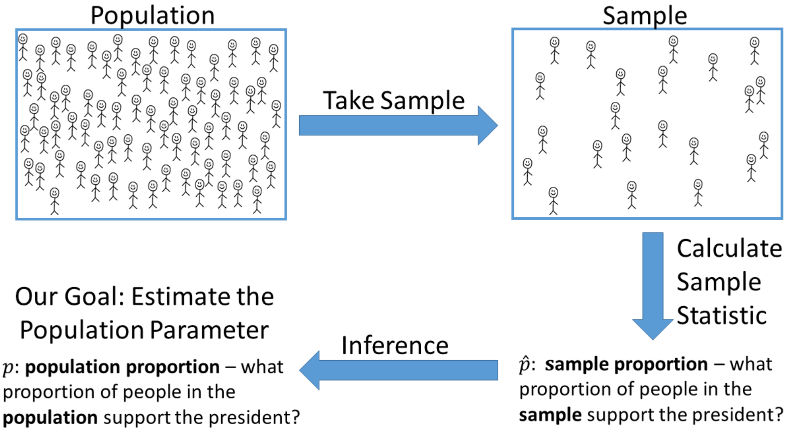
Narayanan Rengaswamy
University of Arizona, Tucson

The field of quantum information and quantum computation has made tremendous progress towards maturing from a theoretical curiosity into a physical reality. Today there are real quantum systems being developed in academia and industry that could soon help us solve some practical problems beyond the reach of the best classical supercomputers. But what is fundamentally different in quantum information science compared to conventional "classical" information and computation?

Sep 17, 3:15PM-4:30PM on Zoom

The talk will only assume some familiarity with linear algebra and basic probability.

MHC Math/Stat Virtual Tea Session 4


Population
Take Sample
Sample
Calculate Sample Statistic
Our Goal: Estimate the Population Parameter
Inference
p: **population proportion** – what proportion of people in the **population** support the president?
$\hat{p}$: **sample proportion** – what proportion of people in the **sample** support the president?

;

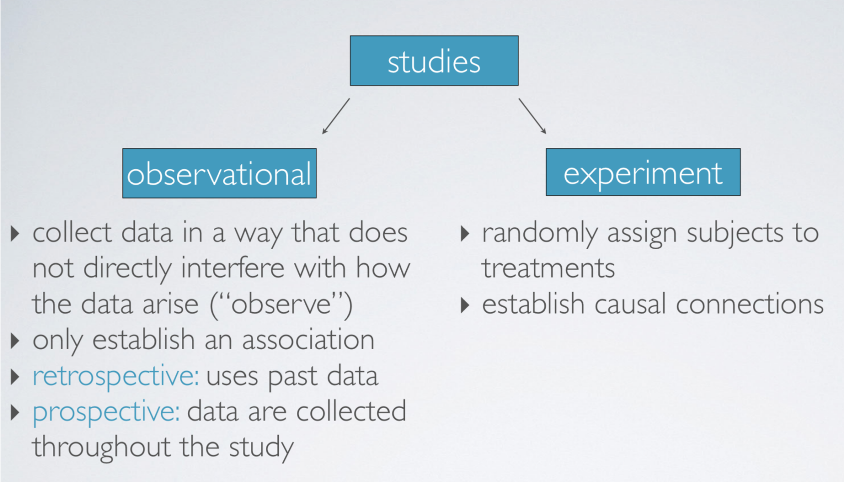


Photo taken from Coursera

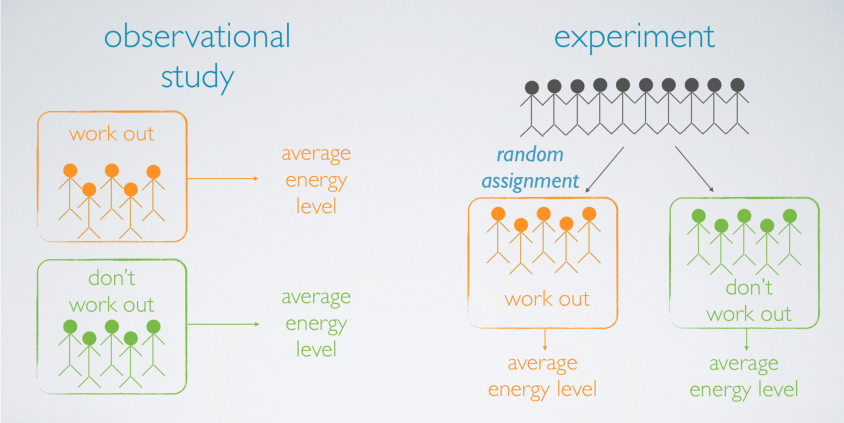


Photo taken from Coursera

Goal: identify causal associations

- factor: a variable to manipulate
- response variable

Does changing the value of the factor(s) cause the value of the response variable to change?

- Experimental units: the individuals on whom or which we experiment (subjects or participants if human)
- Level: The specific values that the experimenter chooses for a factor
- Treatment: The combination of specific levels from all the factors that an experimental unit receives is known as its treatment.
  e.g.: workout and meditation, only workout, only meditation, no intervention

The gold standard for a statistical study is the double-blind, randomized, controlled experiment.

A study is **controlled** if one group receives the treatment and another group does not. (In medicine, that group usually gets either a placebo, or standard medical care, or both.)

A study is **double-blind** if neither the subjects nor the scientists know who is assigned to which group until after the data are collected. This

- prevents subjects in different groups from behaving in different ways;
- prevents scientists from introducing any unconscious bias into the data collection process.

A study is **randomized** if the control group and the treatment group are chosen at random.

Without randomization, the groups may differ in a systematic way. For example, surgeons used to assign only the healthiest patients to receive an experimental new surgical treatment, since those patients could best withstand the invasive procedure. But the outcomes for those patients are not a reliable forecast for how normal patients would respond.

In an observational study, the researcher does not get to determine who receives the treatment.

For example, people who smoke get lung cancer at a higher rate than those who do not smoke. Does smoking cause lung cancer?

Obviously, it would be ethically problematic to do a randomized controlled experiment (one would have to assign 14-year-olds at random to smoke heavily for the rest of their lives). And it would be hard to make this double-blind—people know if they smoke.

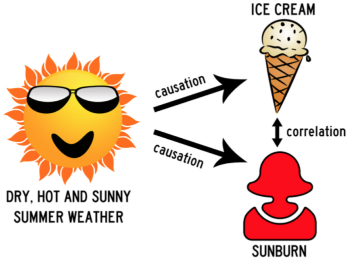
The tobacco lobby used to say no, arguing that:

- there might be a gene that predisposes people to both enjoy smoking and get cancer;
- people who like to smoke may tend to follow unhealthy lifestyles (e.g., alcohol use), and that may be the real cause of lung cancer;
- no randomized, controlled, double-blind experiment (on humans) has shown causation.

The other two arguments from the tobacco lobby carry more weight. The differences between lung cancer rates in the smokers and non-smokers may be due to smoking, or they may be due to a **confounding factor**. In this case, tobacco lobbies suggested two possible confounding factors: genes and lifestyle.

In week 2, we looked at the correlation between ice cream sales and sunburn.

Just because there is a linear relationship between two variables does not mean we have evidence that one variable causes the other.



A confounding factor is associated with both:

- outcome (lung cancer)
- group membership (smoke or not smoke)

One way to try to handle confounding is to make subgroup comparisons that control for possible confounding effects. For example, one could compare the lung cancer rates for smokers who have a healthy life style against smokers who do not.

Do seatbelts save lives?

Seatbelt studies are usually observational (why?). One compares the fatality rates in accidents in which seatbelts were worn to the fatality rate in accidents without seatbelts.

But one has to worry about confounding factors. For example,

- People who don't wear seatbelts may drive more recklessly.
- People who don't wear seatbelts may prefer cars that are not designed with safety in mind.

Some researchers try to control for this by comparing the fatality rates among seatbelt wearers and non-wearers in similar cars, or cars that are thought to have been traveling at the same speed. But this is awkward to do and invites criticism.

In order to control for a confounding factor, one has to guess what it is. But that can be hard and you are never sure that you have thought of everything.

In contrast, with a randomized design, the random assignment of people to the treatment and control groups ensures that there is almost no chance of a systematic difference between the groups. You are unlikely to get most of the people with good genes for lung cancer in one group and all those with bad genes in the other.

1. Designed experiments
2. Observational studies