



## M1-05: Observational Studies and Confounders

Part of the "Basics of Data Science with Python" Learning Badge

Video Walkthrough: <https://discovery.cs.illinois.edu/m1-05/>

### Controlled Experiments vs. Observational Studies

| Controlled Experiments   | Observational Studies   |
|--|---|
| <p>The researcher decides how to divide the subjects into treatment and control groups.</p> <p>-Can be <u>randomized</u> or nonrandomized<br/>           -Randomized controlled double-blind experiments do not have confounders since there should be no differences between the treatment and control groups</p> | <p>The researcher has NO power over assignment into treatment and control groups.</p> <p>-The subjects themselves or "fate" determines who gets the treatment and who does not<br/>           -The researcher just observes what happens.</p> |

Observational studies are done out of necessity. Whenever possible, it's better to do a randomized controlled double-blind experiment. Why?

**Main Problem with Observational Studies:** They can show an association, but it's difficult to make conclusions about causality. Since the treatment and control groups just "happened" they are often very different from each other. These differences confound (mix up) the results when you try to reach a conclusion.

An underlying difference between the two groups (other than the treatment) is called a confounding variable. *Confounding variables (or confounders) are common in observational studies. Good studies take great care to reduce confounding.*

Confounders make it difficult to prove causation by association or correlation.

-Did the treatment cause the response or is the treatment simply associated with the response?

-Maybe both treatment and response were caused by a third confounding factor.

\*If the treatment truly does cause the response, there will be a causal link explaining how or why the treatment itself is causing the response.

\*If the treatment does not cause the response, then there could be a confounder that's making it look like the treatment is causing the response.



## M1-05: Observational Studies and Confounders

Part of the "Basics of Data Science with Python" Learning Badge

Video Walkthrough: <https://discovery.cs.illinois.edu/m1-05/>

---

**Puzzle #1:** Coffee and lung cancer. Studies in the 1960s showed that coffee drinkers had higher rates of lung cancer than those who did not drink coffee. Because of this, some people identified coffee as a cause of lung cancer. Do you think this is true?

**Puzzle #2:** A recent study was done and found that people who wear sunscreen regularly were **more likely** to get skin cancer than those who do not. This goes against what doctors had originally thought. Answer the following questions below based on the information given.

- a) What type of study is this?
  
- b) Does the study show that sunscreen causes skin cancer?
  
- c) Based only on the info below, classify the following as confounders, causal links, or neither.
  - i) Chemicals in Sunscreen- When applying sunscreen to your skin, you're exposing yourself to certain harmful chemicals that may cause skin cancer.
  
  - ii) Genetics- Cancer is known to run in some families, making it more likely for you to get skin cancer whether you wear sunscreen or not.
  
  - iii) Health problems- Studies have shown that too much sun exposure is bad for your health.
  
  - iv) Fair Skin- People with fair or lighter skin easily get sunburned and therefore wear more sunscreen to prevent this. Also, people with fair skin have a higher risk of getting skin cancer.