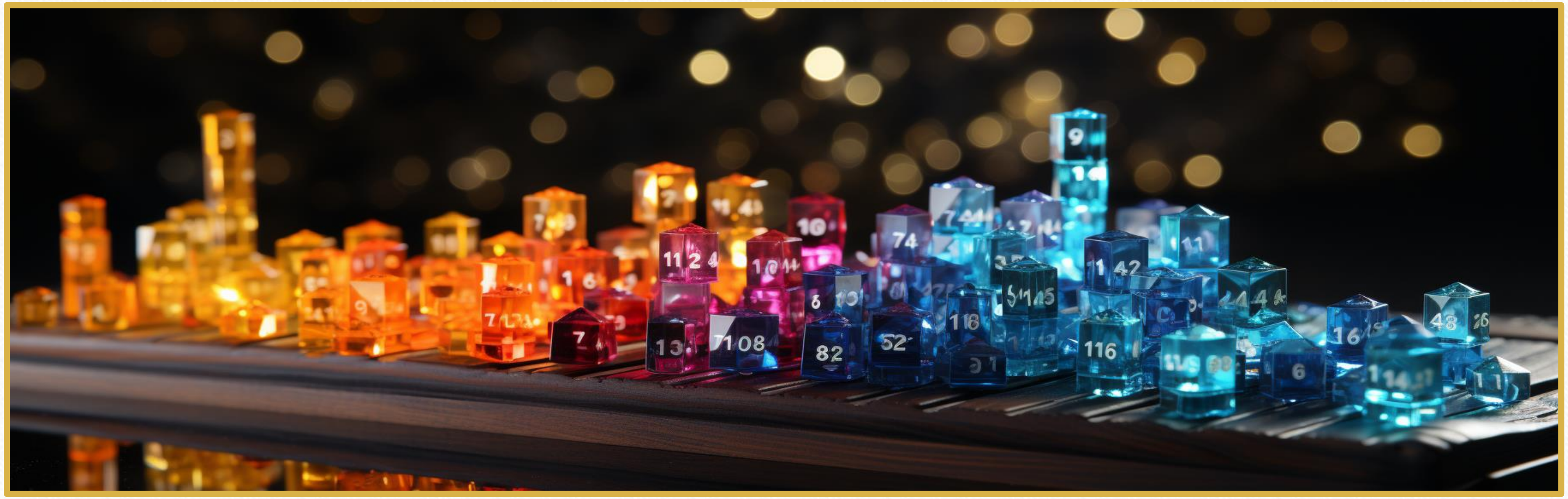
UXG1205 Lecture
13. Correlation (Part 2)

LIN QINJIE

Outline

- Observational studies
- Experiments



Observational studies

- Types of observational studies

Observational Studies

- In the field of statistics, an **observational study** serves as one technique to ascertain the existence of a correlation, for instance, between exercise frequency and cardiovascular health.
- In an **observational study**, as the name implies, researchers gather information by simply observing, without exerting any direct influence on the subjects under investigation.

- Examples that are NOT observational studies:
  - Assessing the impact of a new learning method on students struggling with mathematics, for instance, does not fall under the category of an observational study, as the introduction of the method directly affects the participants.
  - Assessing the impact of a novel pain reliever on individuals experiencing chronic pain is not classified as an observational study, as administering the medication directly affects the participants.
- A study where the investigators actively modify the conditions for the participants is termed an **experiment**.

- Examples of observational study vs. experiment:

Observational Study	**Experiment**
Analyzing the dietary habits of individuals with high cholesterol	Testing the impact of a new drug on weight loss
Monitoring the sleep patterns of people working night shifts	Investigating the effects of different fertilizers on plant growth
Reviewing the relationship between screen time and eye strain	Conducting a trial to determine the efficacy of a new vaccine
Studying the link between outdoor activity and mental well-being	Examining the impact of different learning methods on memory retention
Investigating the association between income levels and life satisfaction	Evaluating the effect of temperature variations on battery performance

Observational Studies

- There exist two kinds of observational studies.
  - In a **retrospective study**, researchers enlist participants and then examine past outcomes of interest.
  - In a **prospective study**, researchers bring in participants and then observe future outcomes of interest.

- Consider a study in which 1,000 individuals with diabetes and 1,000 without are enlisted. Each participant is also tested for a specific lifestyle factor that might be linked to the development of diabetes.
- This is a **retrospective observational study** as the outcome of interest (diabetes) has already occurred prior to the research being conducted.

- Now envision a study where 1,000 individuals with a specific lifestyle factor and 1,000 without are enlisted. Each participant is observed for ten years to see if they develop diabetes.
- This would exemplify a **prospective observational study**, as the outcome of interest (diabetes) has not occurred yet.

- Examples of retrospective vs. prospective:

Retrospective Study	**Prospective Study**
100 individuals are surveyed to determine if they have switched jobs in the last two years.	100 students are observed for one semester to monitor their study habits and academic performance.
Researchers review medical records of patients to investigate the relationship between smoking and lung disease in the past five years.	100 adults are followed over a period of two years to observe the development of exercise habits.
Individuals are asked about their dietary habits in the past year and their current health status.	A group of non-smokers is observed for ten years to see if they develop any respiratory issues.
Medical records are examined to analyze the impact of dietary habits on the development of heart disease in past patients.	A group of participants is monitored for a duration to study the impact of a new fitness regimen on weight loss.
A survey is conducted on people's past exposure to sunlight and their current skin condition.	Individuals with a family history of diabetes are followed to observe the onset of the condition over several years.

- A significant limitation of an observational study is the existence of **lurking or confounding variables**. Due to such variables, it becomes impossible to isolate a single factor as the cause of an outcome.
  - For example, if an observational study determines that individuals who engage in regular physical exercise tend to live longer than those who do not, we can only state that there is a correlation between regular physical exercise and increased longevity, we cannot establish causation → It could be that individuals who exercise regularly have other healthy habits that contribute to their longevity, and these habits are **confounding variables**.
- Observational studies are not capable of establishing causation; they can only show correlation.

Observational Studies

- Research spanning several decades has observed the habits and health of both smokers and non-smokers.
- Even though lung cancer is notably more prevalent in smokers, it can't be conclusively stated by researchers that smoking is the cause of lung cancer.

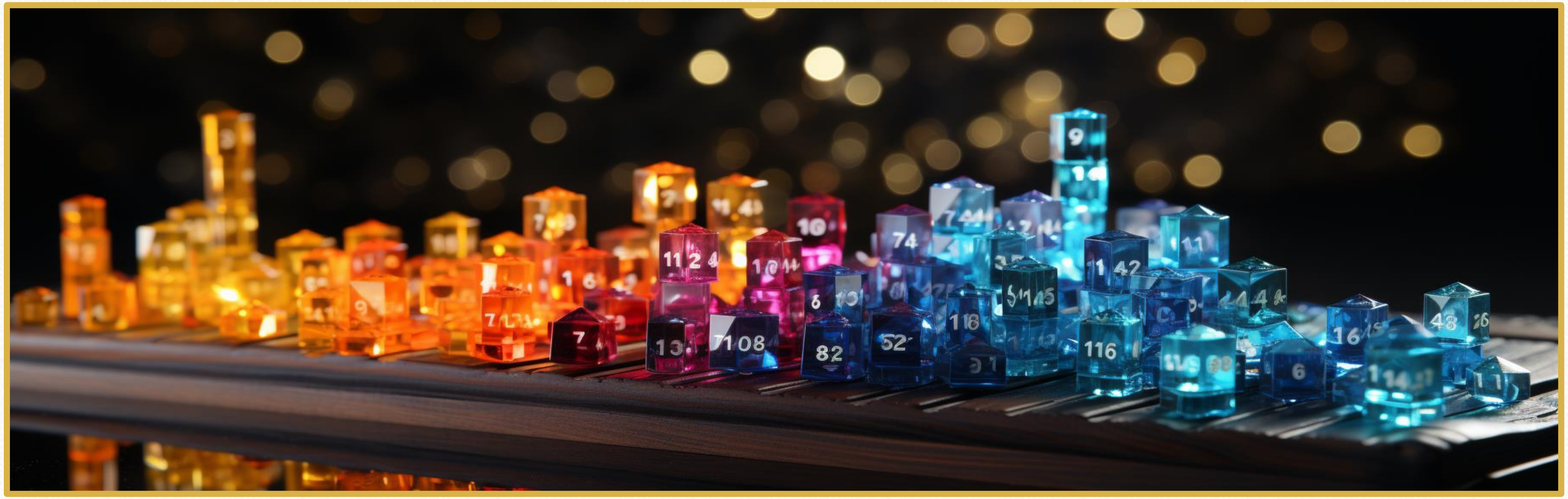
- Much data reveals a positive correlation between the rate of smoking and the incidence of lung cancer, as one increases so does the other.
- However, it's crucial to understand that correlation does not equate to causation.
  - There could be a confounding factor that both induces lung cancer and influences the inclination to smoke or the difficulty in quitting smoking.

- The strength of the correlation between smoking and lung cancer points towards a likely causal relationship.
- But since we only have observational studies as evidence, we cannot conclusively prove it.
  - In fact, attempting to do so would be unethical. Do you know why?
- However, this doesn't mean we should ignore the significant health risk involved! The strong correlation alone is sufficient reason to abstain from or quit this habit.

- Recently, there has been a discussion on the significance of establishing causation when there is an exceptionally strong correlation.
- This is especially with regards to the inability to assert definitive statements regarding medical conditions.
  - For e.g., overemphasizing the necessity for proving causation provides the tobacco industry with grounds to contest anti-smoking policies 😱

- Only **experiments** can establish whether a cause-and-effect relationship exists between two variables.
- Nonetheless, if an observational study reveals the absence of a correlation between two variables, conducting an experiment would be an unnecessary expenditure of time and resources.

- Should an observational study suggest the presence of a correlation, conducting an **experiment** could be considered — provided it is ethical, feasible, and economically viable!
- And that's what we'll explore in the next topic.



Experiments

- Experiments and causation
- Placebo effect

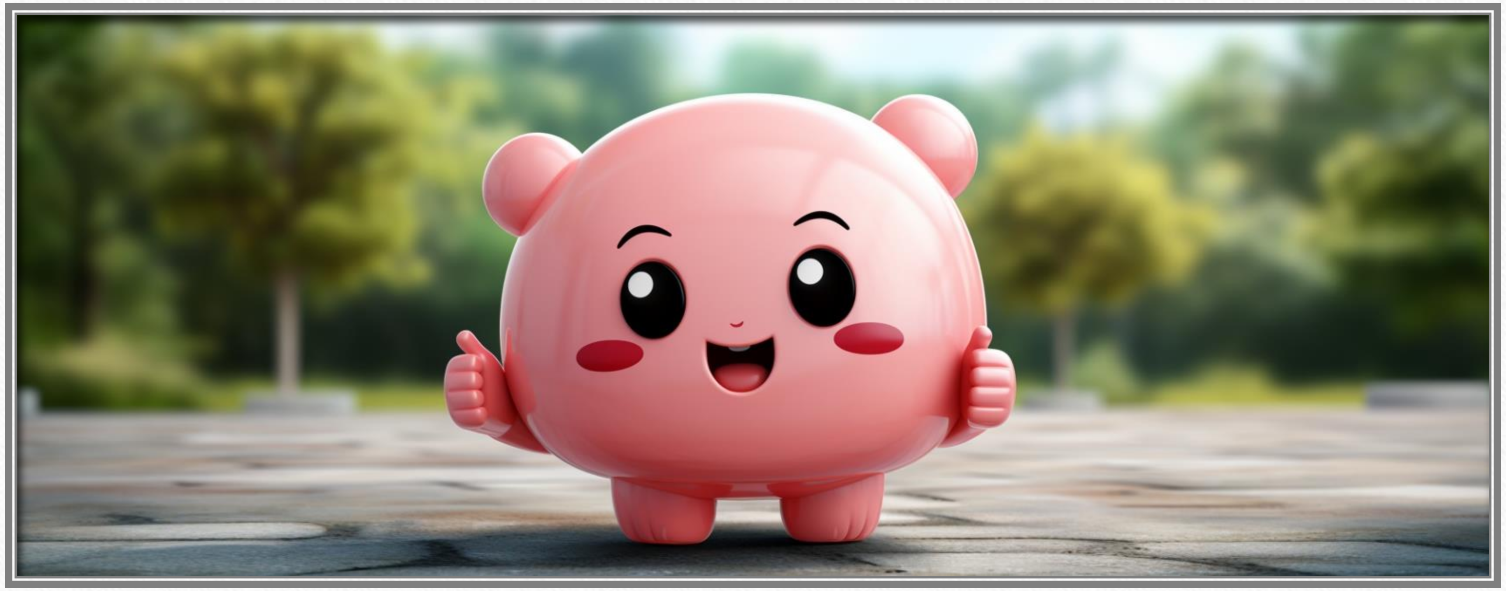
Experiments

- In contrast to observational studies, **experiments** allow a statistician to ascertain causation.
- An **experiment** is characterized by researchers exerting direct influence on the subjects under study.

- The main goal of an experiment is to assesses whether a certain change (the **independent variable**) influences a result (the **dependent variable**). For e.g.:
  - if you want to find out if increasing study hours (independent variable) can improve exam scores (dependent variable).
  - If you want to find out if using a different fertilizer (independent variable) can enhance plant growth (dependent variable).

- In an experiment, participants are randomly assigned to either a **treatment group** or a **control group**. Ideally, the sole distinction between the groups should be the independent variable.
  - **Treatment Group**: The group in an experiment that receives the intervention.
  - **Control Group**: The group in an experiment that does not receive the intervention, used as a baseline to compare with the treatment group.

- An example scenario: Testing the effect of a new fertilizer on plant growth
    - **Treatment Group**: This group consists of 20 plants that receive the new fertilizer. Researchers will monitor these plants to observe any differences in growth, health, and productivity compared to the control group.
    - **Control Group**: This group also consists of 20 plants but does not receive the new fertilizer. These plants are given the usual care and are observed for the same characteristics as the treatment group.
- In the scenario, the only variable that should differ is the application of the new fertilizer.
- By comparing the growth and health of the plants in the treatment group to those in the control group, researchers can determine if the new fertilizer has a significant effect on plant growth.



- Since the sole distinction between the treatment and control groups is the independent variable, any confounding variables are eradicated.
- Consequently, an experiment can establish causation.

Experiments

- Regrettably, just carrying out the experiment can introduce confounding factors, like the **placebo effect**.
- The **placebo effect** happens when the thought of undergoing a test or treatment triggers a response.
- How can we counteract this confounding element?

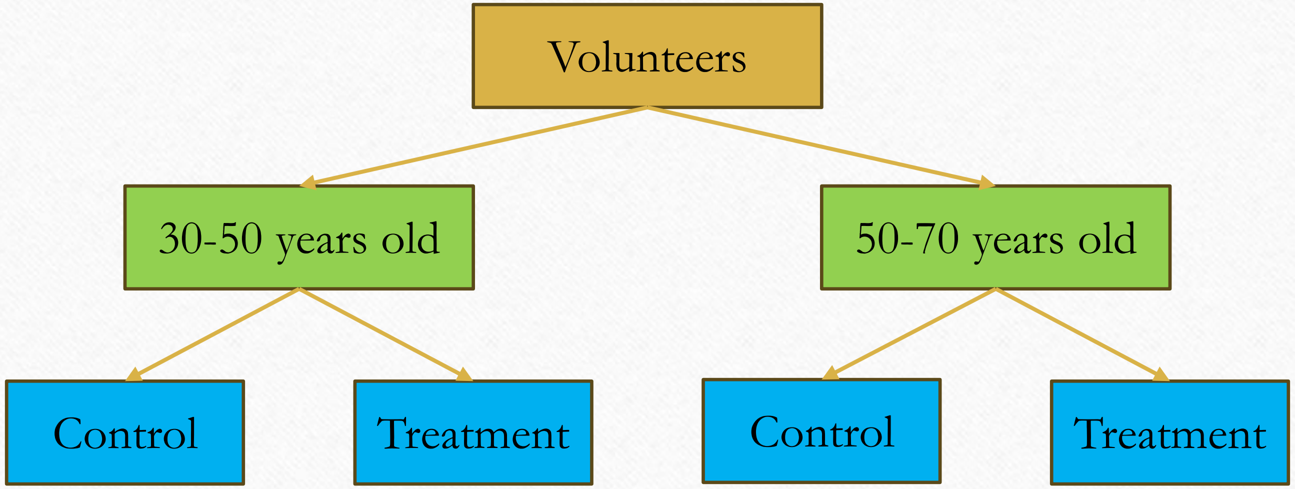
- A straightforward method to address a confounding factor is to ensure participants in both the treatment and control groups encounter it.
- How do we ensure all participants feel the placebo effect?
  - Provide the control group with fake treatments that resemble the actual ones 🤭
  - These fake treatments are also called **placebos**.

Experiments

- To trigger the placebo effect, participants must be unaware of whether they are in the control or treatment group. This type of study is known as a **blind experiment**.
- The control group gets the placebo.
- The treatment group gets the actual intervention.

- But what if the researcher unintentionally treats the control group differently because they are aware they took a placebo?
- Therefore, to ensure the researcher remains unbiased, it's also important to ensure that the researchers don't know which group each participant is in.
- When both the participants and the researchers are unaware of the group assignments, this is termed a **double-blind study**.

- Besides the placebo effect, other typical confounding factors might be age, gender, and income, among others.
- If we identify these variables beforehand, we can use a method called **blocking** to address them.



- **Blocking** functions by grouping subjects based on a confounding factor and then randomly assigning members from each group to either the treatment or control groups.
- This ensures that if outcomes vary based on factors like age, the experiment has already considered that variable and remains valid.

- By removing all confounding factors, an experiment can establish causation.

- If only experiments can determine causation (and observational studies cannot), then why even conduct observational studies?
  - Difficult to find volunteers
  - Ethics

The End