UXG1205 Lecture
12. Correlation (Part 1)

LIN QINJIE

Outline

- Introduction to correlation

- Causation

Introduction to Correlation

- Positive and negative correlation

- **Correlations** can be observed everywhere around us.

- Statisticians examine them to identify associations between actions and results, for e.g.

  - the relationship between smoking and lung cancer or

  - the link between supply, demand, and pricing.

- What do we mean by **correlation**?

- Correlation arises when the occurrence of one event offers a reasonable forecast of another event.

- For example, a student's study hours might give insights into their exam performance; therefore, study time is correlated to exam scores.

- Correlations can be straightforward, like the connection between a light switch and the illumination of a bulb.

- Other examples:

  - the volume knob on a radio and the sound's loudness

  - the relationship between a car's fuel gauge and the distance it can travel

- Question:
  - Do you think property price and property tax are correlated?
  - Do you think the size of a watermelon is correlated to the number of seeds it has?

- When property value increases, property taxes follow suit. Similarly, as the size of a watermelon enlarges, the number of seeds it contains typically increases.

- The above examples illustrate **positive correlations**.

  - Essentially, when one variable experiences an increase, the other is likely to exhibit a corresponding increase as well.

  - When one variable experiences a decrease, the other is likely to decrease as well.

- How about apple harvests and apple prices? Are they correlated?

- When there are lots of apples, the price goes down because there are plenty for everyone. But, if bugs ruin a bunch of apples and there aren't many left, the price goes up because everyone wants them.

- This shows a **negative correlation**.

  - When one variable increases, the other decreases.

  - When one variable decreases, the other increases.

- Below are some examples of positive/negative correlations.

Variable 1	Variable 2	Type of Correlation
Number of Passengers	Available Seats on the Bus	Negative
Hours of Study	Exam Performance	Positive
Amount of Rainfall	Number of Umbrellas Sold	Positive
Speed of a Car	Fuel Consumption	Positive
Daily Exercise Time	Risk of Heart Disease	Negative
Unemployment Rate	Consumer Spending	Negative
Age of Car	Value of Car	Negative
Amount of Sunlight	Plant Growth	Positive

- The sequence of events has no influence on correlation.

- If property values are correlated with property taxes, then property taxes are correlated with property values.

  - Correlation is bidirectional; it functions in both directions.

- Some events might be interconnected through a 3rd variable, like time, leading to a correlation that may have little or no logical significance.

- Such a 3rd variable is known as a **lurking variable** or **confounding variable**.

- For e.g: From 2000 to 2020, there was a significant increase in global internet usage. Concurrently, the number of HDB flats in Singapore also saw a substantial rise. Is there a correlation between these two trends?

  - Even though they appear to be unrelated, the increase in global internet usage and the number of HDB flats in Singapore do exhibit a **positive correlation**.

  - Both occurrences showed a predictable change over time. As the years progressed, the utilization of the internet worldwide expanded, and the construction of HDB flats in Singapore increased.
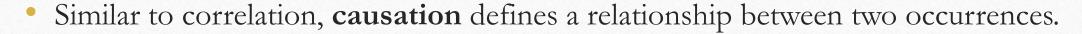
Causation

- Relationship with correlation

- Similar to correlation, **causation** defines a relationship between two occurrences.

- **Causation** occurs when one of the two correlated events is the cause of the other event.

- Pouring water on a plant causes it to grow and increase in height. Therefore, the height of the plants correlates with the amount of water they receive.

- The growth of a plant does not cause the pouring of water. However, the amount of water received still correlates with the height of the plants.

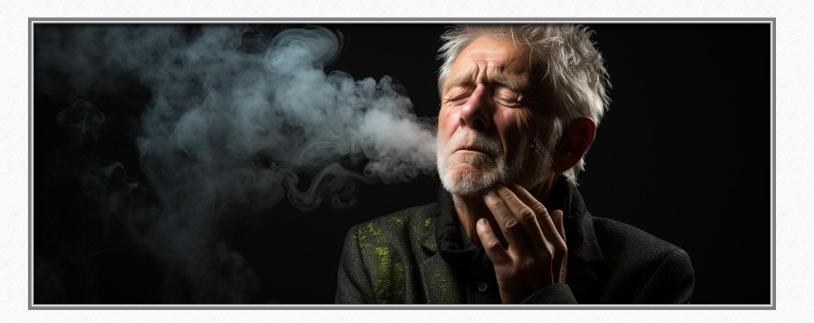 - Correlation can be present even in the absence of causation.

- **Causation**, in contrast to correlation, operates in one direction only. If X causes Y, Y does not cause X.

- Let X represent "It is raining" and Y represent "The ground is wet." Which of the following statements are accurate?

  A. X caused Y.

  B. Y caused X.

  C. X is correlated with Y.

  D. Y is correlated with X.

- Correlation may not necessarily suggest causation. However, causation suggests correlation.

- Example: Smoking vs lung cancer.

  - Lung cancer is correlated with smoking, but lung cancer does not cause smoking to happen.

  - Smoking causes lung cancer, hence they are both correlated.

- Examples:

Example	Type of Relationship
Increase in exercise frequency vs. decrease in body weight	Possible causation
Increase in number of trees planted vs. increase in oxygen production	Possible causation
Increase in coffee prices vs. increase in smartphone users.	Correlation only
Increase in global temperature vs. increase in ice melting at polar caps	Possible causation
Speed of a car vs. number of people at a mall	Correlation only
Increase in fast-food consumption vs. increase in obesity rates	Possible causation

The End