

Lecture 18: Chapter 5

Section 5E: Correlation and Causality

What is Correlation?

Definition 1. Given two variables X and Y , the **correlation** of X and Y is a measure of the dependence of Y on X and vice versa.

Example 1. The variables X = weight gain, and Y = eating at McDonalds, have a high correlations.

Example 2. The variables X = dogs, and Y = cereal have no correlation.

Definition 2. A **scatter diagram** is a graph in which each point represents the values of two variables.

Example 3. The following is a scatter diagram (scatter plot) for the variables, “life expectancy” and “infant deaths per 1000 live births”

Types and Causes of Correlation

We say there is **no correlation** if there is no apparent relationship between two variables.

We say there is **positive correlation** if both variables increase with one another. (The variables have a proportional relationship.

We say there is **negative correlation** between two variables if one increases in the opposite direction with the other.

We may also speak of the strength of the correlation between two variables.

Causes of correlation:

- i) Coincidence
- ii) Both variables may be directly influence by some common underlying cause.
- iii) One variable may actually cause the other.

Guidelines for Establishing Causality

To investigate whether a suspected cause actually causes an effect:

1. Look for situations in which the effect is correlated with the suspected cause even while other factors vary.
2. Among groups that differ only in the presence or absence of the suspected cause, check that the effect is similarly present or absent.
3. Look for evidence that larger amounts of the suspected cause produce larger amounts of the effect.
4. If the effect might be produced by other potential causes, make sure that the effect still remains after accounting for these other potential causes.
5. If possible, test the suspected cause with an experiment. If the experiment cannot be performed with humans for ethical reasons, consider doing the experiment with animals, cell cultures, or computer models.
6. Try to determine the physical mechanism by which the suspected cause produces the effect.

Causes of Causality

There are three terms often used in everyday rhetoric related to causality.

Possible Cause: A correlation between two variables has been discovered, but it cannot be determined if the correlation implies causality.

Probable Cause: There is good reason to suspect the correlation between two variables is causal.

Cause beyond reasonable doubt: Correlation implies causality to such a high degree that it would be unreasonable to doubt the causal relationship between the given variables.