1. We've all heard the phrase **"correlation does not imply causation."** Please explain what this phrase means, what a causal effect is, and how factors can be correlated but not causally related (from a logical perspective--no need for math at this point).

Ans - Correlation and causation are terms which are mostly misunderstood and often used interchangeably. Understanding both the statistical terms is very important not only to make conclusions but more importantly, making correct conclusion at the end. In this blogpost we will understand why correlation does not imply causation.

**Correlation** is a statistical technique which tells us how strongly the pair of variables are linearly related and change together. It does not tell us why and how behind the relationship but it just says the relationship exists.

**Causation** takes a step further than correlation. It says any change in the value of one variable will **cause** a change in the value of another variable, which means one variable makes other to happen. It is also referred as cause and effect.

When a person is exercising then the amount of calories burning goes up every minute.

**Ice cream sales is correlated with homicides in New York (Study)**

As the sales of ice cream rise and fall, so do the number of homicides. Does the consumption of ice cream causing the death of the people?

No. Two things are correlated doesn’t mean one causes other.

**Correlation does not mean causality or in our example, ice cream is not causing the death of people.**

In Majority of the cases correlation, are just because of the coincidences. Just because it seems like one factor is influencing the other, it doesn’t mean that it’s actually does.

There is no causal relationship between the ice cream and rate of homicide, sunny weather is bringing both the factors together. And yes, ice cream sales and homicide has a causal relationship with weather.

“**Correlation tests for a relationship between two variables. However, seeing two variables moving together does not necessarily mean we know whether one variable causes the other to occur. This is why we commonly say “correlation does not imply causation.”**

1. Sometimes, even in the absence of a causal relationship, correlations can still be extremely useful. Please provide an example of when this is the case.

Ans - Sometimes, even in the absence of a causal relationship, correlations can still be extremely useful. Symptoms of illness are vital in arriving at a diagnosis; certain economic indicators may presage a recession; a student’s declining grades may be a sign of problems at home. In each of these cases, one or more “markers” can be used to identify an underlying condition — be it an illness, an economic slump, or a family problem. Changing the marker itself may have no effect on the condition. For example, fever often precedes full-blown chickenpox, but while medications to reduce the fever may make the patient feel better, they have no impact on the infection.

1. **Selection bias** is a distortion in a measure of association (such as a risk ratio) due to a sample **selection** that does not accurately reflect the target population. How does random assignment help solve selection bias?

Ans - Selection bias, also known as sampling bias, usually refers to groups (e.g., experimental, control) that are systematically different prior to experimental manipulation or intervention due to the assignment of participants to groups. In other words, variations detected during a study are attributable to group differences due to selection bias or the independent variable (e.g., manipulated variable). Selection bias can occur during participant selection, assignment, and/or during the study. The bias that occurs during participant selection is generally identified as a threat to external validity, whereas bias that occurs during assignment is known as a threat to internal validity. During a study, if a significant number of participants withdraw without completing the study, selection bias can also occur. This entry examines the context in which ...

Randomized experiments solve the problem of selection bias by generating an experimental [control group](https://energypedia.info/wiki/Control_Groups) of people who would have participated in a program but who were randomly denied access to the program or treatment. The random assignment does not remove selection bias but instead balances the bias between the participant and nonparticipant samples. In [quasi-experimental designs](https://energypedia.info/wiki/Quasi-Experimental_or_Non-Experimental_Designs), statistical models (for example, matching, double differences, instrumental variables) approach this by modeling the selection processes to arrive at an unbiased estimate using nonexperimental data. The general idea is to compare program participants and nonparticipants holding selection processes constant. The validity of this model depends on how well the model is specified.[[2]](https://energypedia.info/wiki/Selection_Bias#cite_note-Baker.2C_J._L._.282000.29:_Evaluating_the_Impact_of_Development_Projects_on_Poverty._A_Handbook_for_Practioners._The_World_Bank.2C_Washington.2C_D.C.:_http:.2F.2Fsiteresources.worldbank.org.2FINTISPMA.2FResources.2Fhandbook.pdf_.28p.5.29-1)

1. When you're formulating a business question, try to avoid questions like this: **What was the most consumed piece of marketing collateral during Q3?**How would you reframe the question if your goal was to build a model for future content-marketing plans? Please provide the reframed question and a short explanation of why you chose that new question.