LECTURE 1 Linear Statistical Modelling and Social Science
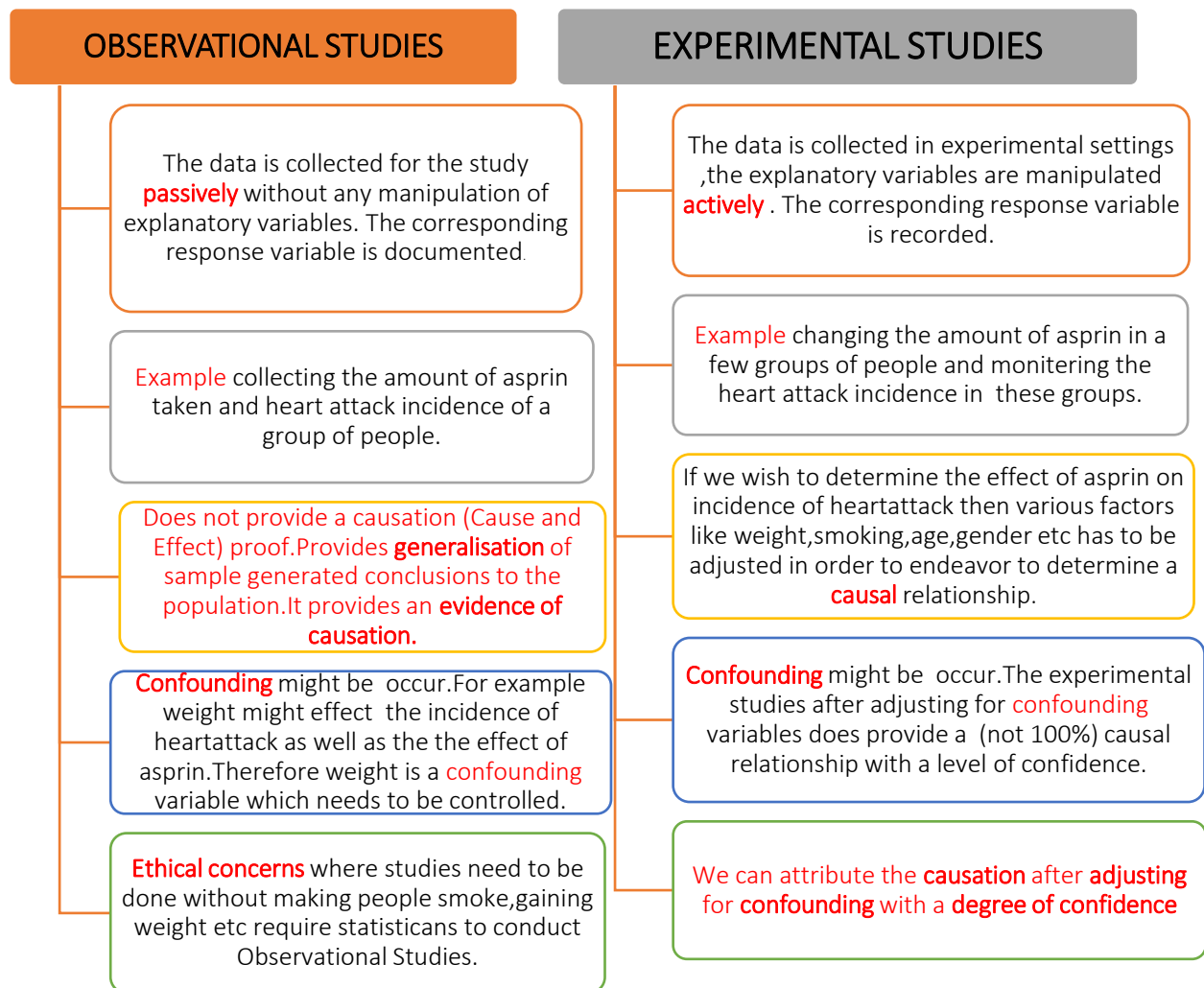
Objectives:

- *To differentiate between social reality and statistical models that represent social realities.*
- *To juxtapose/expound the advantages and disadvantages of observational and experimental studies*
- *To understand the importance of identifying the population of interest, good sampling techniques and collection of a representative sample.*

Statistical Models:

- Are the simplistic descriptive summaries representing the complex social realities perceived in the world.
- Are not the social processes that involve complex social realities. Example a person life trajectory to a particular profession is a process.
- Statistical models are fitted to the data in order to understand the relationships between these models , the social realities and the social theories. The residuals, the parts of response variables that cannot be explained by the explanatory variables help us evaluate these relationship.
- Statistical processes help identify the significant aspects of a process. For example a person might be earning a high income because of a combination of factors like good education, parental influence, school etc.
- Statistical model is substantive if it provides an accurate description of the data. If the relationship between education and income is non-linear then it cannot be depicted as linear.
- Statistical data analysis helps answer vital question in the contemporary social context. Example "Is there income equality amongst males and females "

Observational Studies and Experimental Studies

OBSERVATIONAL STUDIES	EXPERIMENTAL STUDIES

The data is collected for the study **passively** without any manipulation of explanatory variables. The corresponding response variable is documented.

The data is collected in experimental settings ,the explanatory variables are manipulated **actively** . The corresponding response variable is recorded.

Example collecting the amount of asprin taken and heart attack incidence of a group of people.

Example changing the amount of asprin in a few groups of people and monitering the heart attack incidence in these groups.

Does not provide a causation (Cause and Effect) proof.Provides **generalisation** of sample generated conclusions to the population.It provides an **evidence of causation.**

If we wish to determine the effect of asprin on incidence of heartattack then various factors like weight,smoking,age,gender etc has to be adjusted in order to endeavor to determine a **causal** relationship.

Confounding might be occur.For example weight might effect the incidence of heartattack as well as the the effect of asprin.Therefore weight is a confounding variable which needs to be controlled.

Confounding might be occur.The experimental studies after adjusting for confounding variables does provide a (not 100%) causal relationship with a level of confidence.

Ethical concerns where studies need to be done without making people smoke,gaining weight etc require statisticans to conduct Observational Studies.

We can attribute the **causation** after **adjusting for confounding** with a **degree of confidence**

Key Points

- To obtain credible research associations or causations for observational and experimental studies it is important to statistically control for all the potential confounding variables. Practically sometimes this not possible because confounders might not be identifiable.
- If a variable is not affecting the response then there is no need to control for it.
- Experimental Studies can be ambiguous. For example the recovery due to a drug might not necessarily be solely due to the active ingredient but due to the positivity of the patient.
- Randomised controlled double blind experiments are the best experiments since they eliminate the chance of subjectivity. This entails random selection and random assignment.

- In observational studies there might be associations between three or more variables whose relationship might be effected by their effect on each other. The relationship between explanatory variable and response variable can be ascertained by controlling for an intervening variable.
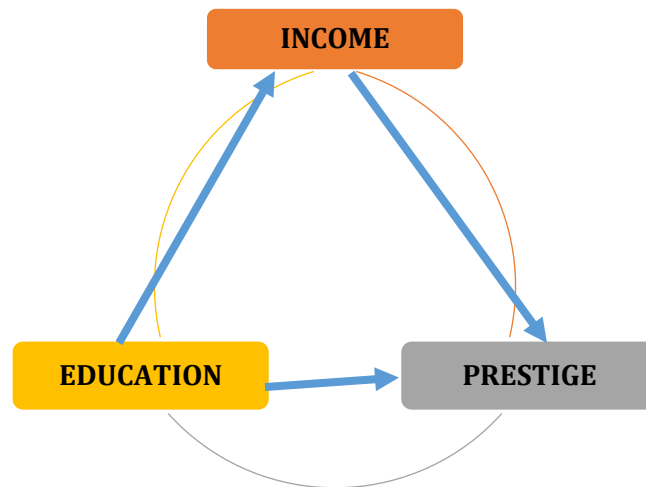
- Causal model above shows that the education has effects on the income and the prestige of a person. Income in turn affects the prestige but the degree of this association is decreased (when education is controlled) because both prestige and income are mutually dependent on Education. The association between income and prestige is not causal (spurious).
- The relationship between education and prestige is intervened by the income and this relationship can be ascertained by controlling for income variable.
- Crossectional Observational studies are less reliable to provide causal interpretation as compared to Longitudinal studies though sometimes longitudinal studies might be practically uneconomical or not feasible. This can be due to uncontrolled confounding variables.
- Experimental studies can attribute causation more reliably as the explanatory variable can be manipulated in practice or principle. Sometimes the variable cannot be manipulated in principle or practice. Example Gender Job and Income
- **Causation is affected by the confounding variables whether it is observational studies or experimental studies**
- **Non Manipulative explanatory variables** like age, gender etc should also be factored in the study to make it non restrictive.
- To conduct statistical studies and to identify cogent statistical inferences it is important to identify the population of interest, use the technique of randomization and good sampling design to collect a representative sample from the population of interest.

****