Confounding

Srikar babu Pilli

August 2022

The concept of confounding is a very important one.Let us say,you are very ill and someone has given you an allopathic medicine saying that it is a very good medicine and some other one has given you a homeopathic medicine saying that it is a very good medicine,now we want to know which one is a good medicine.

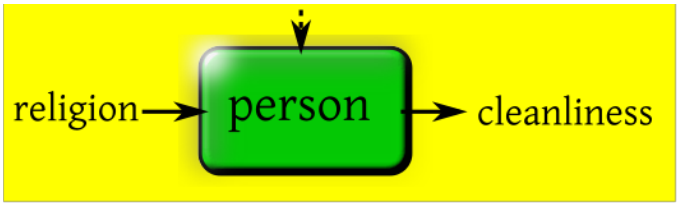
So what you do is that you swallow both the medicines and say,after three days you get fine.Now from this,we can't really say that the allopathic medicine cured you.we can't really say that the homeopathic medicine cured you either.It is quite possible that the two chemicals mixed together cured you.It is even possible that the allopathic medicine cured you and the homeopathic medicine has no effect and similarly it is possible that the homeopathic medicine cured you and the allopathic medicine has no effect.So,it is quite possible that you would possibly have been cured by only one medicine.It is even possible that you have been cured by no medicine at all and these two medicines kept you ill longer than you should have been.

In short,we can't make any decision regarding these two medicines simply because you have taken them together.What you should have done is to use these medicines on two different persons of identical nature,only then you can say that the person who got the allopathic medicine got well in 3 days and the guy who got the homeopathic medicine got well in 2 days etc. and now we can compare.This is the idea of confounding.

Confounding means that if we change two things simultaneously,then any change in output can't be clearly attributed to this input or that input,we have to try to change the inputs one at a time,if we do not do that then the two inputs are said to be confounded.This is a concept which is often represented in various disguised forms.It becomes tricky especially when we are confounding between a treatment and a factor and a control factor or between two control factors i.e./ factors which are beyond our control.

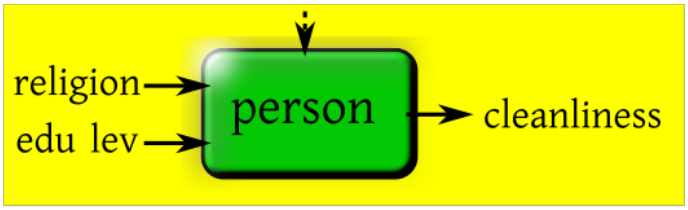
Let us take an example,many people say that,in many cases Muslims are uncleanly people,if we go to any place that is infested by Muslims,they are dirty places and that will be indeed borne out by actual fact.So it seems to bear

out this allegation that somehow this religion is connected with cleanliness.so roughly speaking,we have a black box in our mind where we have a person as the unit and we have got religion as the input and we have cleanliness as the output.

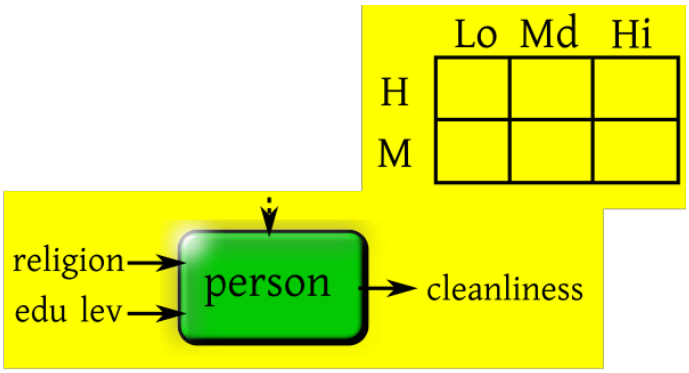


This is just a non-mathematical example and we are not going to go into how we measure cleanliness as a continuous variable.suppose,religion has got two levels namely Hindus and Muslims and if we just go about,say various parts of India,we will really see that Muslim infested parts are very uncleanly and the Hindu infested parts are generally cleaner than that.However,if we look carefully,we will observe that there is yet another factor that is playing a role and that is the educational level.

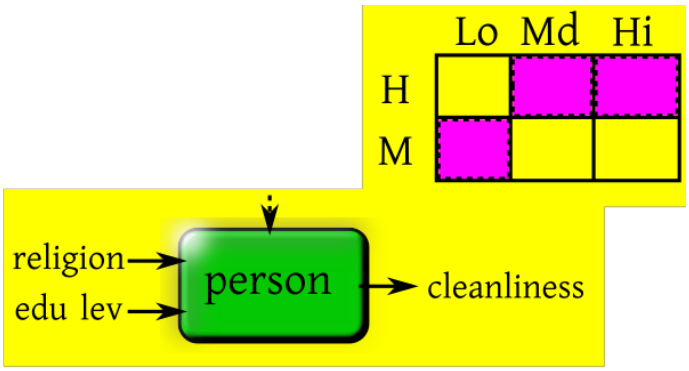
Now,most of the Muslims also happen to be of the low education group class because the Muslims are generally poor people,they can't afford good education,as a result,when we are looking at a general Muslim,most of them are with low education level.So,it is quite possible that the uncleanness that we see is because of low educational level whereas when we look at a bulk population of Hindus,there are less education people as well as higher education people.so,the difference that we see has nothing to do possibly with religion,it can be explained using the difference in educational levels.



In this case,these two factors are basically confounded.When we just do an observational study,take social study,we observe whatever we get to observe,in that case,we will see that whenever religion becomes from Hindu to Muslim,the educational level also generally falls.So,it is basically like say,we take three educational levels,we call them low,medium and high and we have got two levels in religion namely Hindus and Muslims,so we have got 6 cells.



So,we can draw a meaningful conclusion only when we know all these 6 cells are nicely represented.But if we just observe whatever we can observe,in that case,we will see that typically only the purple boxes are represented in our sample and so we have only low education level Muslims and medium level and high level Hindus,so the bulk comes from this groups.So we see the moment the religion changes so does the education.so,we have really not looked at all the cells,so we are basically comparing a low education Muslim with a high education Hindu and trying to say whatever difference we see is because of the religion,that will not be fair.So.from this,we can't really conclude anything.



If our data set is like this then we have a confounding between these two factors.The point here is not that we can conclude either for Hindu against Hindu or for Muslim against Muslim. The linear algebra and the R will remain good but the conclusion will not be meaningful.So,we have to make sure that all these cells are properly represented and we are changing one factor at a time.If we have some observation corresponding to Hindu and low education group,then we can compare the cleanliness of low educated Hindus and Muslims to get an idea about the difference between Hindus and Muslims,but here there are no less educated Hindus,so we are comparing more educated Hindus with less educated Muslims which means we are basically changing both the religion as well as the education level and we don't know that the change that we see is because of the

education level or the religion or because of the simultaneous presence of both the changes,and we have no way of knowing it.That is called confounding and it is a good idea to try to avoid it.