SR HW 6

Emilio Horner

2022-10-13

6e1.

multicollinearity, post-treatment bias, and collider bias

6e2.

Collider bias happens when you control for the third variable, this creates a statistical but not necessarily causal association among causes. An example could be money being causal with age. If one controls for jobs if could be assumed aging causes increase in wealth when people its not that just getting older makes one richer but instead getting a job that happens when one gets older makes one ricer.

6e3.

Fork:

For a fork the variable Z is a common cause of the other two variables and this is what is generating the correlation

Pipes: the pipe is the X -> Z -> Y line and basically if you condition for Z you would block the path of understanding X to Y

Colliders:

In the collider it is the act of conidtioning on Z that porduces the causal influence. X and Y do not really have causal influence

Descendants

For descendants, a variable is infuenced (or descends) frm another variable conditioning on that variable will partially conidtion on the desdending variable. I feel like this is similar to the collider except with a previous variable impacting the collider variable

6e4.

How is a biased sample like conditioning on a collider? Think of the example at the open of the chapter.

The collider bias is effectively a biased sample. The problem comes when trying to generalize the oberservation in the data to the relationship of the variables. For example, if there's doctors at a Mountain Goat concert that only went because they have money and can go, and poor people who are huge mountain goats fans one could mistakenly think poor people are bigger fans and doctors don't really like music. This might be true at the particular concert but is inaccurate as a generalized statement.

6m1.

C ← V → Y

the paths now from X to Y are

X -> Y X <- U -> A <-C ->Y X <-U -> A <-C <-V ->Y X <- U -> B <-C -> Y X <-U ->B <-C <-V -> Y

A is the variable that should be conditioned on

6m2.

```
Data <- 4
```

```
X <- Data |>
  rnorm(mean =0, sd =1)
```

```
Z <- Data |>
  rnorm(mean =0, sd =1)
```

```
Y <- Data |>
  rnorm (mean = Z, sd =1)
```

```
cor(X,Z)
```

```
## [1] 0.8722359
```

the variables here are correlated since they’re the same basically

6m3.

for the upper left one, z is a fork on x -> y so its noncausal and X <- Z <-A -> Y is some sort of weird pipe thing that should be closed

for the upper right one, i think its fine

for the lower left one, Z is a collider so you need to conidtion on Z

for the lower right one, A is a fork on X and Z so should be conidtioned on.