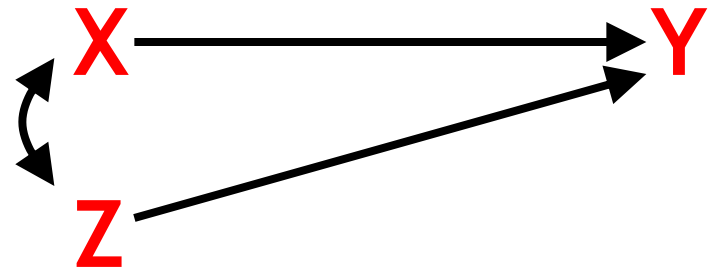
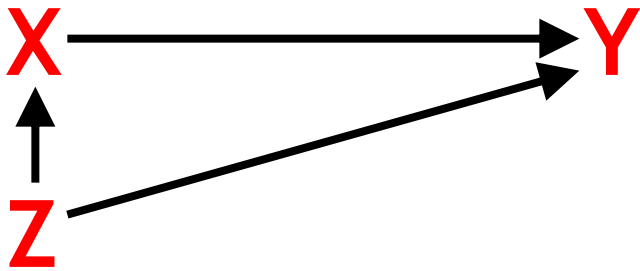


SOC 3811/5811:  
BASIC SOCIAL STATISTICS

Two-Way ANOVA and Three-Variable Relationships

# Three Variable Relationships

In assessing the effect of X on Y, we have primarily thought of Z as a **confounder**.



Does this mean that we should just automatically control for Z to assess the causal effect of X on Y?

# Three Variable Relationships

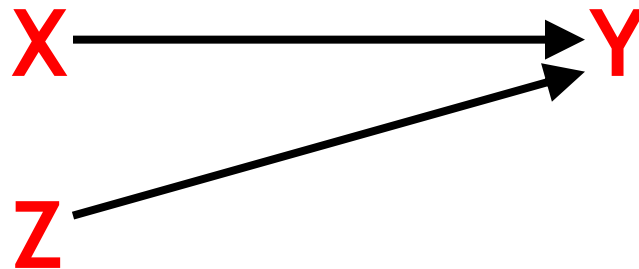
**No.** There are variety of ways that Z can play into the relationship between X and Y. Sometimes statistically controlling for Z is a great idea, sometimes it makes no difference, and sometimes it is a terrible idea.

**Theory** and **prior evidence** should guide our decision about how Z plays into the relationship between X and Y

**Depending on our theoretical understanding of how Z plays into the relationship between X and Y, we might or might not want to statistically control for Z**

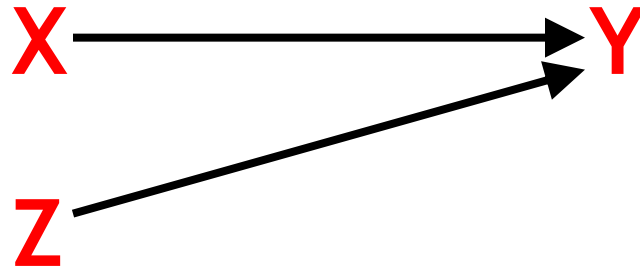
# Three Variable Relationships

Under the scenario depicted below, the association between X and Y is unaffected by the presence of third variable(s) Z



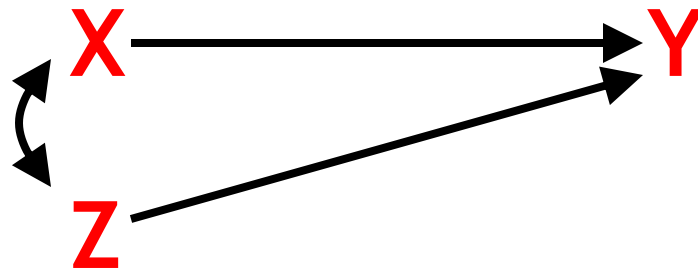
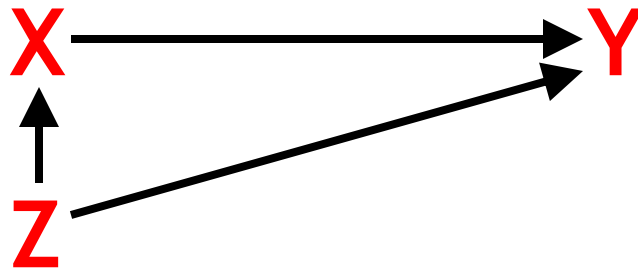
# Three Variable Relationships

Statistically controlling for Z has no bearing on our assessment of the association between X and Y ... so there is no need to do so (but it doesn't hurt anything)



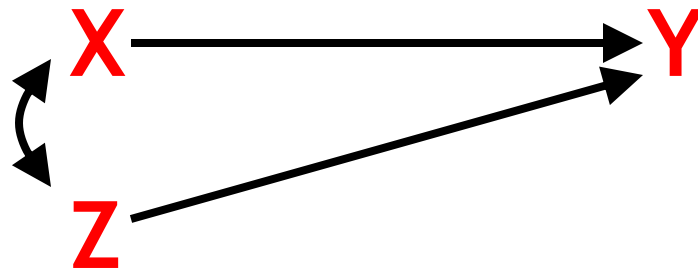
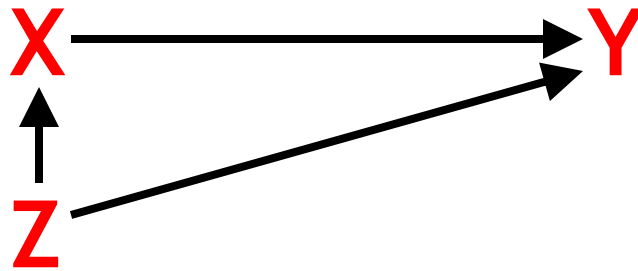
# Three Variable Relationships

Under both scenarios depicted below, the association between X and Y is—at least partly—spurious owing to the influence of **confounding variable(s) Z**



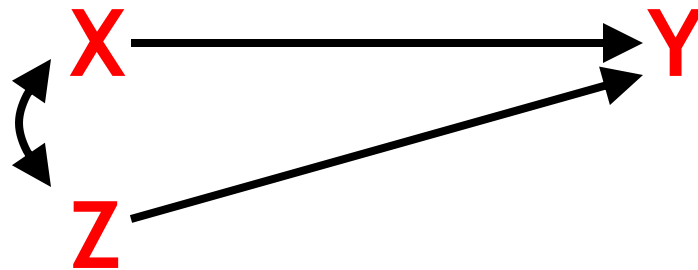
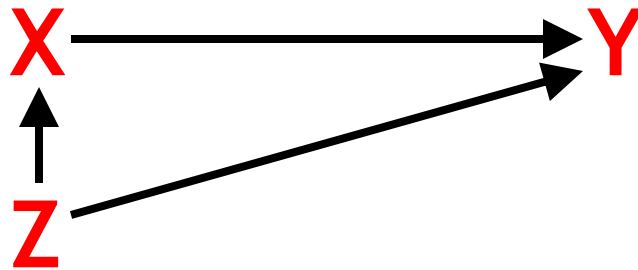
# Three Variable Relationships

Under each of these scenarios, statistically controlling for Z is designed to estimate the *independent* association between X and Y ... the association “net of” Z



# Three Variable Relationships

Here, the independent association between X and Y ... the association “net of” Z ... represents the **direct effect** of X on Y (but don't forget about the other criteria for establishing causality)

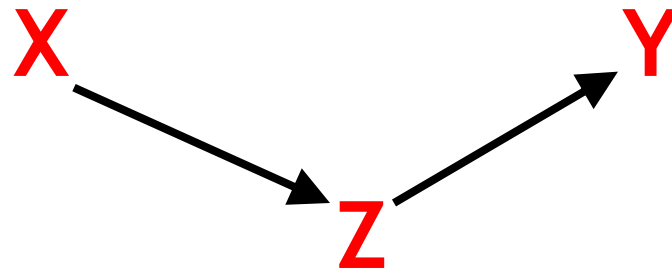
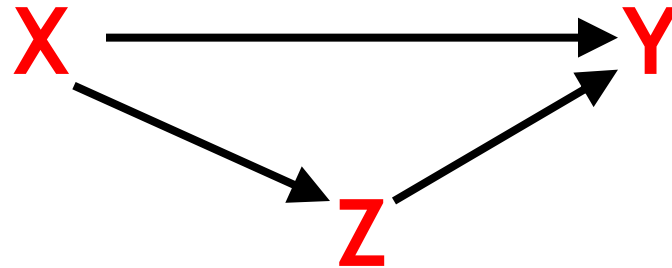




# Three Variable Relationships

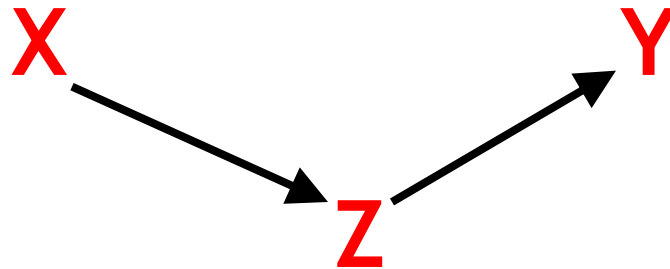
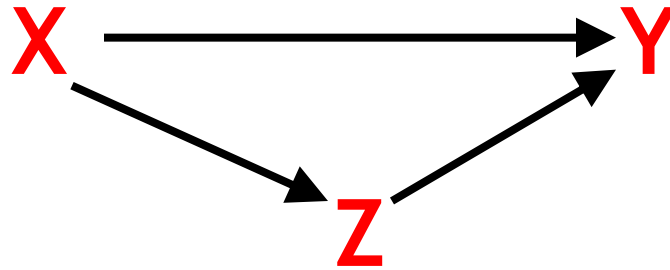
Under the scenarios depicted below, Z is associated with both X and Y but does **not** induce spuriousness

Z is a **mechanism** or **pathway** through which X affects Y ... Z is known as a **mediating variable**



# Three Variable Relationships

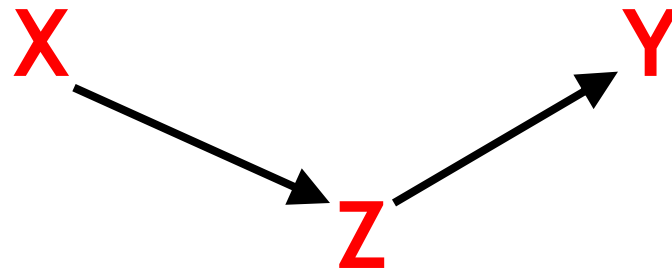
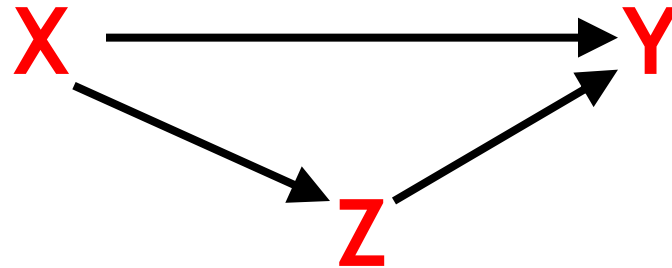
Statistically controlling for Z under these scenarios is also designed to estimate the independent association between X and Y ... the association “net of” Z



# Three Variable Relationships

Conceptually, in these cases the independent association between X and Y ... the association “net of” Z ... represents the **direct effect** of X on Y

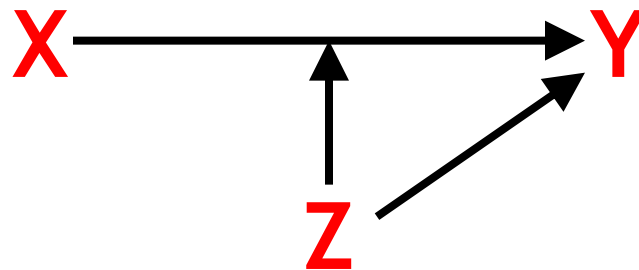
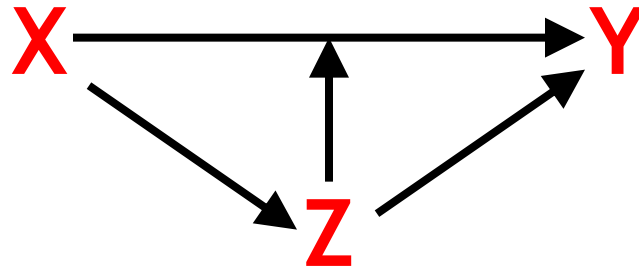
Is that really what we want to estimate? **It depends...**



# Three Variable Relationships

In some cases the association between X and Y may be different across levels of Z

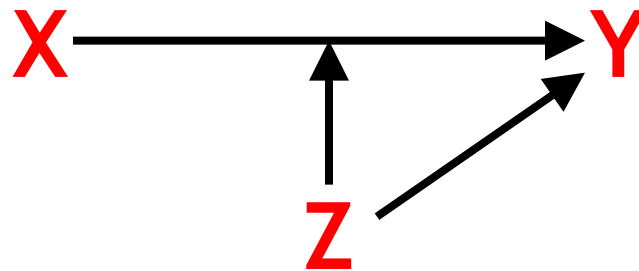
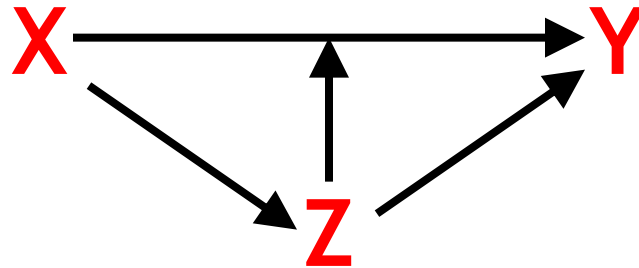
In these cases we say that there is an **interaction** between X and Z ... Z is known as a **moderating** variable



# Three Variable Relationships

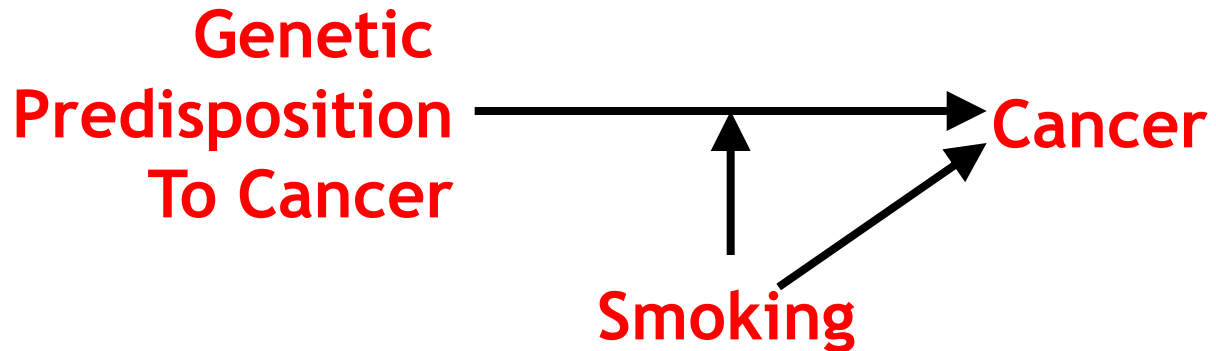
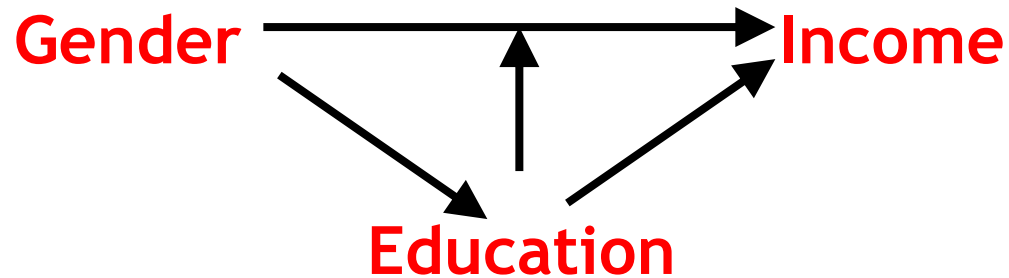
If we simply statistically adjust for Z in these situations, then the resulting “net” association between X and Y isn’t very meaningful

We must allow for different associations by level of Z



# Three Variable Relationships

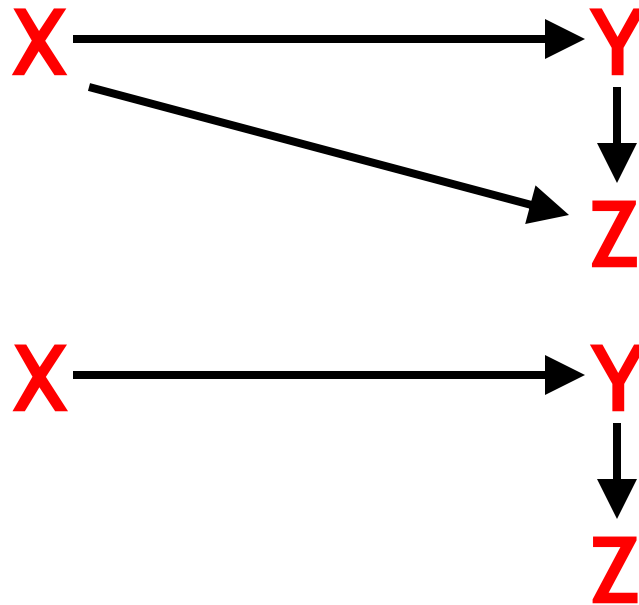
Example



# Three Variable Relationships

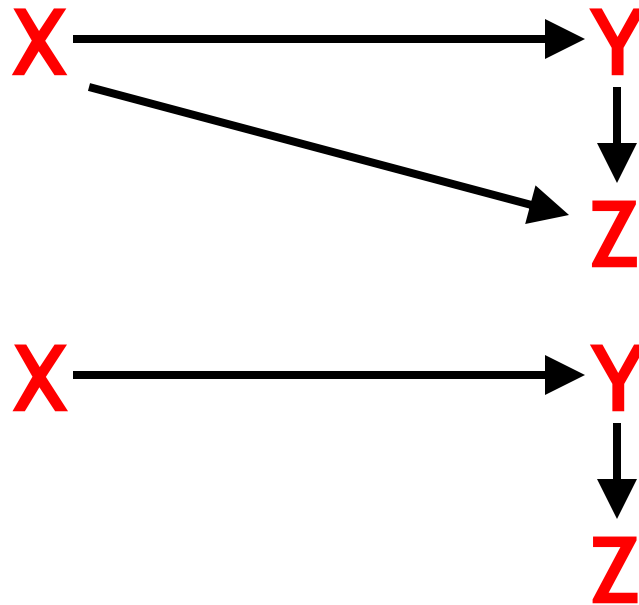
Under the scenarios depicted below, Z is affected by Y (and also possibly X)

Here, Y is a mechanism through which X affects Z



# Three Variable Relationships

If we are trying to assess the independent association between X and Y, then in these situations statistically controlling for Z is a **terrible idea**





# Three Variable Relationships

In assessing the causal effect of X on Y, is Z...

...totally irrelevant to the association between X and Y?

...a **confounding** variable which must be controlled?

...a **mediating** variable which we might want to control?

...a **moderating** variable which we must treat with care?

...a variable that it would be a mistake to control for?

It depends on the situation ... so carry out analyses based on **theory** and on evidence from **prior research**

How much does the number of hours you spend per week on this class (**X**) affect your grade (**Y**)?

1. If we simply looked at the association between X and Y, would the result be an accurate estimate of the causal effect?
2. Theoretically speaking, what role might each of the following variables (**Z**) play in the association between X and Y? Is Z a **confounder**? A **mediator**? A **moderator**? Totally **irrelevant**?
3. Based on that answer, what would it mean to “control for Z”?

**Z** = Like math?

**Z** = Had a stats class before?

**Z** = Completed all assignments

**X** = Hours Spent  
Per Week on Class



**Y** = Expected  
Grade



**Z** = Like Math

**Z** = Had a Stats  
Class Before

**Z** = Completed All  
Assignments

# Worksheet

According to today's class data: How much does the number of hours you spend per week on this class (**X**) affect your grade (**Y**)?