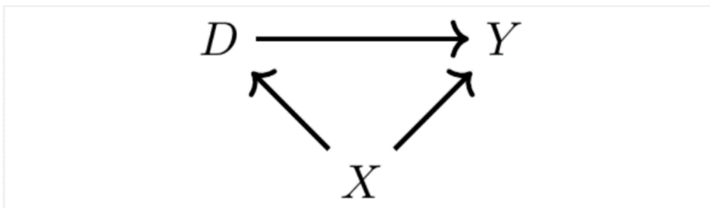
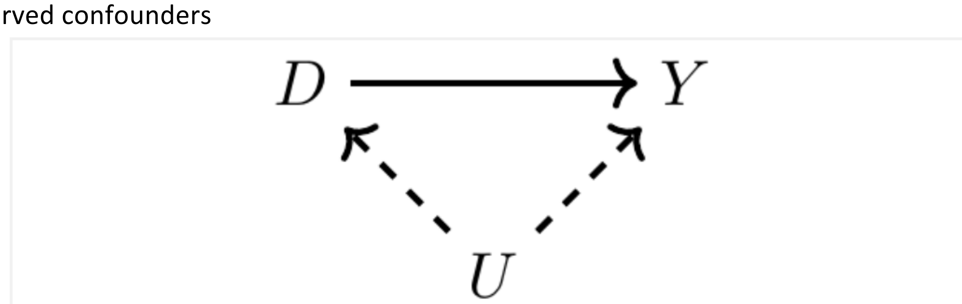
Week 3 “1-pager”

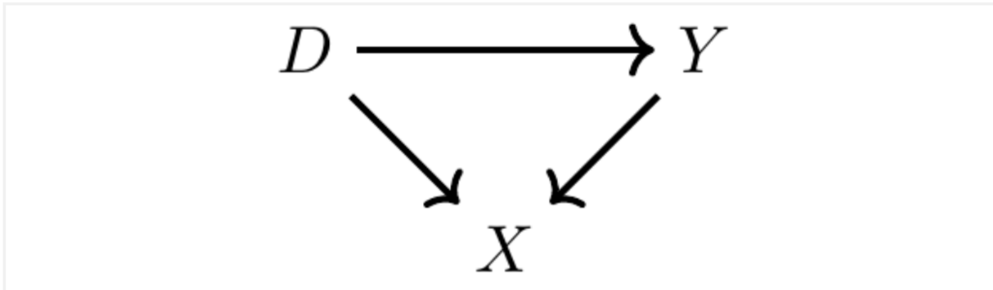
Directed Acyclic Graphs

1) Directed Acyclic Graphs (DAGs) are graphical forms of chains of causal effects
    a. DAGs shows all causal relationship that are relevant to the effect of D on Y
    b. Developed by theory, literature, or deep institutional knowledge
2) Causal effects occur in two ways:
    a. Direct Pathway: $D \rightarrow Y$
    b. Indirect (mediated by a third variable): $X \rightarrow D \rightarrow Y, D \leftarrow X \rightarrow Y, etc.$
    c. Assumptions are not displayed as causal relationship in the DAG
3) Simple DAG
    a. 

    b. We have 3 random variables: X, D, and Y
        i. D affects Y
        ii. X affects Y and D
    c. There are two paths affecting Y:
        i. Direct Pathway: $D \rightarrow Y$
        ii. Backdoor Pathway: $D \leftarrow X \rightarrow Y$
        iii. The direct path is causal while the backdoor pathway is not causal but a process that creates a spurious correlation between D and Y that is driven by fluctuations in X
4) Backdoor Criterion
    a. Backdoor pathways matter because they create systematic noncausal correlations between the causal variables of interest and the outcome of interest
    b. Our Goal: Close all of the backdoor pathways and identify the direct pathway of D on Y using an identification strategy or credible research design
        i. Conditioning or controlling for the confounder or the appearance of a collider closes a backdoor pathway
5) Confounders
    a. In the Simple DAG, X is an observed confounder
    b. When X mediates the values D and Y, X is considered a ***<u>confounder</u>***
    c. Unobserved confounders
        i. 

ii. Since U is unobserved, there remains an open backdoor pathway from D to Y

a.

6) Colliders
   a. Colliders differ from confounders
   b. 

   c. There are two pathways
      i. Direct Pathway: $D \rightarrow Y$
      ii. Backdoor Pathway: $D \rightarrow X \leftarrow Y$
   d. When two variables cause a third variable along some path, it is a collider
   e. D and Y collide at X and fluctuations in D and Y affect X
   f. Colliders mean that the backdoor pathway is closed; when left alone the backdoor pathway is closed in the presence of a collider

Regression Review

1) The expected value of beta-hat is beta if OLS is unbiased if assumptions hold
   a. $E(\hat{\beta}) = \beta$
   b. First Assumption: Linear in Parameters
      i. In the population model, the dependent variable y is related to the independent variable and the error term u: $y = \beta_0 + \beta_1 x + u$
   c. Second Assumption: Random Sampling
      i. We have a random sample of size n with a set of number $\{(x_i, y_i): i = 1, 2, \dots, n\}$ following the population model $y = \beta_0 + \beta_1 x + u$
      ii. Where i is a random sampling draw from the population
      iii. Such that $y_i = \beta_0 + \beta_1 x_i + u_i, i = 1, 2, \dots, n$
   d. Third Assumption: Sample Variation in Explanatory Variables
      i. The sample outcomes on x, $\{x, i = 1, 2, \dots, n\}$ are not the same value
      ii. If there is no variation in x then this assumption fails
   e. Fourth Assumption: Zero Conditional Mean
      i. The most crucial assumption in causal inference and the one most likely to fail
      ii. $E(u|x) = 0$
      iii. We can still calculate beta-hat even if the fourth assumption fails, but it means that beta-hat will not be an unbiased estimator
2) Bias of Expected Value of OLS
   a. If there is an omission of a confounding variable or the model is underspecified we will have omitted variable bias: When $E[U|X] \neq 0$
      i. $\hat{\beta}_1 = \beta_1 + \beta_2 \frac{C(U,X)}{V(X)} = \beta_1 + \frac{C(Y,U)}{V(U)} \frac{C(U,X)}{V(X)}$

b. The direction of the bias depends on the direction of the covariance among Y, X, and U

Direction of Bias	$C(X,U) > 0$	$C(X,U) < 0$
$C(Y,U) > 0$	Upward Bias	Downward Bias
$C(Y,U) < 0$	Downward Bias	Upward Bias

3) Robust Standard Errors
   a. Whenever we assume homoskedasticity
      i. $V(u|x) = E(u^2|x) = \sigma^2 = E(u^2)$
   b. When the fifth assumption is violated, the variance of the estimator beta is biased and no longer has the minimum mean squared errors
      i. We need robust standard errors $V(\hat{\beta}_1) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 \hat{u}_i^2}{SST_x^2}$
4) Clustered Robust Standard Errors
   a. In addition to heteroskedastic errors, we need to be considered if errors are correlated at an aggregate level even if they don't affect individuals directly
   b. When use the cluster option in the regression by the cluster_id, the standard errors in the clustered data are similar to their distribution in the unclustered data set
   c.