Covariates HW

Adon Rosen

1. Assume that you have the confounder Z and that you are interested in the effect of X on Y

a. Given a fully randomized trial the relationship between the predictor variable (X) and the confounder variable (Z) is independent. This independence is attained by removing the confounder variables ability to influence the possible levels of X. Thus an ANOVA approach is appropriate for this design. Assuming X is a factor, Z is a continuous variable, and Y is a continuous variable.

b. When participants are randomly assigned X and Z become independent of one another, this occurs because all levels of Z are equally likely to be placed into the various levels of X. Therefore, this relationship is controlled for by controlling for the possible distribution of Z that is observed in X.

c. If the randomization is not possible then an ANCOVA approach is more appropriate. The ANOVA model would not be able to protect against the possible influence of the continuous confounder, for instance if school (X) is being used to predict performance on a test (Y), the canonical ANOVA has no way to protect against the bias of a continuous confounder such as SES (Z); thus, an ANCOVA approach is more appropriate for this instance.

2. In this situation assume that the treatment effects the covariate

a. Assuming the model is: Y ~ X + Z + e ; then this model is testing the effect of X (school) on Y (performance) when holding Z (SES) constant; however, due to possible collinearity issues it may make it difficult to asses the directional of these variables, although this would yield the direct effect of X on Y.

b. Assuming the model is: Y ~ X + e; then this model is testing the relationship between X (School) and Y (performance) without controlling for any confounding variables. This would yield the total effect of X on Y.

3. In the following situation,we are dealing with a variable that is affected both by the outcome and the treatment.

a. If trying to asses the relationship between X and Y using a randomized trial an ANOVA approach is appropriate as the relationship between X and Y is not influenced by Z. If the design is not randomized

the presence of Z does not need to be accounted for and the model can be trained using an ANOVA technique.

b. Assuming the model is: Y ~ X + Z + e; this technique would not be appropriate as it assumes that Y is caused by Z; wherein the relationship is actually Y is causing Z. While the estimation technique may be the Best Linear Unbiased Estimator, the model would be biased. This is because the model does not speak to the theory, and a collider variable is being used as the criterion variable