

Assignment 1

Andaleeb Hassan
AST 432

August 27, 2025

All confounders should be controlled, but not all controlled variables are confounders.

Explanation:

Confounders are variables that are statistically associated with both the independent variable (exposure of interest) and the dependent variable (outcome of interest) and does not come in the causal pathway. If not controlled, they can bias the estimated effect and distort the true association between the variables.

Example : In studying the effect of *education* (X) on *income* (Y), *ability* (Z) might be a confounder. Because smarter people tend to get more education and smarter people tend to earn more. Thus, *ability* (Z) might statistically be associated with both *education* (X) (exposure of interest) and *income* (Y) (outcome of interest).

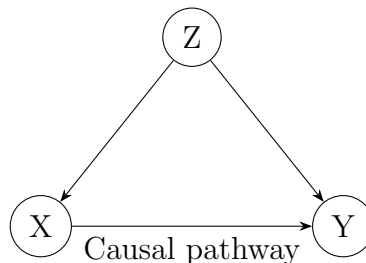


Figure 1: schematic of Confounding variable

Controlled variables are any variables that researchers decides to adjust for in their analysis to reduce bias or increase precision. Not all of them need to be confounders. Some are included to reduce variance or because of study design, even if they're not technically confounding the main relationship.

A control variable is any variable that is included in a regression model to hold its effect constant

Example: In a study to understand the salary structure of college faculty members, sex could be a potential confounder or effect modifier. To understand it's potential effects, we consider three regression models.

$$\text{Model 1: } Income = \beta_0 + \beta_1 \text{ year} + \varepsilon$$

$$\text{Model 2: } Income = \beta'_0 + \beta'_1 \text{ year} + \beta'_2 \text{ sex} + \varepsilon$$

$$\text{Model 3: } Income = \beta''_0 + \beta''_1 \text{ year} + \beta''_2 \text{ sex} + \beta''_3 \text{ year sex} + \varepsilon$$

We can demonstrate this idea using the (Data file :*salary*) which contains salary of faculty in a small Midwestern college in the early 1980s.. {Fox, John. alr4: Data to Accompany Applied Linear Regression (4th Edition). R package version 0.5.1, 2023. <https://CRAN.R-project.org/package=alr4>}.

Variables	Model 1	Model 2	Model 3
Intercept	18166.1***	18065.4***	18222.6***
year	752.8 ***	759.0***	741.0 ***
sex		201.5	-570.85
year:sex			169.1
R-squared	0.50	0.50	0.5

Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 1: Comparison of Three Regression Models

.scatterplot of the models

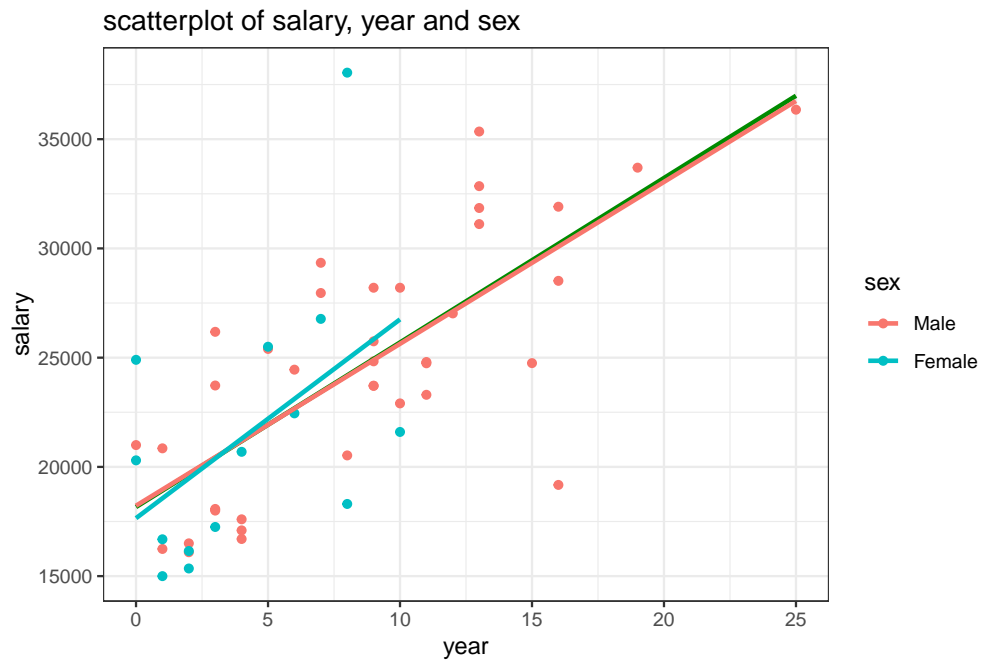


Figure 2: Regression **model 1** and **model 3**

The plot visually displays our findings in table 1, as both the sex regression coefficient and interaction coefficient are insignificant but we can still control for it.