

Omitted Variable Bias Calculation

Full Model: $TS \sim B0 + B1(CS) + B2(SES) + e1$

Naïve Model: $TS \sim b0 + b1(CS) + e2$

Policy variable: Classroom Size

Omitted variable: Socio-Economic Status

```
URL <- "https://raw.githubusercontent.com/DS4PS/cpp-523-fall-2019/master/labs/class-size-seed-1234.csv"
dat <- read.csv( URL )

m1 <- lm( test ~ csize, data=dat )
m2 <- lm( test ~ csize + tqual, data=dat )
m3 <- lm( test ~ tqual + ses, data=dat )
m4 <- lm( test ~ csize + ses, data=dat )
m5 <- lm( test ~ csize + tqual + ses, data=dat )

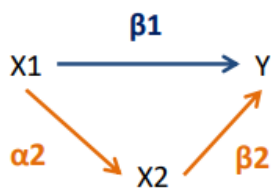
m.auxiliary <- lm( ses ~ csize, data=dat )
summary( m.auxiliary )

# lm(formula = ses ~ csize, data = dat)
# Coefficients:
# -----
#               Estimate Std. Error t value Pr(>|t|)
# (Intercept)  4.469458    0.009033   494.8   <2e-16 ***
# csize       -0.094876    0.000326  -291.0   <2e-16 ***
# -----

# b1 = B1 + bias
# b1 - B1 = bias
b1 <- -4.22
B1 <- -2.67
b1 - B1

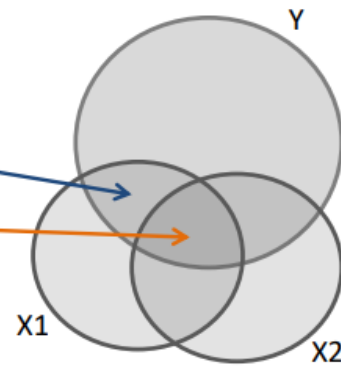
# bias = a1*B2
a1 <- -0.0949
B2 <- 16.34
a1*B2
```

	Dependent Variable: Test Scores				
	Model 1	Model 2	Model 3	Model 4	Model 5
	(1)	(2)	(3)	(4)	(5)
Classroom Size	-4.22 ^{***} (0.18)	-3.91 ^{***} (0.03)		-2.67 (1.63)	-2.22 ^{***} (0.23)
Teacher Quality		55.01 ^{***} (0.25)	55.03 ^{***} (0.26)		55.01 ^{***} (0.25)
Socio-Economic Status			40.94 ^{***} (0.27)	16.34 (17.10)	17.77 ^{***} (2.40)
Intercept	738.34 ^{***} (4.88)	456.70 ^{***} (1.48)	272.91 ^{***} (1.39)	665.29 ^{***} (76.57)	377.26 ^{***} (10.82)



Direct Effect

Indirect Effect



$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon_1 \quad (\text{full regression})$$

$$X_2 = \alpha_0 + \alpha_1 X_1 + \varepsilon_2 \quad (\text{auxiliary regression})$$

$$\text{bias} = \beta_2 \alpha_1 \quad (\text{path diagram for } X_1 \rightarrow X_2 \rightarrow Y)$$