

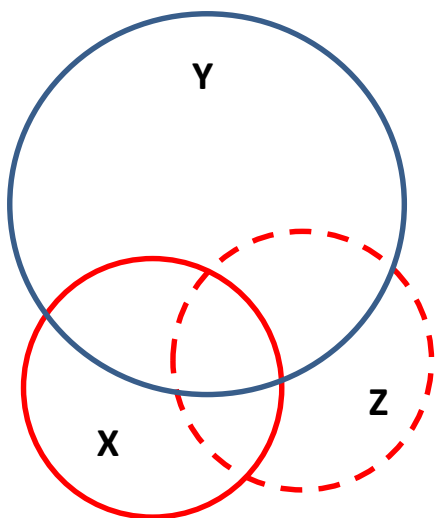


SEVEN SINS OF REGRESSION

Fundamentals of
PROGRAM EVALUATION

JESSE LECY

(1) OMITTED VARIABLE BIAS



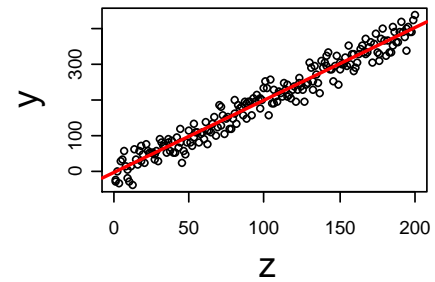
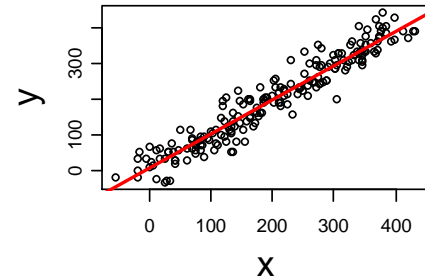
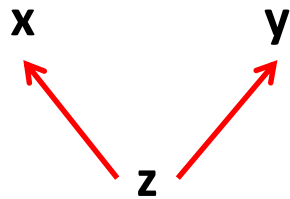
$$Y = b_0 + b_1X + e$$

$$Y = \beta_0 + \beta_1X + \beta_2Z + \varepsilon$$

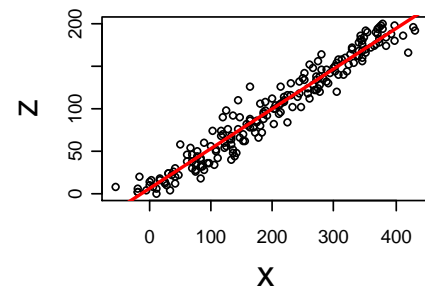
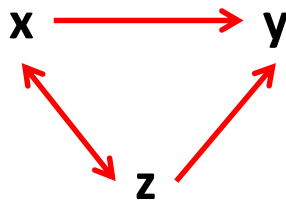
$$b_1 \neq \beta_1$$

When it matters:

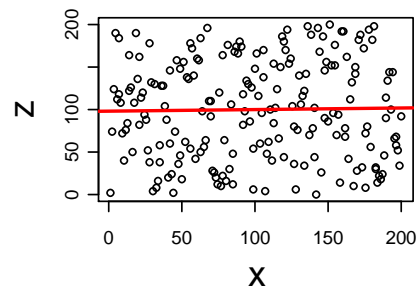
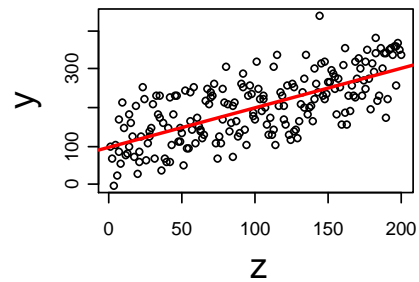
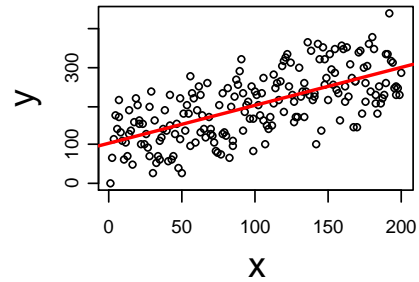
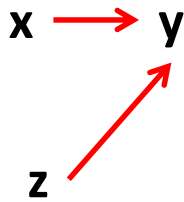
Spurious
Correlations



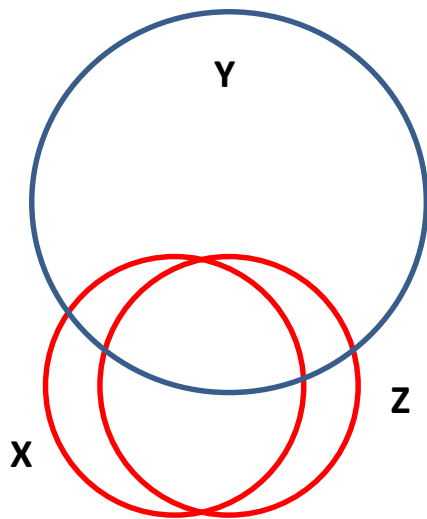
Indirect
Effects



When it doesn't matter:



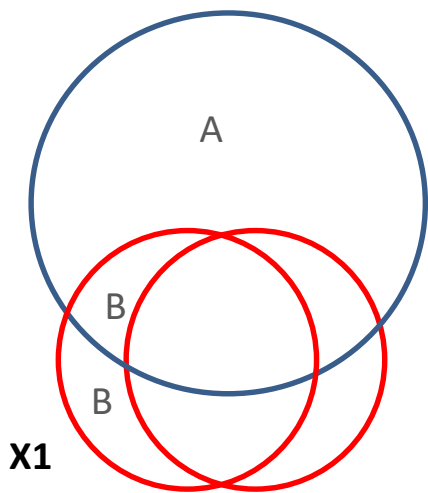
(2) MULTICOLLINEARITY



Cannot tell the independent effects of either variable.

The standard errors of each slope will be inflated.

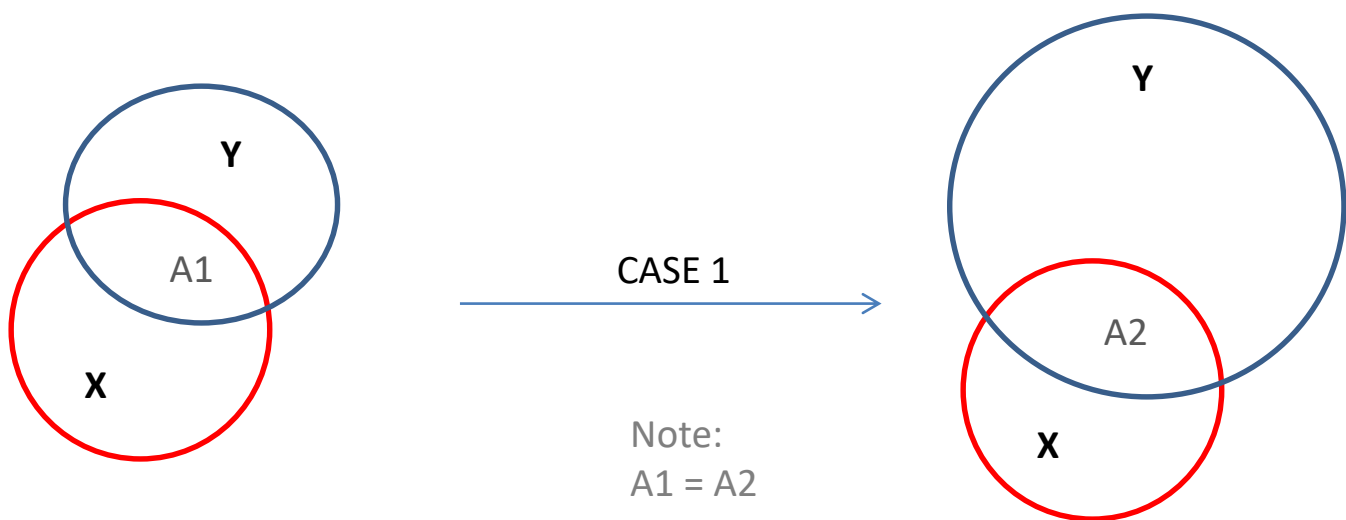
(2) MULTICOLLINEARITY



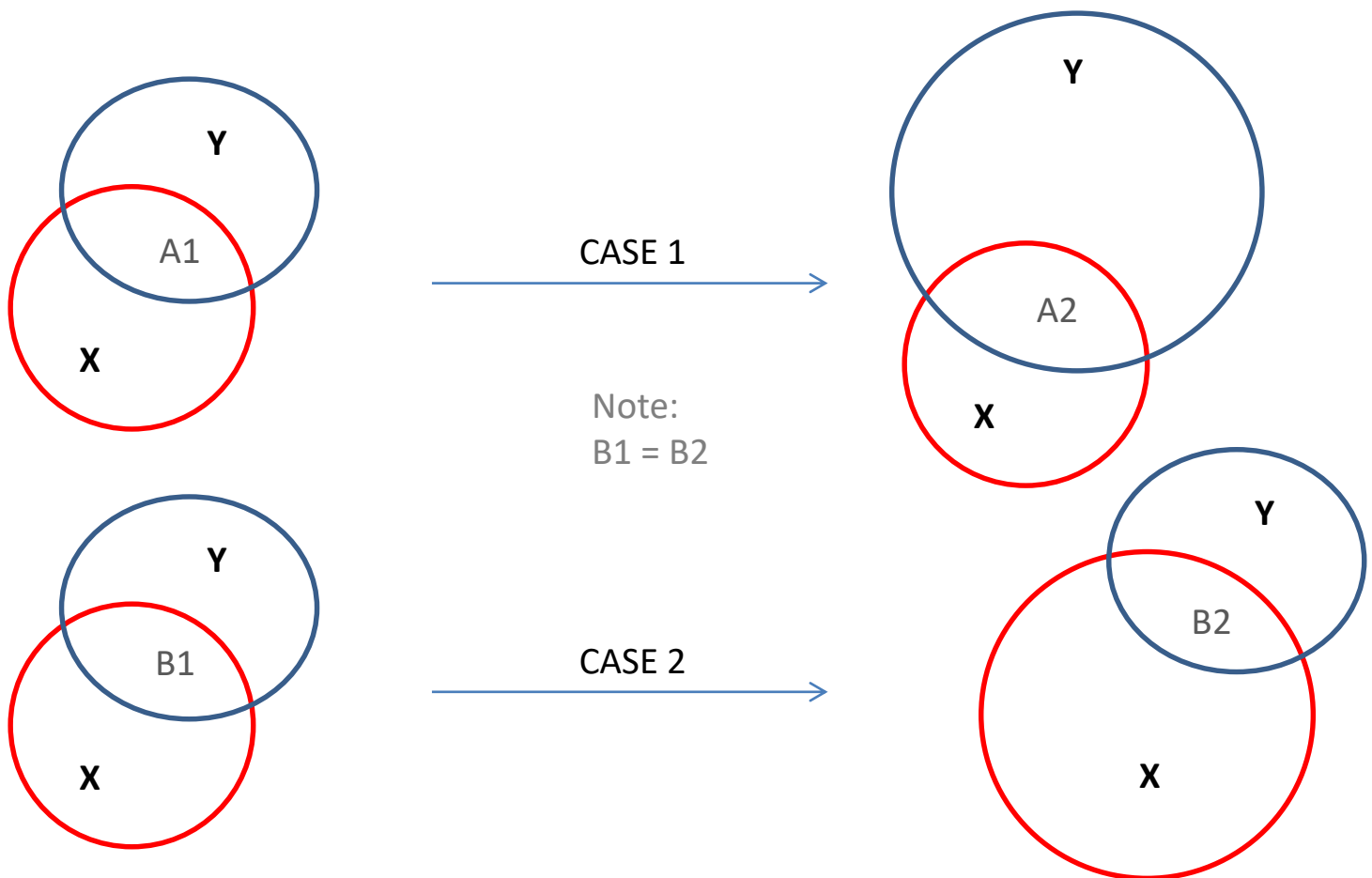
$$SE_{b_1} \sim \frac{A}{B}$$

The higher the multicollinearity, the smaller B will be, which means larger the standard errors. When standard errors are large the confidence intervals are bigger and it is less likely that the slope will be statistically significant.

(3) MEASUREMENT BIAS



(3) MEASUREMENT BIAS

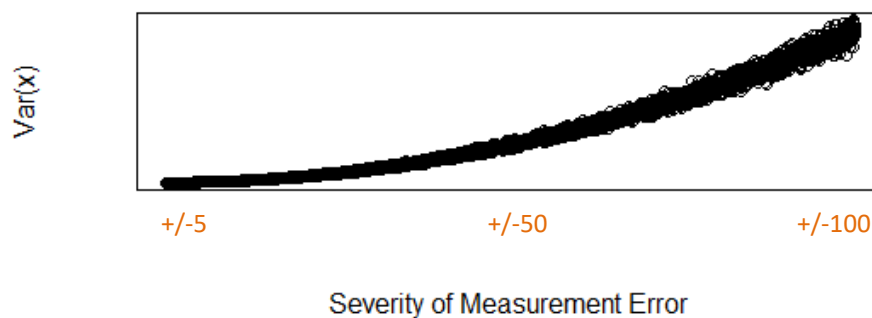


(3) MEASUREMENT BIAS

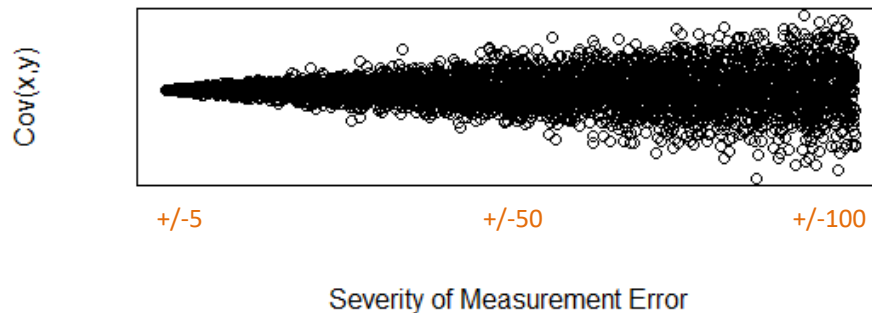
- Measurement Error is random error.
- Measurement error in the dependent variable only affects the standard errors.
- Measurement error in the independent variables causes attenuation.
- Attenuation always pushes the slope towards zero, no matter if the relationship is positive or negative.

RANDOM MEASUREMENT ERROR EXAMPLE

Variance of X as a Function of Measurement Error



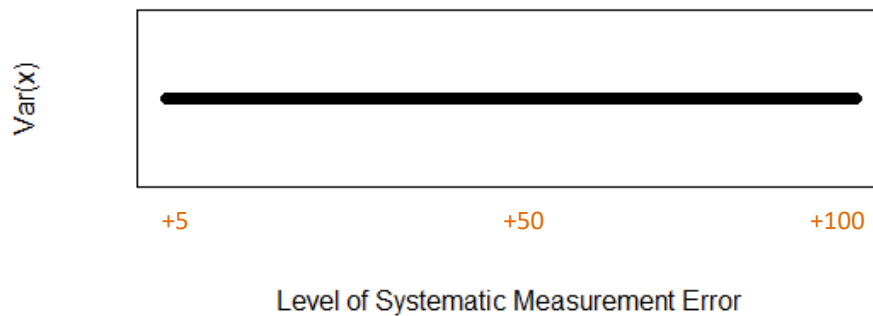
Covariance X,Y as a Function of Measurement Error



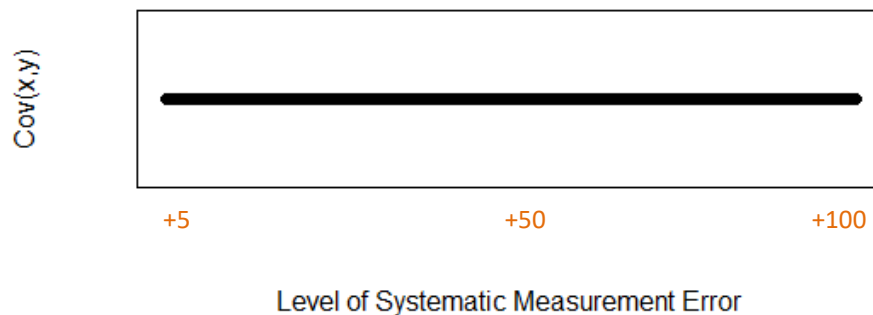
Incrementally add Measurement Error to X,
Hold Y constant

NON-RANDOM MEASUREMENT ERROR EXAMPLE

Variance of X as a Function of Measurement Error

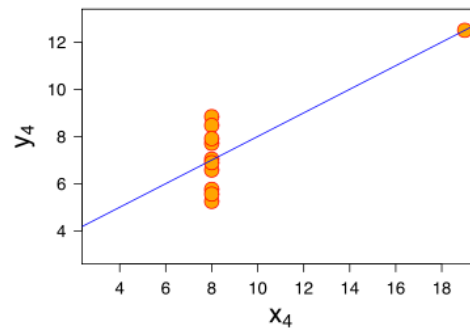
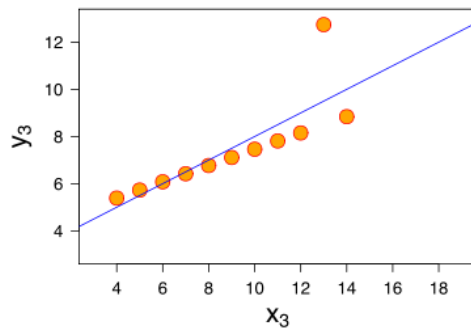
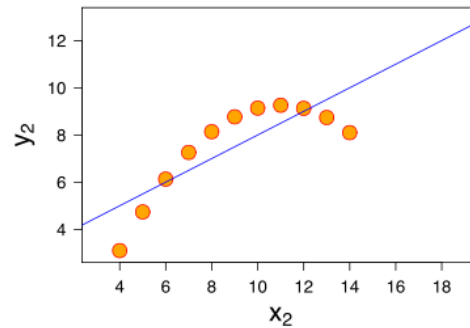
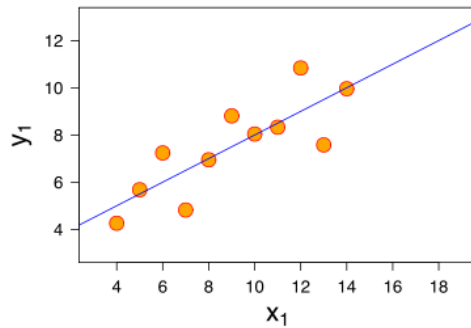


Covariance X,Y as a Function of Measurement Error



Systematic Mis-Measurement
(add same amount to each observation)

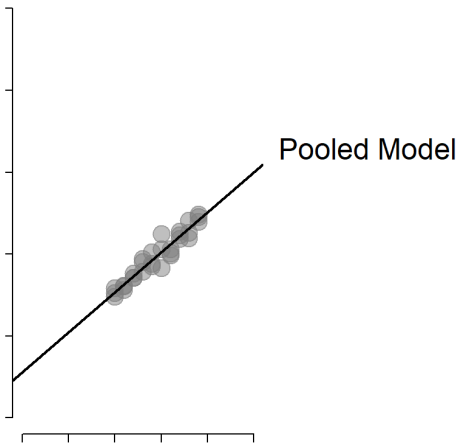
(4) MISSPECIFICATION BIAS



(5) GROUP DIFFERENCES (HETEROGENEITY BIAS)

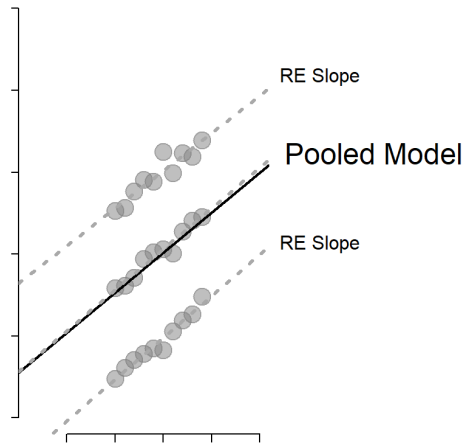
- If you have natural group structures in your data and there are **innate differences in the groups that are correlated with your study variable** then you will likely end up with heterogeneity bias in your estimates if you do not include the groups in your model.
- A group can be many individuals in one or more time-periods. A “group” can also be one individual measured over time.

Scenario 1



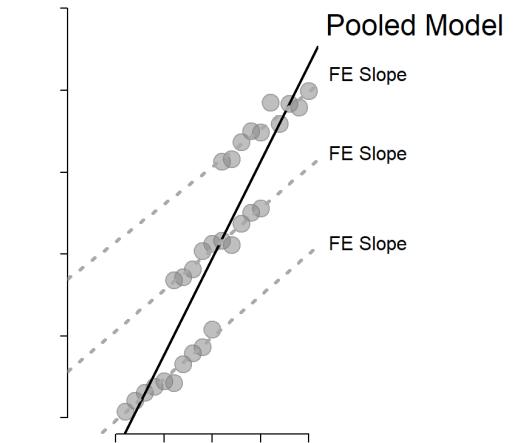
Pooled Model is UNBIASED + EFFICIENT

Scenario 2



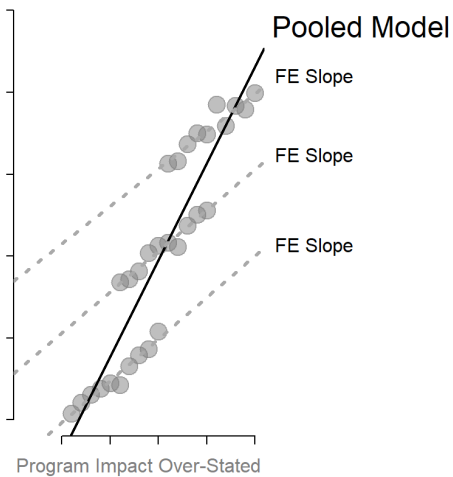
Pooled Model is UNBIASED + INEFFICIENT

Scenario 3



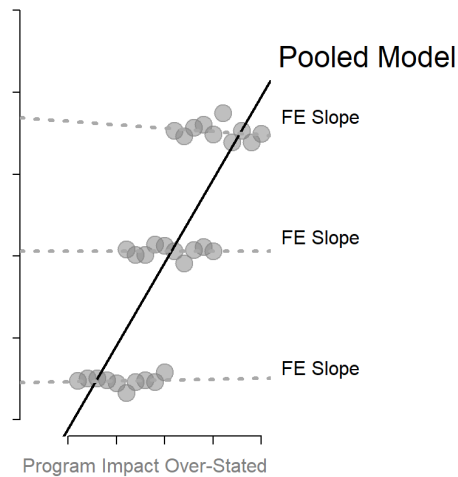
Pooled Model is BIASED + INEFFICIENT

Matthew Effect



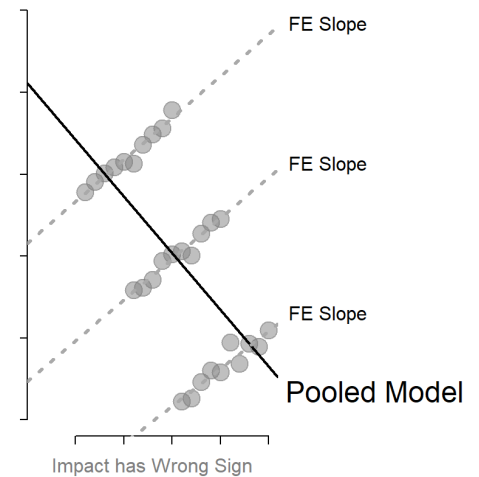
Program Impact Over-Stated

Artificial Program Impact



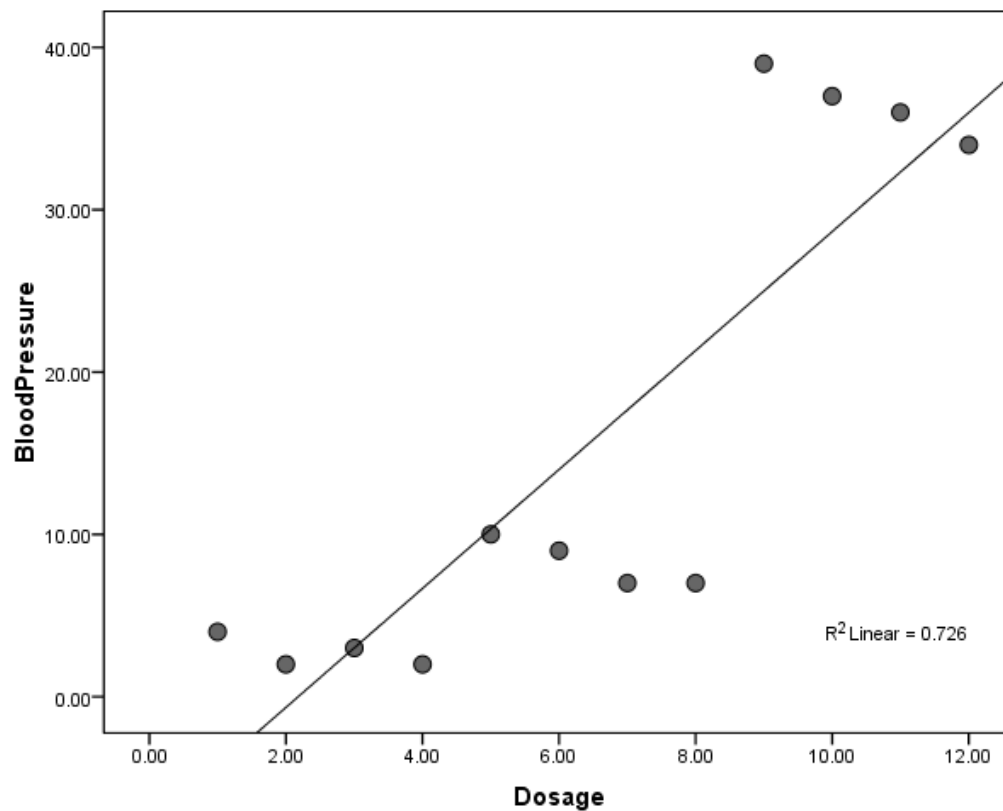
Program Impact Over-Stated

Simpson's Paradox



Impact has Wrong Sign

CROSS-SECTION VARIATION



REGRESSION MODEL:

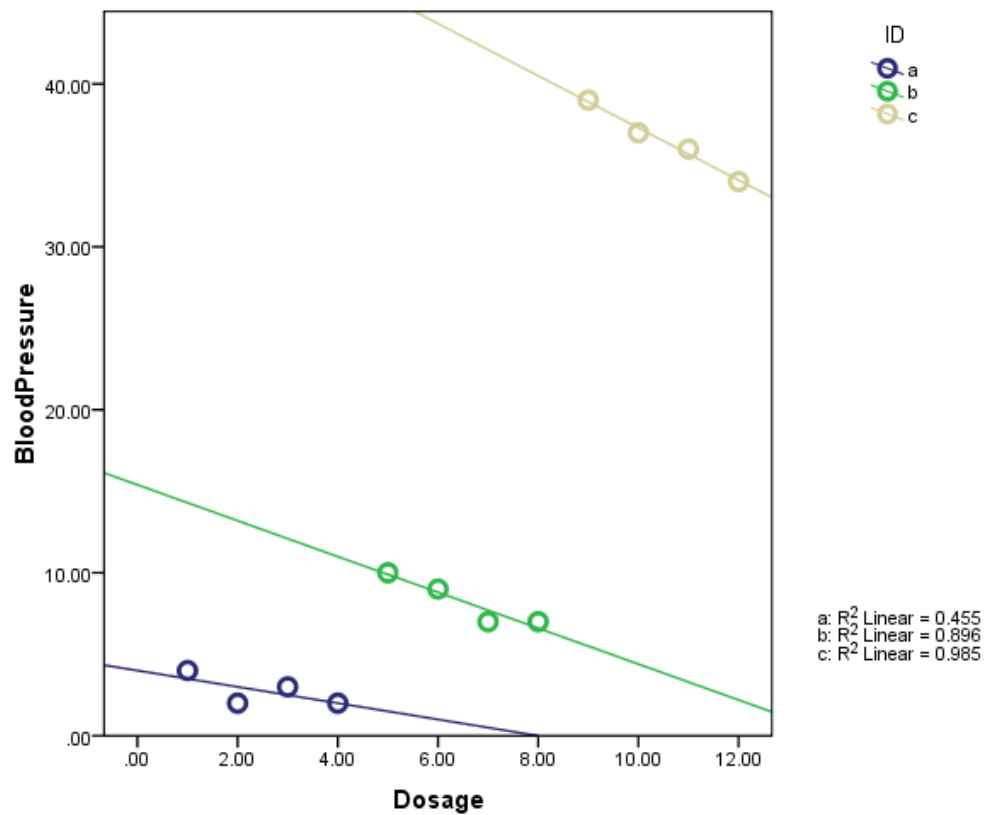
$$BP = b_0 + b_1 Dosage + e$$

Coefficients^a

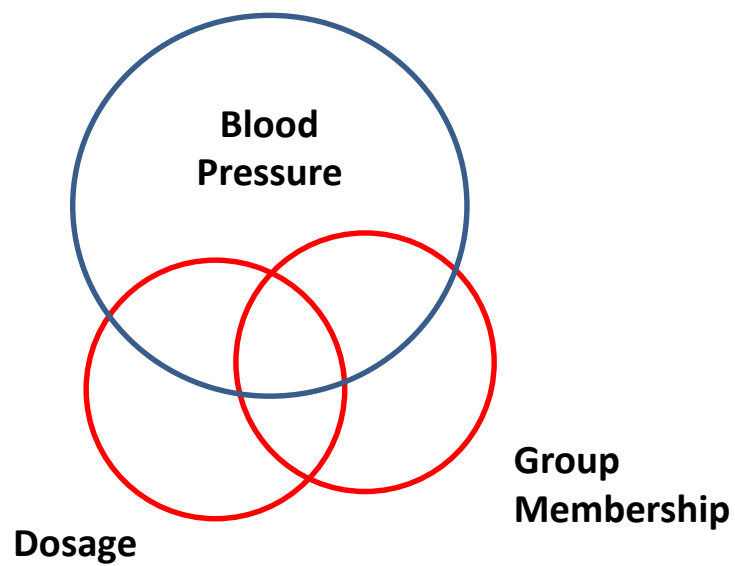
Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	-7.985	5.242		-1.523	.159
Dosage	3.664	.712	.852	5.144	.000

a. Dependent Variable: BloodPressure

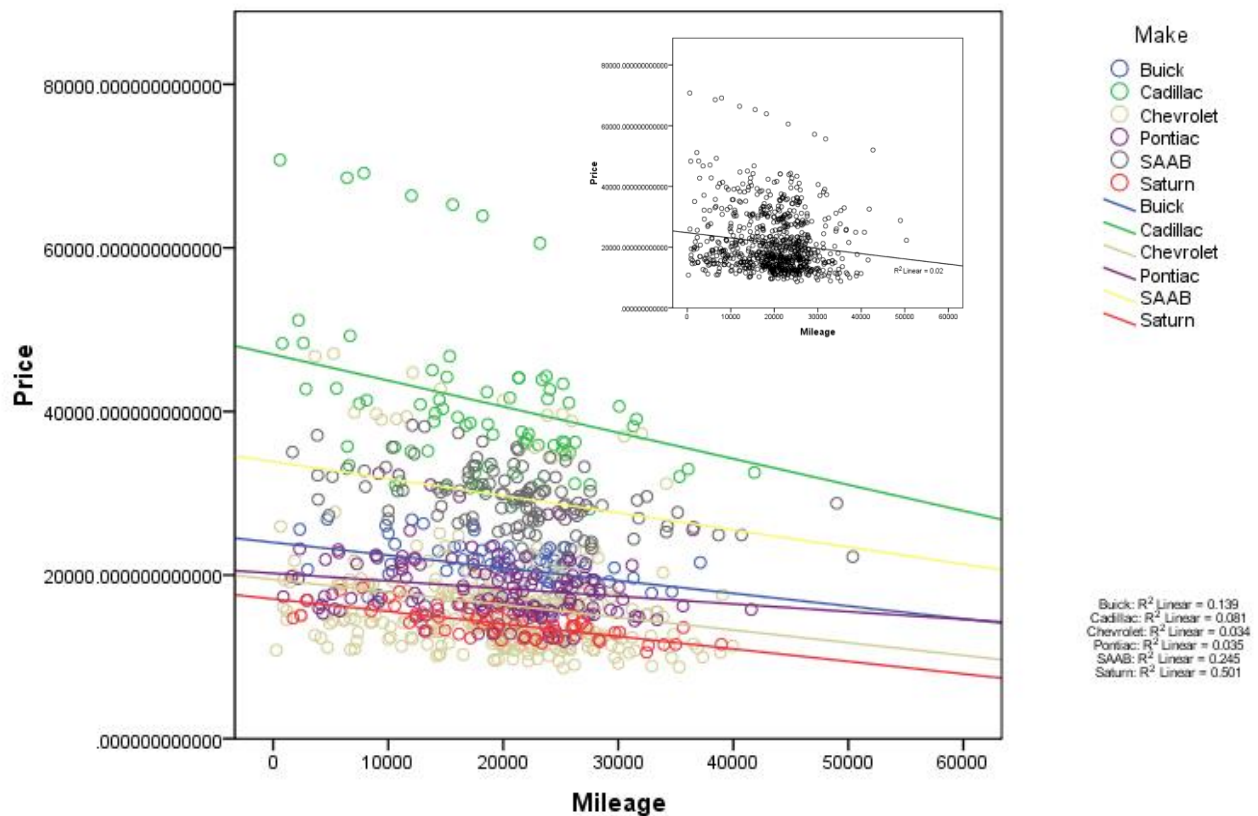
WITHIN GROUP VARIATION



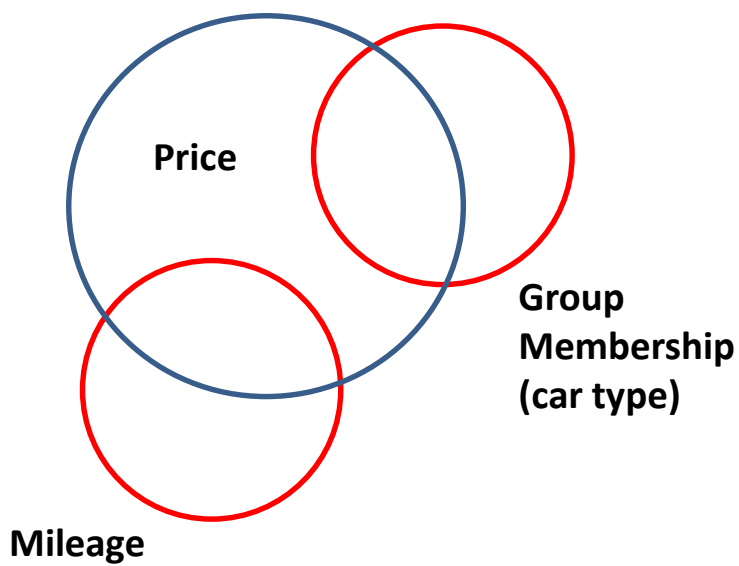
HETEROGENEITY BIAS



ANOTHER EXAMPLE: USED CAR SALES



HETEROGENEITY BIAS



$$bias = \beta_2 \alpha_1$$

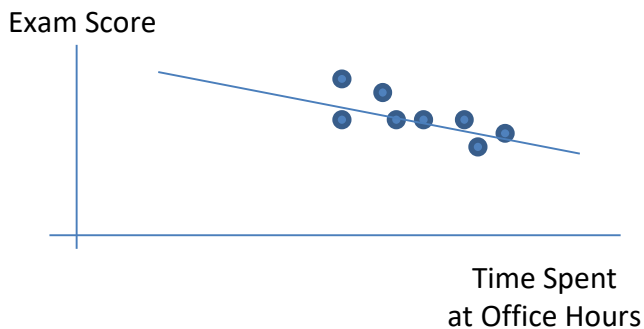
$$\alpha_1 = 0$$

$$\Rightarrow bias = 0$$

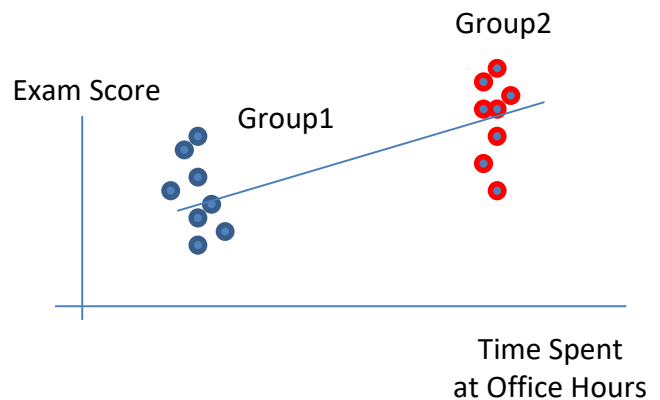
(6) BIAS VIA SELECTION

Does extra tutoring improve exam scores?

Self-Selected Group

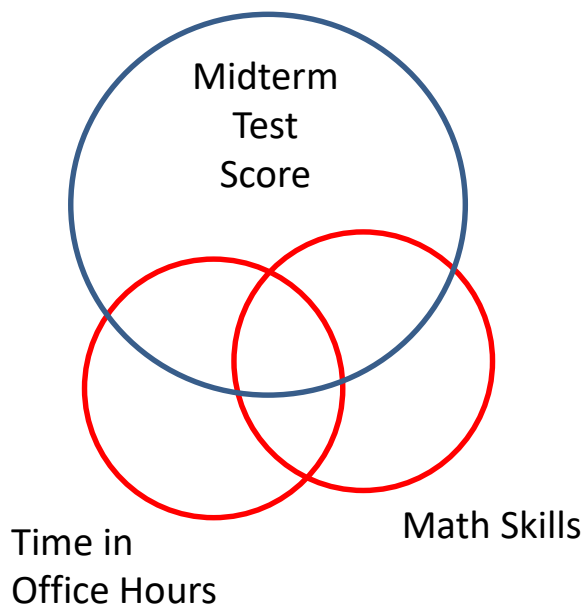


Assigned Groups

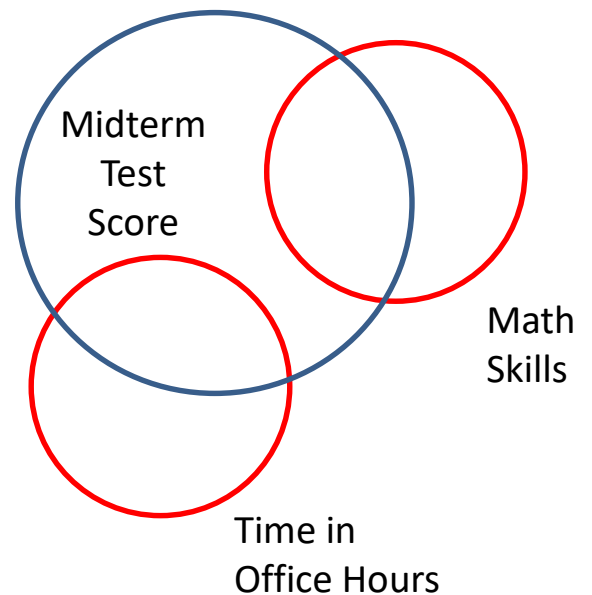


SELECTION BIAS: EXAMPLE 1

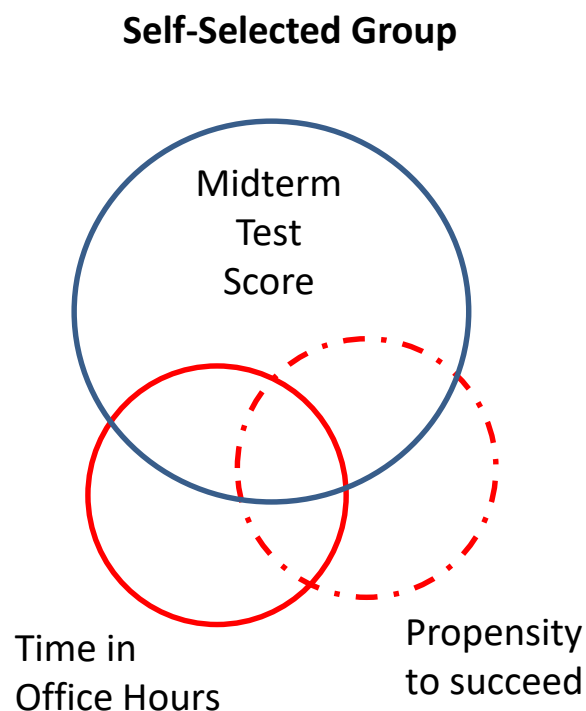
Self-Selected Group



Assigned Groups

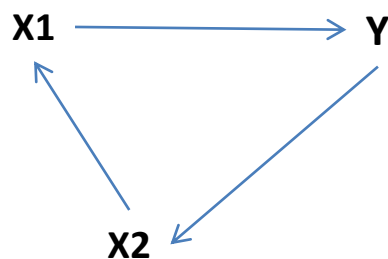


SELECTION BIAS: EXAMPLE 1



If ability is negatively or positively correlated with the decision to participate in the program then those in the treatment group (utilize service / join program) will be different than those in the control group

(7) SIMULTANEITY BIAS



When the causal structure forms a feedback loop. For example, monetary policy at the national level (changing interest rates impacts inflation, which impact growth, which impacts interest rates).

It is very difficult to determine the independent effects in this case.

We will not cover an example of this in class.

Just make a mental note, when you hear simultaneity it comes from this feedback structure.