# Partial correlations

# Today

- path diagrams
- partial and semi-partial correlations

# Causal relationships

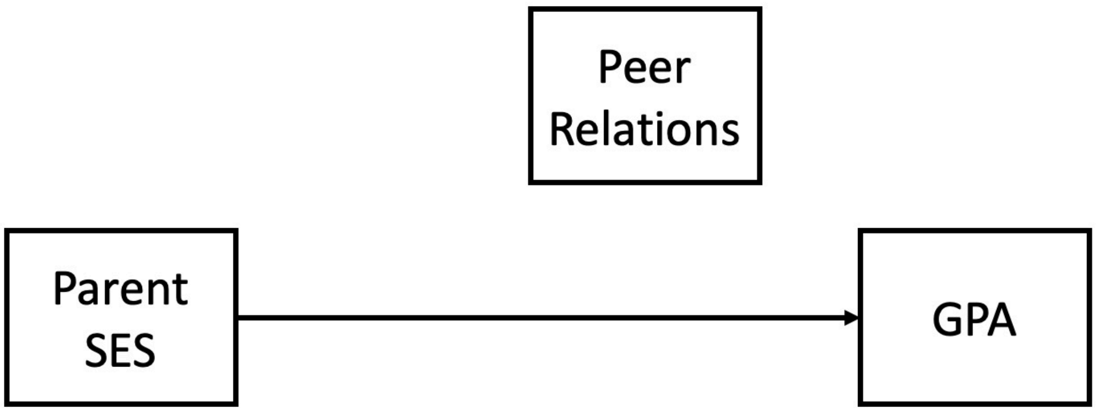
Does parent socioeconomic status *cause* better grades?

- $r_{GPA,SES} = .33, b = .41$

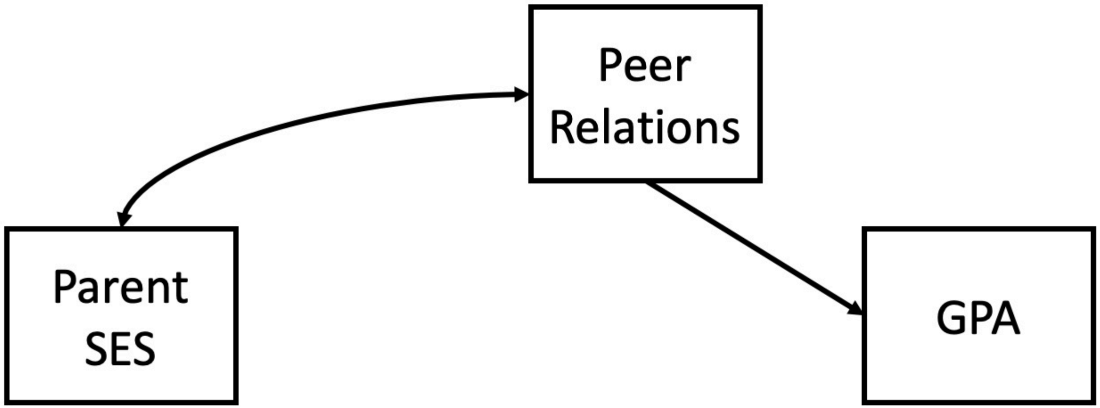
Potential confound: Peer relationships

- $r_{SES,peer} = .29$
- $r_{GPA,peer} = .37$

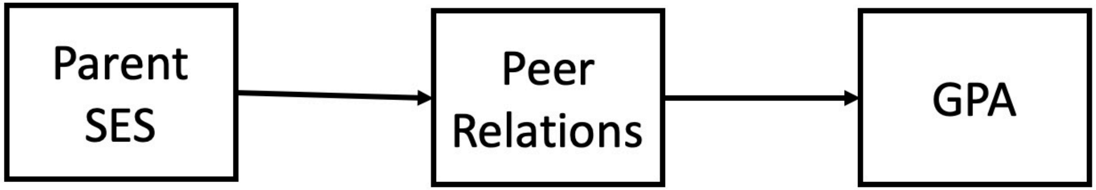
# Does parent SES cause better grades?



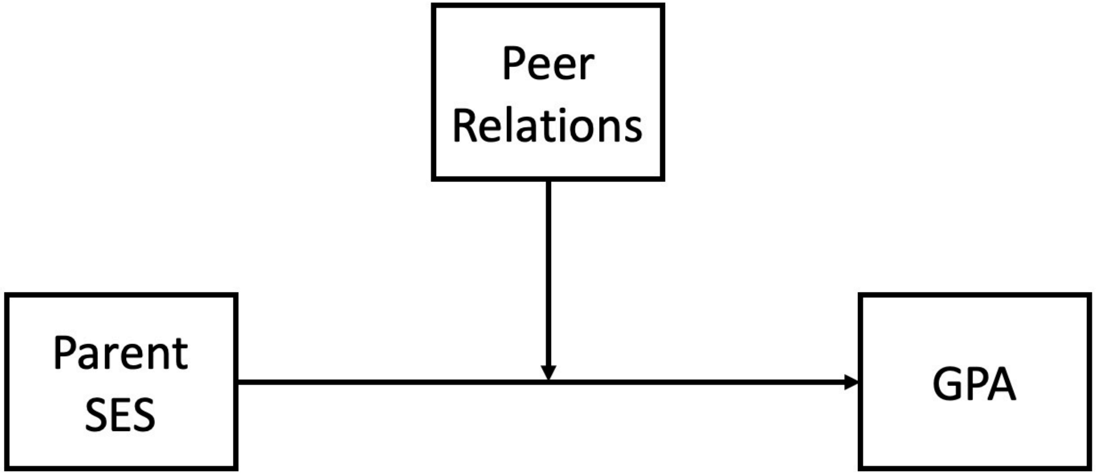
# Spurious relationship



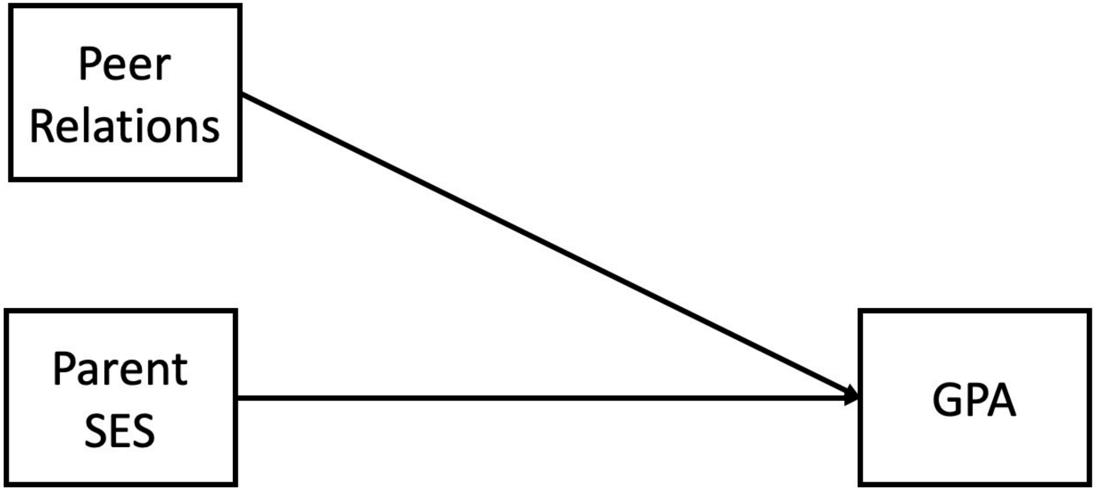
# Indirect (mediation)



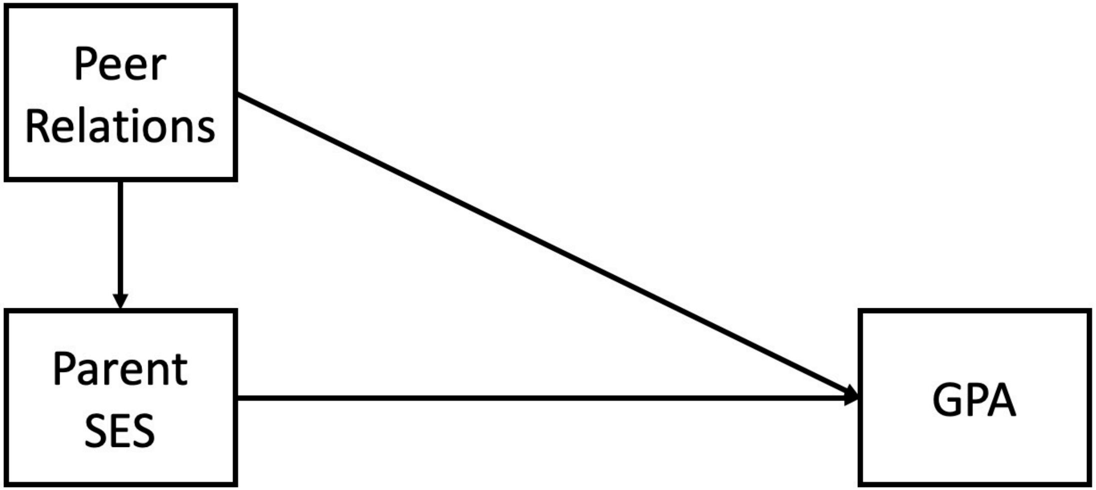
# Interaction (moderation)



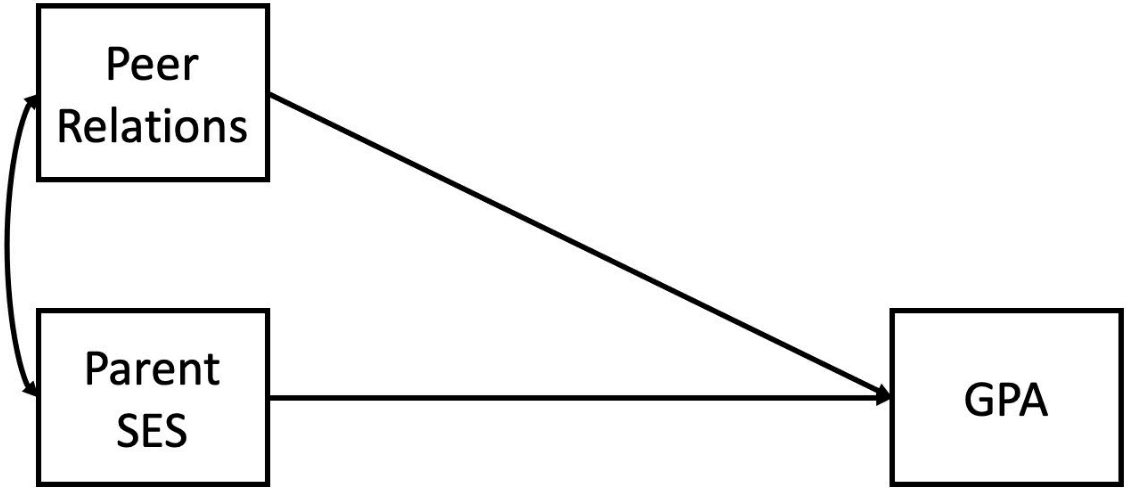
# Multiple causes



# Direct and indirect effects



# Multiple regression



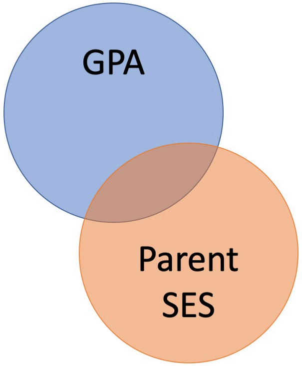
# General regression model

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \cdots + b_k X_k$$

This is ultimately where we want to go. Unfortunately, it's not as simply as multiplying the correlation between Y and each X by the ratio of their standard errors and stringing them together.

Why?

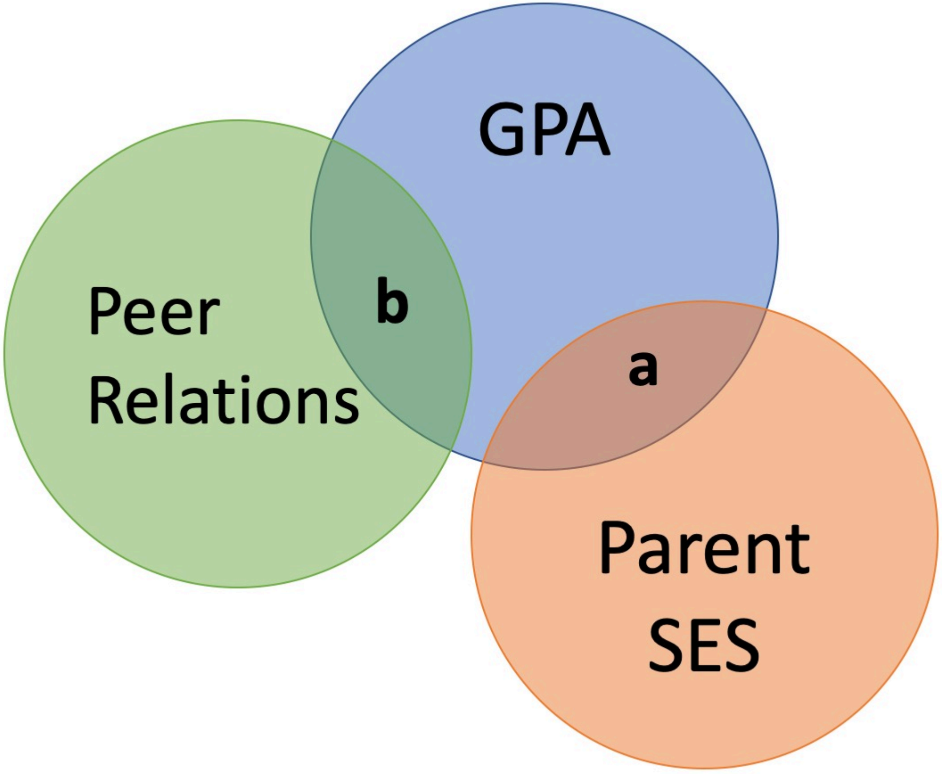
# What is $R^2$?

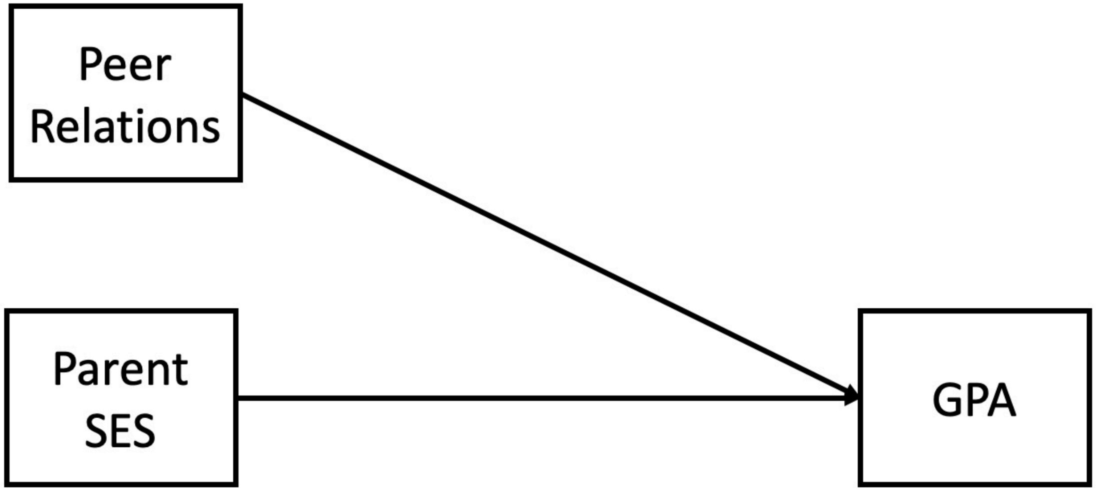


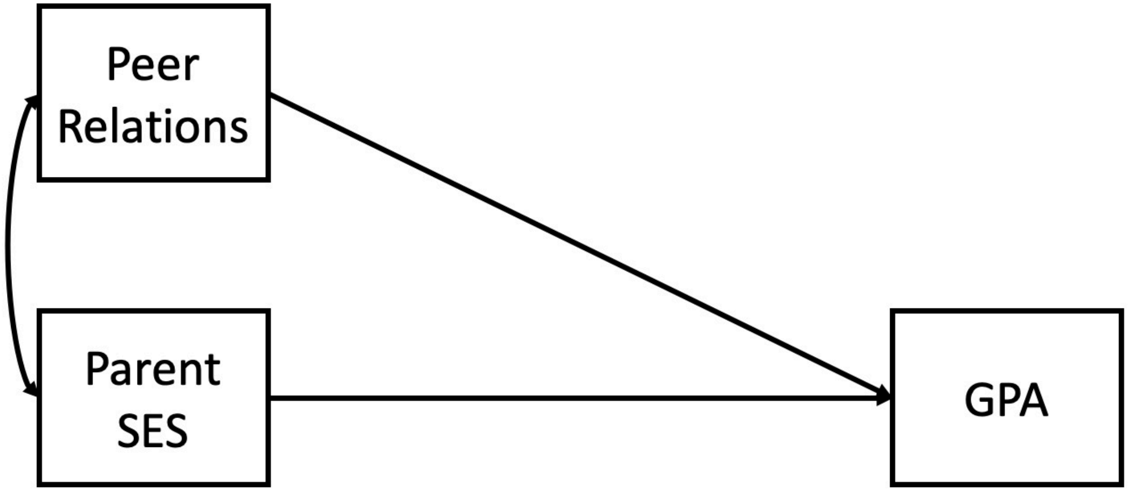

$$R^2 = \frac{s^2_{\hat{Y}}}{s^2_Y}$$

$$R^2 = \frac{SS_{\text{Regression}}}{SS_Y}$$

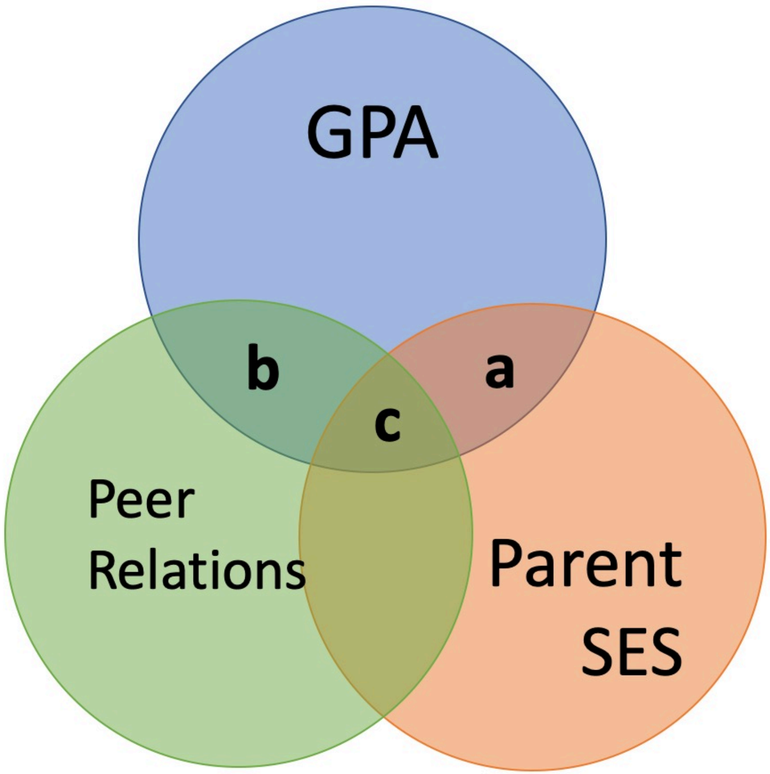
# GPA = SES + Peer Relationships



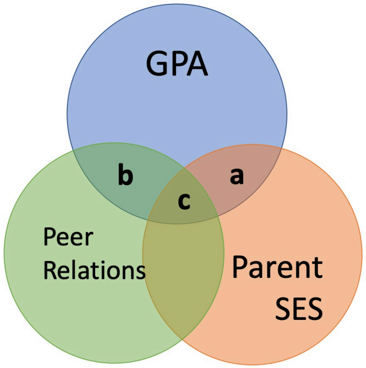

Peer Relations
Parent SES
GPA


Peer Relations
Parent SES
GPA

# What is $R^2$?



# What is $R^2$?




$$R^2_{Y.12} = a + b + c$$

# How do we control for something?

## Experimental Control

- Control variances (via equal groups)
- Randomly assign people to a group
- Conditions are the same except the IV

## Statistical Control

- Control variance by removing unwanted variance from the other variables
- Control by **"partialling"**

# Types of correlations

Pearson product moment correlation

- Standard correlation measure
- Ignores all outside variables
- aka bi-variate correlation
- Zero-order correlation
- Only two variables are X and Y

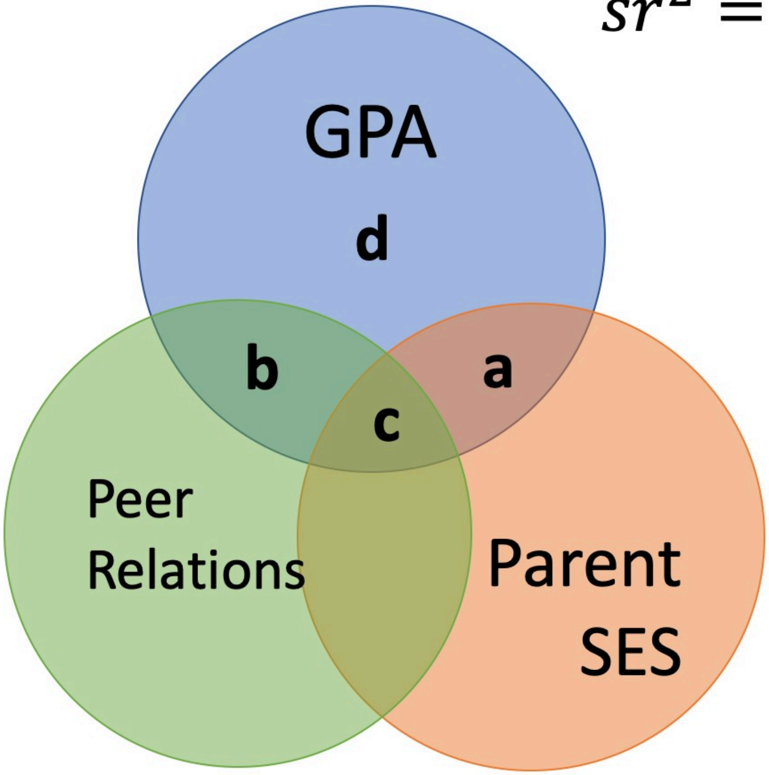
# Semi-partial correlations

- This correlation assess the extent to which the part of $X_1$ *that is independent of* of $X_2$ correlates with *all* of Y

- This is often the estimate that we refer to when we talk about **controlling for** another variable.

- Notation: $sr$ or $r_{Y(1.2)}$; $sr^2$ is the semi-partial correlation, squared

$$sr^2 = \frac{a}{a+b+c+d}$$



# Semi-partial correlations

$$sr_1 = r_{Y(1.2)} = \frac{r_{Y1} - r_{Y2}r_{12}}{\sqrt{1 - r_{12}^2}}$$

$$sr_1^2 = R_{Y.12}^2 - r_{Y2}^2$$

Notation:

- $R_{Y.12}^2$ is your standard $R^2$ -- variance explained in what is before the dot by what is after the dot. *"Variance explained in Y by X1 and X2"*
- $r_{Y(1.2)}$ is another way to write the semi-partial correlation.

# Numbers to these correlations

$$r_{Y1} = .3 \, r_{Y2} = .3 \, r_{12} = .2$$

```
sr1 = (.3 - (.3*.2)) / sqrt(1 - (.2^2))
```

$$r_{Y1} = .3 \, r_{Y2} = .3 \, r_{12} = .5$$

```
sr2 = (.3 - (.3*.5)) / sqrt(1 - (.5^2))
sr1
```

```
## [1] 0.244949
```

```
sr2
```

```
## [1] 0.1732051
```

# Types of correlations

## Pearson product moment correlation

## Semi-partial correlation

- This correlation assess the extent to which the part of $X_1$ *that is independent of* of $X_2$ correlates with all of Y
- This is often the estimate that we refer to when we talk about **controlling for** another variable.
- Notation: $sr$ or $r_{Y(1.2)}$; $sr^2$ is the semi-partial correlation, squared

# Types of correlations

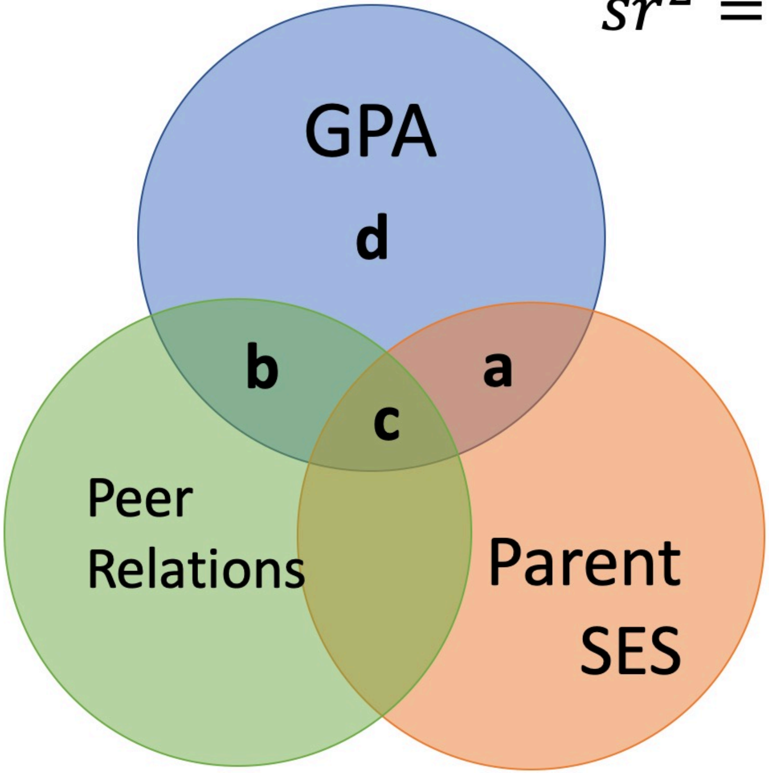
## Pearson product moment correlation

## Semi-partial correlation

## Partial correlation

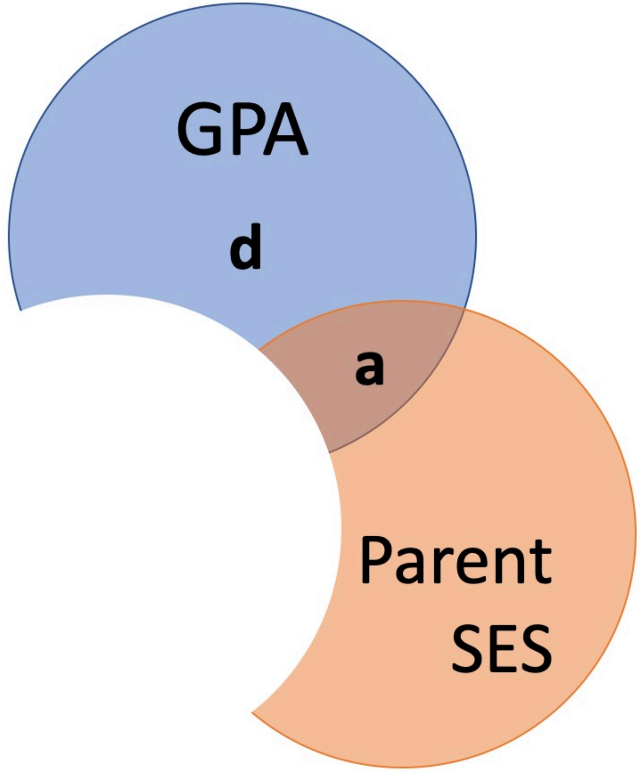
- The extent to which the part of $X_1$ *that is independent* of $X_2$ is correlated with *the part of Y that is also independent of* $X_2$.
- Notation: $pr$ or $r_{Y1.2}$; $pr^2$ is the partial correlation, squared

# Partial Correlations




$$sr^2 = \frac{a}{a+b+c+d}$$

$$pr^2 = \frac{a}{a+d}$$


GPA
d
a
Parent
SES

# Partial correlations

$$pr_1 = r_{Y1.2} = \frac{r_{Y1} - r_{Y2}r_{12}}{\sqrt{1 - r_{Y2}^2}\sqrt{1 - r_{12}^2}}$$

$$= \frac{r_{Y(1.2)}}{\sqrt{1 - r_{Y2}^2}}$$

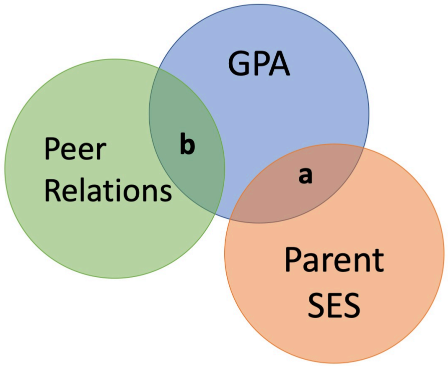
# Partial correlation

$$pr^2 = \frac{R^2_{Y.12} - r^2_{Y2}}{\sqrt{1 - r^2_{Y2}}}$$

$$sr^2 = R^2_{Y.12} - r^2_{Y2}$$

# What happens if X1 and X2 are uncorrelated?

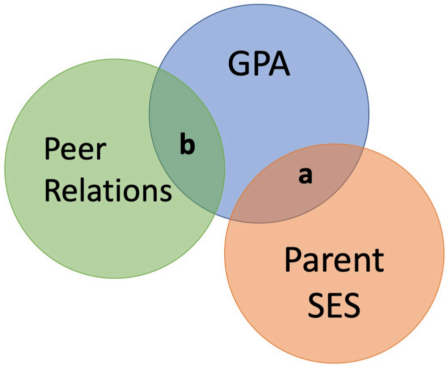
How does the semi-partial correlation compare to the zero-order correlation?



# What happens if X1 and X2 are uncorrelated?

How does the semi-partial correlation compare to the zero-order correlation?

$$r_{Y(1.2)} = r_{Y1}$$

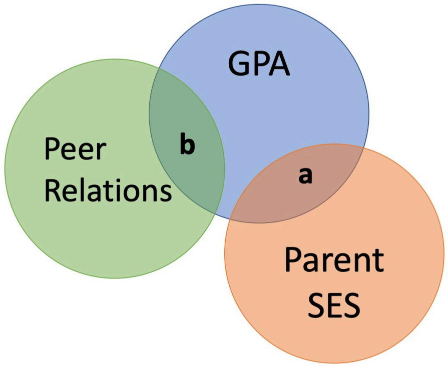


## What happens if X1 and X2 are uncorrelated?

How does the semi-partial correlation compare to the zero-order correlation?

$$r_{Y(1.2)} = r_{Y1}$$

How does the partial correlation compare to the zero-order correlation?



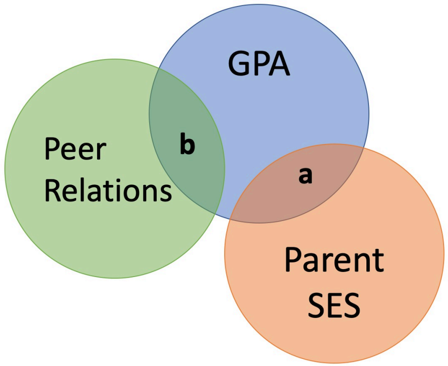
# What happens if X1 and X2 are uncorrelated?

How does the semi-partial correlation compare to the zero-order correlation?

$$r_{Y(1.2)} = r_{Y1}$$

How does the partial correlation compare to the zero-order correlation?

$$r_{Y1.2} \neq r_{Y1}$$



## When we use these?

Semi-partial: used when we want to show that some variable adds incremental variance in Y above and beyond another X variable. How much unique variance an IV adds to the TOTAL amount of variance.

- e.g., predicting Alzheimer's

Partial: used when some third variable, Z, is a plausible explanation of the correlation between X and Y. How much unique variance adds to the unaccounted variance in Y.sq

- e.g., predicting grades

# Example with Flowers

Iris dataset from before. Sepal.Length, Sepal.Width, Petal.Length

```
library(ppcor)
# zero-order correlation
round(cor(iris[,c("Sepal.Length", "Sepal.Width", "Petal.Lengt
```

```
##              Sepal.Length Sepal.Width Petal.Length
## Sepal.Length         1.00       -0.12         0.87
## Sepal.Width         -0.12        1.00        -0.43
## Petal.Length         0.87       -0.43         1.00
```

# Example with Flowers

Iris dataset from before. Sepal.Length, Sepal.Width, Petal.Length

```
# semi-partial correlation
round(spcor(iris[,c("Sepal.Length", "Sepal.Width", "Petal.Len
```

```
##              Sepal.Length Sepal.Width Petal.Length
## Sepal.Length         1.00        0.28         0.91
## Sepal.Width          0.52        1.00        -0.67
## Petal.Length         0.83       -0.33         1.00
```

# Example with Flowers

Iris dataset from before. Sepal.Length, Sepal.Width, Petal.Length

```
# partial correlation
# semi-partial correlation
round(pcor(iris[,c("Sepal.Length", "Sepal.Width", "Petal.Leng
```

```
##              Sepal.Length Sepal.Width Petal.Length
## Sepal.Length         1.00        0.58         0.92
## Sepal.Width          0.58        1.00        -0.67
## Petal.Length         0.92       -0.67         1.00
```

## $e$ are part of $Y$ that is independent of $X$

$$\hat{Y} = b_0 + b_1 X$$

$$e_i = Y_i - \hat{Y}_i$$

We can use this to construct a measure of $X_1$ that is independent of $X_2$:

$$\hat{X}_{1.2} = b_0 + b_1 X_2$$

$$e_{X_1} = X_1 - \hat{X}_{1.2}$$

We run a regression where X2 predicts X1. then we take the error from that, which is the part of X1 that is independent of X2.

We can either correlate that value with Y, to calculate our semi-partial correlation:

$$r_{e,x_1,Y} = r_{y(1.2)}$$

Or we can calculate a measure of Y that is also independent of $X_2$ and correlate that with our $X_1$ residuals.

$$\hat{Y} = b_0 + b_1 X_2$$

$$e_y = Y - \hat{Y}$$

$$r_{e_{X_1},e_y} = r_{y1.2}$$

# Example

```
# create measure of Sepal.Length *independent* of Sepal.Width
# model where X1 = SLength, X2 = SWidth
# error is part of SLength that is independent of SWidth
mod.sepals = lm(Sepal.Length ~ Sepal.Width, data = iris)
indep.Sepals = broom::augment(mod.sepals)

# correlate with Y = PLength to get semi-partial
round(cor(indep.Sepals$.resid, iris$Petal.Length), digits = 2
```

```
## [1] 0.83
```

```
# Measure of Y = PLength that is also independent of X2 = SWi
# create measure of Petal.Length independent of Sepal.Width s
mod.petals = lm(Petal.Length ~ Sepal.Width, data = iris)
indep.Petals = broom::augment(mod.petals)

# Correlate those residuals with other residuals to get parti
round(cor(indep.Sepals$.resid, indep.Petals$.resid), digits =
```

```
## [1] 0.92
```

# Who cares? Where are we going?

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_p X_p$$

- Regression coefficients are "partial" regression coefficients

- Predicted change in $Y$ for a 1 unit change in $X$ *holding all other predictors constant*

- Similar to semi-partial correlations -- represents part of each $X$

# Next time...

Multiple regression!!