Mediation

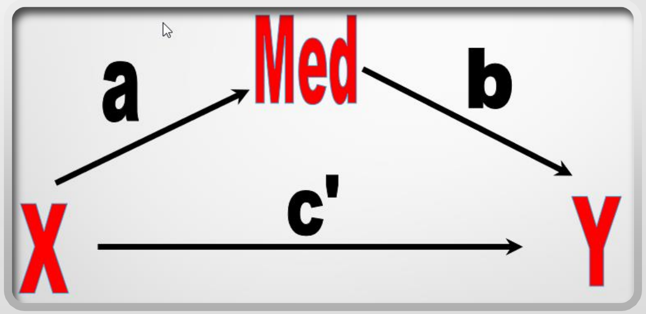
More specifically … Statistical Mediation

Introduction to Mediation

- Mediation involves exploring a variable (mediator) that might explain the relationship between a predictor/IV and outcome/DV
  - There could be multiple different mediators, one mediator could mediate multiple relationships, etc.
- A mediating variable is the *mechanism* by which an effect occurs between a predictor and an outcome

Mediation Model

- *c* represents the effect of *X* on *Y*
  - E.g., unstandardized regression coefficient
- *a* represents the effect of X on the mediator
  - E.g., unstandardized regression coefficient
- *b* represents the effect of the mediator on *Y*, after controlling for *X*
  - Partial coefficient
- *c*' represents the effect of *X* on *Y* after controlling for the mediator
  - Partial coefficient

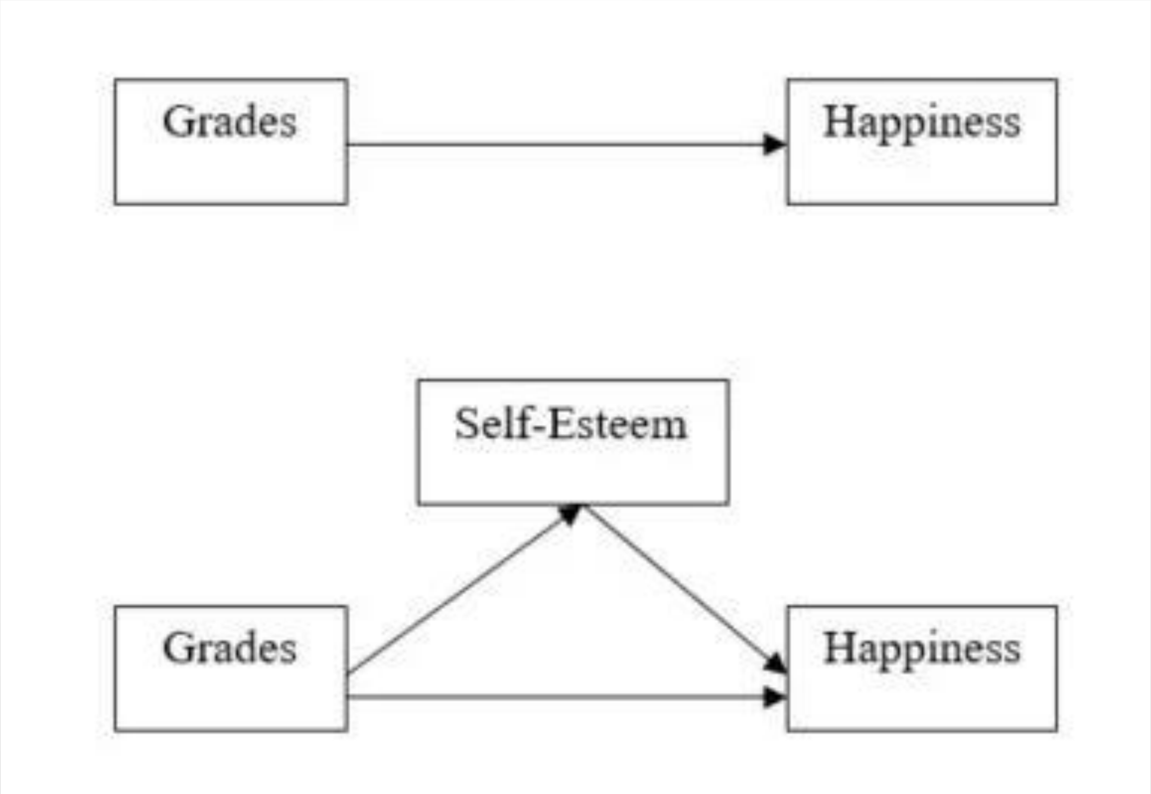


Research Question in Mediation

- Rephrasing the important question:
  - Does the simple relationship between *X* and *Y* (*c*) shrink in magnitude or disappear when the mediator is present in the model
    - i.e., $|c'| < |c|$, or $c' = 0$
  - A relationship that shrinks indicates partial mediation, whereas a relationship that disappears indicates full/complete mediation
    - Rarely would a mediator control the entire relationship (full mediation), but might control part of the relationship (partial mediation)

Mediation Example

- Self-esteem is a ‘mechanism’ that controls how grades affect happiness



Decomposition of Effects

- Total Effect = Indirect Effect + Direct Effect
- Indirect Effect (*a*b*)
  - Product of the *a* and *b* coefficients (*ab*)
  - A one unit increase in X leads to an *a* unit change in M, which hence leads to an *a*b* unit change in Y
  - E.g., $a = 2$, $b = 3$
    - Increasing *X* by one unit is expected to increase *M* by 2 units
    - A 1 unit change in *M* is expected to increase *Y* by 3 units
    - Thus, a 1 unit increase in *X* increases *M* by 2 units, and a 2 unit increase in *M* increases *Y* by $2 \times 3 = 6$ units (which equals $a*b$)
  - I.e., we must multiply (not add) *a* and *b* to properly estimate the indirect effect

Decomposition of Effects

- Direct Effect (c′)
  - The effect of X on Y, after controlling for M
- Total Effect (c) = Indirect Effect (ab) + Direct Effect (c′)
  - $c = ab + c'$
  - Thus:
    - $ab = c - c'$
- These relationships hold for most, but not all, relationships (e.g., this may not hold for multilevel models, logistic models)

Baron & Kenny

- The most famous reference to mediation is Baron & Kenny (1986)
    - Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173-1182
- This paper has been cited more than *120000* times!!

Why was Baron & Kenny so popular?

- The article laid out mediation in a set of simple steps:
  1) Test X -> Y (path c)
  2) Test X -> M (path a)
  3) Test M -> Y, controlling for X (path b)
  4) Test X -> Y, controlling for M (path c´)

  - If steps 1 to 3 are significant, and c´ is reduced then we have partial mediation (or if c´ = 0 we have full mediation)

Criticisms of the Baron & Kenny Steps

- The steps encouraged hypothesis testing, rather than effect size estimation
- A significant X to Y (*c*) path is not necessarily required (e.g., the direct and indirect effect may be of opposite signs, or X and Y might be separated by a long period of time)
- The M to Y (*b*) path may not be significant when X and M are highly correlated (multicollinearity)
- A statistical test of X to Y in the last step is meaningless, since it cannot prove partial or full mediation

Interpreting/Testing the Indirect Effect?

- The most important effect in mediation is the *indirect effect (ab)*
  - The effect on Y of increasing X by one unit, through M
- From a null hypothesis testing perspective, there are two popular approaches for testing the indirect effect (i.e., whether $c' < c$)
  - Sobel Test (traditional method)
    - Reasonable when sample sizes are large (>100)
  - Bootstrapping (modern approach)
    - Good test for small or large sample sizes

Sobel Test

- The Sobel test assesses the statistical significance of the indirect effect (*ab*)
    - In other words, it assesses whether a significant part of the relationship between X and Y is mediated
  - Let s_a and s_b represent the standard errors of a and b, and reject H_0: $ab = 0$ if $z > z_{1-\alpha}$, where:

$$z = \frac{ab}{\sqrt{a^2 s_b^2 + b^2 s_a^2}}$$

Problem with the Sobel Test

- The distribution of *ab* is highly positively skewed, but the Sobel test assumes a normally distributed sampling distribution
    - Reasonable with large N, but inefficient with smaller N

Bootstrapping

- Bootstrapping allows researchers the opportunity to resample from the data in order to generate an empirical sampling distribution of *ab*
  - Assumes that the sample distribution = population distribution
  - Sample N cases <u>with replacement</u> from a distribution of scores, and calculate the coefficient of interest on each sample
    - The average coefficient across bootstrap samples is the estimate of the coefficient, and the *sd* of the coefficients across the bootstrap samples is the standard error
- See ‘bootstrapping_example.R’

Bootstrapping

- For our interests, we want to use bootstrapping to gain information regarding the indirect effect (*ab*)
- Researchers can use the empirical (bootstrap) estimates of *ab* to generate a confidence interval for *ab*
  - This CI will generally be asymmetric

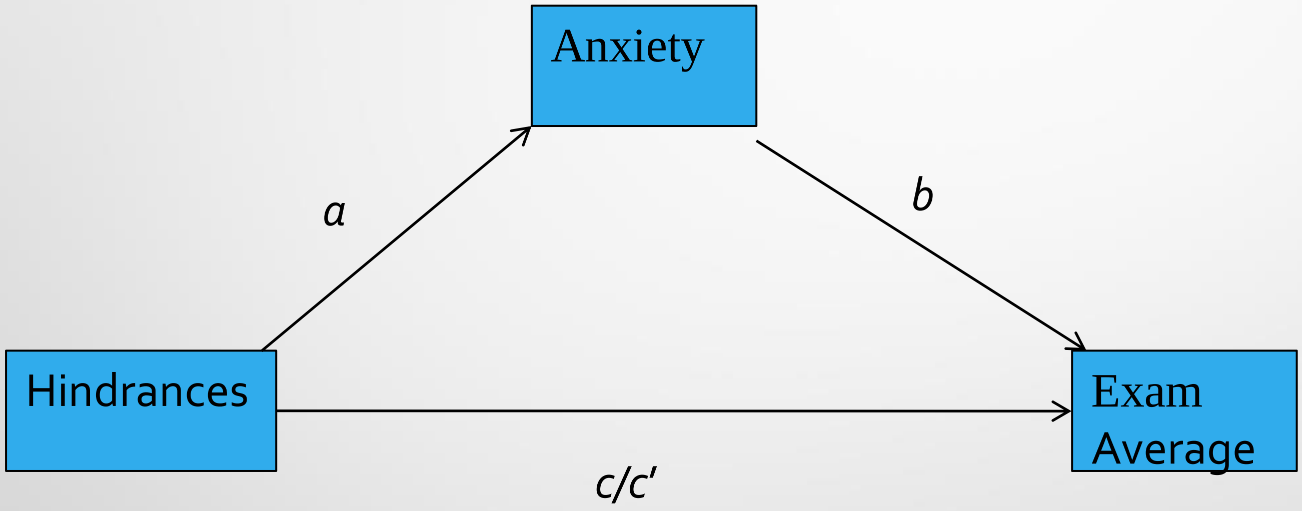
Effect Size Measures

- Upsilon
  - 'Completely Standardized Indirect Effect'
  - Product of the squared Beta (β) coefficients for a and b
- Proportion of the Total Effect that is Mediated
  - Proportion Mediated = ab / c

Example

- Let's say we are interested in determining if anxiety (anx) mediates the relationship between hindrances to doing well in a stats course (hindr) and the exam average (exavg)
  - r(anx,hindr)=.452*
  - r(anx,exavg)=-.367*
  - r(hindr,exavg)=-.303*
- Let's use α = .05

Mediational Model



Regression Equations

- Predicting *exavg* from *hindr (c)*:
    - *exavg* = -3.874 (*hindr*) + 91.825 ($p < .001$)
- Predicting *anx* from *hindr (a)*:
    - *anx* = .552 (*hindr*) + .335 ($p < .001$)
- Predicting *exavg* from both *hindr* (*c'*) and *anx (b)*:
    - *exavg* = -2.203 (*hindr*) -3.027 (*anx*) + 94.352
    - *p*(*hindr*) = .062; *p*(*anx*) = .002
- Historically, since *p(hindr)* > α, some would say (but we won't) that *anx* <u>fully</u> *mediates* the prediction of *exavg* from *hindr*
- Recall though, that the effect of interest is the indirect effect (*ab*)
    - *ab* = .552 * -3.027 = -1.67
        - Increasing *hindr* by one unit is expected to reduce *exavg* by -1.67 units, through *anx*
    - We will explore the CI for this effect via bootstrapping in R

Added Variable Plots

- The relationship between *examavg* and *hindrances* is weaker after controlling for *anxiety*
  - Although it is hard to see because the X axis scale changes across the two graphs