Chapter 3: Single Mediator Model

- Limitations of verbal descriptions
- Single mediator model
- Statistical Mediation Analysis
- Tests of the mediated effect

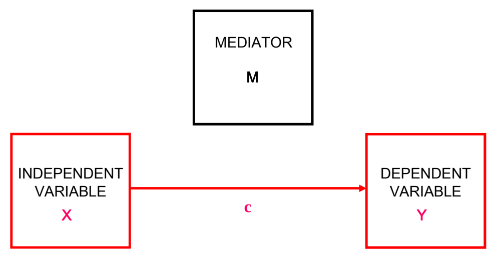
Three ways to specify a model

- Verbal description: A variable M is intermediate in the causal sequence relating X to Y.
- Diagram
- Equations

Mediation Regression Equations

- Tests of mediation for a single mediator use information from some or all of three equations.
- The coefficients in the equations may be obtained using methods such as ordinary least squares regression, covariance structure analysis, or logistic regression.

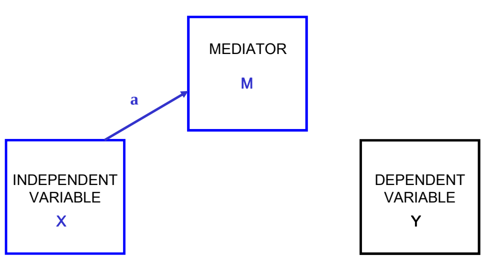
Equation 1



1. The independent variable is related to the dependent variable:

$$Y = i_1 + \hat{c}X + e_1$$

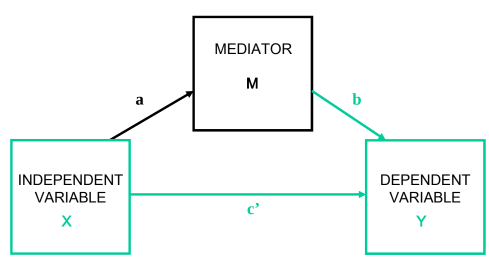
Equation 2



2. The independent variable is related to the potential mediator:

$$M = i_2 + \hat{a}X + e_2$$

Equation 3



3. The mediator is related to the dependent variable controlling for exposure to the independent variable:

$$Y = i_3 + \hat{c}' X + \hat{b}M + e_3$$

Mediated Effect Measures

Indirect Effect = Mediated effect = ab = c-c'

Direct effect= c' Total effect= ab+c'=c

7

Mediated Effect, $\hat{a}\hat{b}$, Standard Error

Mediated effect= $\hat{a}\hat{b}$ Standard error= $\sqrt{\hat{a}^2 s_{\hat{b}}^2 + \hat{b}^2 s_{\hat{a}}^2}$

Multivariate delta method standard error (Sobel 1982; Folmer 1981)

Test for significant mediation:

$$z' = \frac{\hat{a}\hat{b}}{\sqrt{\hat{a}^2 s_{\hat{b}}^2 + \hat{b}^2 s_{\hat{a}}^2}}$$

Compare to empirical distribution of the mediated effect

8

Reasons for Confidence Limits

- Gives a range of values based on a sample estimate.
- Helps avoid binary, significant or not, approach to research.
- Incorporates variability in the point estimate as well as the point estimate.

9

Confidence Limits for *ab*

Confidence Limits $\hat{a}\hat{b} \pm z_{crit}\ s_{\hat{a}\hat{b}}$

UCL = $\hat{a}\hat{b} + z_{crit}\, s_{\hat{a}\hat{b}}$

LCL = $\hat{a}\hat{b} - z_{crit}\ s_{\hat{a}\hat{b}}$

Where z_{crit} is the z critical value because the standard error is asymptotic. Valid to use t instead of z.

95% Confidence Limits
UCL = $\hat{a}\hat{b} + 1.96\ s_{\hat{a}\hat{b}}$
LCL = $\hat{a}\hat{b} - 1.96\ s_{\hat{a}\hat{b}}$

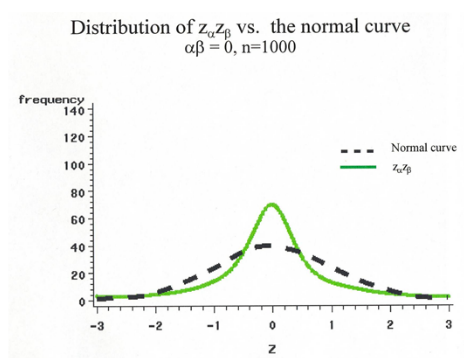
With normal distribution upper and lower critical values have the same value but opposite sign, e.g., 1.96 for $z_{.975}$ and -1.96 for $z_{.025}$

10

Distribution of the Product

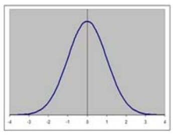
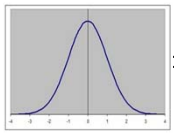
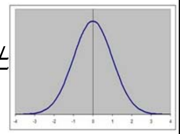
- The mediated effect is the product of two coefficients *a* and *b*. The distribution of the product has a normal distribution only in special cases (MacKinnon et al., 2002).
- At low values of *a* and *b*, the distribution has excess kurtosis and skewness, e.g. when *a* and *b* are both zero, kurtosis is 6. It is not surprising that the confidence limits are inaccurate if the distribution is assumed to be normal.
- One solution is to use the distribution of the product in statistical tests and confidence limits.

11



12

Product of Two Normal Distributions is not always Normal

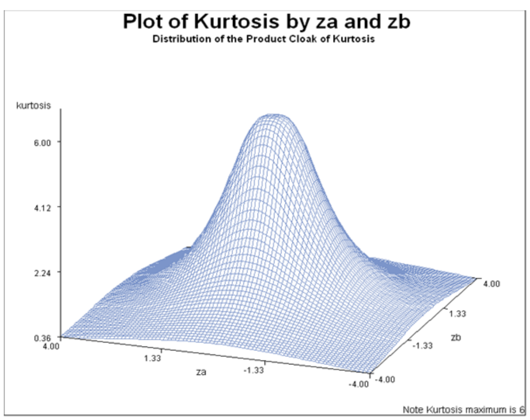
 X  ≠ 

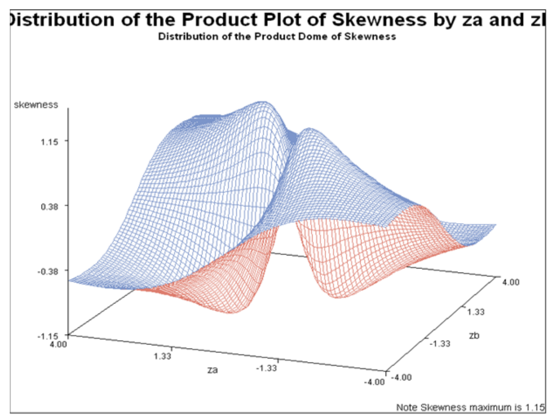
13

Plot of Kurtosis and Skewness of the Distribution of the Product

- The next two plots show the kurtosis and skewness of the distribution of the product as a function of $za= a/s_a$ and $zb=b/s_b$.
- The range of values for za and zb is from -4 to +4 in these plots. Applied research often has these *z* values, that is a z test for *a* and a z test for *b* range from -4 to 4.
- A normal distribution would have skewness and kurtosis of 0 for all values of *za* and *zb*. The distribution of the product has different values of skewness and kurtosis for values of *za* and *zb*.

14





Critical Values for Distribution of the Product

- Because the distribution of the product is not symmetric, there are different critical values for the distribution for each value of a/s_a and b/s_b.
- The critical values are -1.96 and +1.96 for the 95% confidence interval from the normal distribution. There are different upper and lower critical values for the distribution of the product. Confidence limits and significance tests are more accurate using the critical values from the distribution of the product (MacKinnon et al. 2004).

17

PRODCLIN (distribution of the PRODuct Confidence Limits for the INdirect effect)

- MacKinnon, Fritz, Williams, & Lockwood, (2007) describes program to compute critical values for the distribution of the product.
- Web location includes programs in SAS, SPSS, and R that access a FORTRAN program.
  http://www.public.asu.edu/~davidpm/ripl/Prodclin/
- Input estimates $\hat{a}$, $s_{\hat{a}}$, $\hat{b}$, $s_{\hat{b}}$, correlation between $\hat{a}$ and $\hat{b}$, and Type I error rate. Output includes the input values and normal and distribution of the product confidence limits.

18

RMediation

- Tofighi & MacKinnon (2011) describes an R program to find critical values for the distribution of the product that solves some problems in PRODCLIN, can get accurate results for cases where PRODCLIN did not converge, more accurate results for correlated z-values, makes plots of distributions and finds percentiles and probabilities.
- Input estimates $\hat{a}$, $s_{\hat{a}}$, $\hat{b}$, $s_{\hat{b}}$, correlation between $\hat{a}$ and $\hat{b}$, Type I error rate but input is now called mu.x, se.x, mu.y, se.y, rho, alpha, respectively.

19

Assumptions I

- For each method of estimating the mediated effect based on Equations 1 and 3 (*c-c'*) or Equations 2 and 3(*ab*):
  - Predictor variables are uncorrelated with the error in each equation.
  - Errors are uncorrelated across equations (*ab*) .
  - Predictor variables in one equation are uncorrelated with the error in other equation.
- Reliable and valid measures
- No omitted influences.
- Normally distributed variables

20

Assumptions II

- Data are a random sample from the population of interest.
- Coefficients, *a*, *b*, *c'* reflect true causal relations and the correct functional form.
- Mediation chain is correct: Temporal ordering is correct X before M before Y. Any mediation model is part of a longer mediation chain. The researcher decides what part of the micromediational chain to examine.
- Homogeneous effects across subgroups: It assumed that the relation from X to M and from M to Y are homogeneous across subgroups or other characteristics of participants in the study. This means there are not moderator effects.

21

Identification Assumptions

1. No unmeasured X to Y confounders given covariates.
2. No unmeasured M to Y confounders given covariates.
3. No unmeasured X to M confounders given covariates.
4. There is no effect of X that confounds the M to Y relation.

VanderWeele & VanSteelandt (2009)

22

Water Consumption Study Variables

- Stimulus->Organism->Response study
- X is the temperature in degrees Fahrenheit
- M is a self-report of thirst at the end of the first two hours of the study
- Y is the number of deciliters of water consumed during the last two hours of the study
- 50 participants were in a room for four hours doing a variety of tasks including sorting objects, tracking objects on a computer screen, and communicating via an intercom system

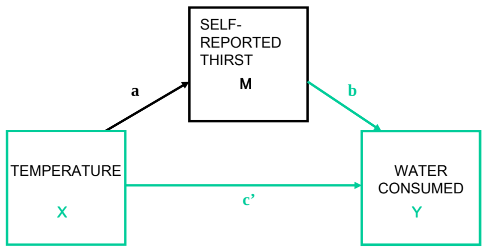
23

Water Consumption Study Purpose

- The purpose of the study was to investigate whether persons can judge their water needs. Temperature should affect self-reported thirst which then should affect water consumption.
- The accuracy of self-reported thirst is important because persons in self-contained environments need to monitor their own hydration.
- The mediated effect of temperature on water consumption through self-reported thirst estimates the extent to which persons were capable of gauging their own need for water.

24

Water Consumption Study



Temperature (X) to self-reported thirst (M) to water consumption (Y).

25

SAS Program

```
proc reg;
model y=x;
model y=x m;
model m=x;
See handout for output.
```

26

SPSS Program

```
regression
   /variables x y m
   /dependent=y
   /enter=x.
regression
   /variables x y m
   /dependent=y
   /enter=x m.
regression
   /variables x y m
   /dependent=m
   /enter x.
See handout for output
```

27

Estimates of *a*, *b*, *c*, and *c’*

(1) Temperature (X) was significantly related to water consumption (Y) ($\hat{c}$=.3604, $s_{\hat{c}}$=.1343, $t\hat{c}$ = 2.683).

(2) Temperature was significantly related to self-reported thirst (M) ($\hat{a}$=.3386, $s_{\hat{a}}$=.1224, $t\hat{a}$=2.767).

(3) Self-reported thirst was significantly related to water consumption controlling for temperature ($\hat{b}$=.4510, $s_{\hat{b}}$=.1460, $t\hat{b}$=3.090).

-The adjusted effect of temperature was not statistically significant, ($\hat{c}’$=.2076, $s_{\hat{c}’}$=.1333, $t\hat{c}’$=1.558) and there was a drop to $\hat{c}’$ = .2076 from $\hat{c}$=.3604.

28

Mediation Models for Water Consumption Data

$$Y = \hat{i}_1 + \hat{c}\,X$$
$$Y = -22.0505 + \underset{(.1343)}{.3604}\,X$$

$$Y = \hat{i}_2 + \hat{c}'\,X + \hat{b}\,M$$
$$Y = -12.7129 + \underset{(.1333)}{.2076}\,X + \underset{(.1460)}{.4510}\,M$$

$$M = \hat{i}_3 + \hat{a}\,X$$
$$M = -20.7024 + \underset{(.1224)}{.3386}\,X$$

29

Mediated Effect Measures

Mediated effect

$\hat{a}\hat{b}$= (.3386) (.4510) = $\hat{c} - \hat{c}'$ =.3604-.2076 =.1527

Standard error = $s_{First} = \sqrt{\hat{a}^2 s_{\hat{b}}^2 + \hat{b}^2 s_{\hat{a}}^2}$

Standard error = $\sqrt{.3386^2(.1460)^2 + .4510^2(.1224)^2} = .0741$

30

Second Order Standard Error

$$\hat{a}\hat{b} = (.3386)(.4510) = \hat{c} - \hat{c}' = (.3604) - (.2076) = .1527$$

$$s_{Second} = \sqrt{\hat{a}^2 s_{\hat{b}}^2 + \hat{b}^2 s_{\hat{a}}^2 + s_{\hat{b}}^2 s_{\hat{a}}^2}$$

$$s_{\hat{a}\hat{b}Second} = \sqrt{.3386^2(.1460)^2 + .4510^2(.1224)^2 + (.1224)^2(.1460)^2} = .0762$$

31

Confidence Intervals for the Mediated Effect First Order

- Confidence intervals are advocated by researchers for several reasons: effect size, range of possible values, not just null hypothesis binary significance testing. For 95% confidence intervals:
- Upper Confidence Interval (UCL) $= \hat{a}\hat{b} + z_{.975}\, s_{\hat{a}\hat{b}}$
- Lower Confidence Interval (LCL) $= \hat{a}\hat{b} + z_{.025}\, s_{\hat{a}\hat{b}}$
- For water consumption data.
- UCL = .1527 + (1.96)(.0741) = .2979
- LCL = .1527 + (-1.96) (.0741) =.0075
- 95% Confidence Interval from .0075 to .2979. The effect is statistically significant because 0 is not in the interval.

32

Confidence Intervals for the Mediated Effect Second Order

- Confidence intervals are advocated by researchers for several reasons: effect size, range of possible values, not just null hypothesis binary significance testing. For 95% confidence intervals:
- Upper Confidence Interval (UCL) $= \hat{a}\hat{b} + z_{.975}\, s_{\hat{a}\hat{b}}$
- Lower Confidence Interval (LCL) $= \hat{a}\hat{b} + z_{.025}\, s_{\hat{a}\hat{b}}$
- For water consumption data.
  UCL = .1527 + (1.96)(.0762) = .3021
  LCL = .1527 + (-1.96) (.0762) =.0033
- 95% Confidence Interval from .0033 to .3021. The effect is statistically significant because 0 is not in the interval.

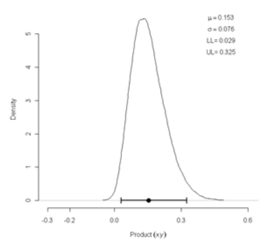
33

Example Calculations using the Distribution of the Product

- For example, $\hat{a} = .3386$, $s_{\hat{a}} = .1224$, $\hat{b} = .4510$, $s_{\hat{b}} = .1460$. Enter these values in the PRODCLIN program.
- PRODCLIN uses the critical values for the 2.5% percentile, M_{lower} =-1.6175 and M_{upper} = 2.2540 the critical value for the 97.5% percentile.
- Use the critical values to calculate upper and lower confidence limits.
- LCL$= \hat{a}\hat{b} + M_{upper}\, s_{\hat{a}\hat{b}}$ = .1527 +(-1.6175) (.0741)
  UCL$= \hat{a}\hat{b} + M_{lower}\, s_{\hat{a}\hat{b}}$ = .1527 + (2.2540)(.0741)
- Asymmetric Confidence Limits are (.0329, .3197) and (.0294, .3245) from new PRODCLIN.

34

Plot and Confidence Limits from RMediation (Chapter 3 data)



35

More Examples

- Was there a significant relation of X to M?
- Was there a significant relation of M to Y adjusted for X?
- Is the mediated effect statistically significant?
- Word Experiment
- PHLAME data
- Fit.txt data

36