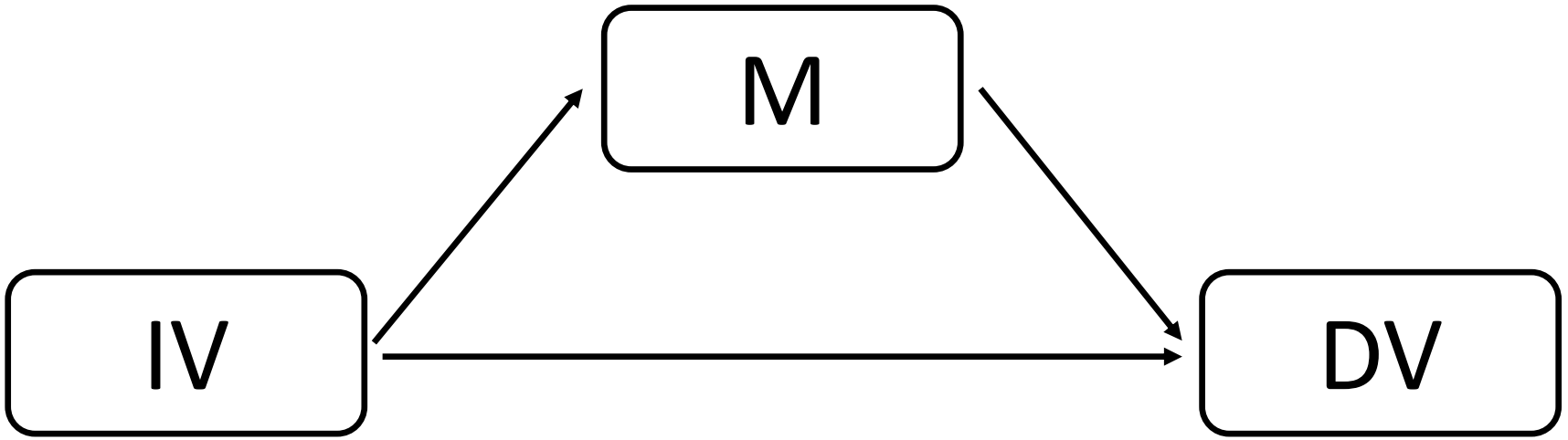
Mediation Analysis in R



Notes:
(1) This PDF is part of YouTube tutorials. This PDF is for individual, personal usage only. Method 1: https://youtu.be/p2XbncjiA6k Method 2: https://youtu.be/MIIF-ICF52Y
(2) The author accepts no responsibility for the topicality, correctness, completeness or quality of the information provided.

Content of This Video

- Method 1:
  - Basic idea of mediation analysis
    - Baron & Kenny (1986) 3 regressions
  - Use mediate() in mediation package for 95% CI
- Method 2:
  - Bootstrapping mediation from scratch
    - Write our own mediation function

Mediation Analysis in R Method - 1

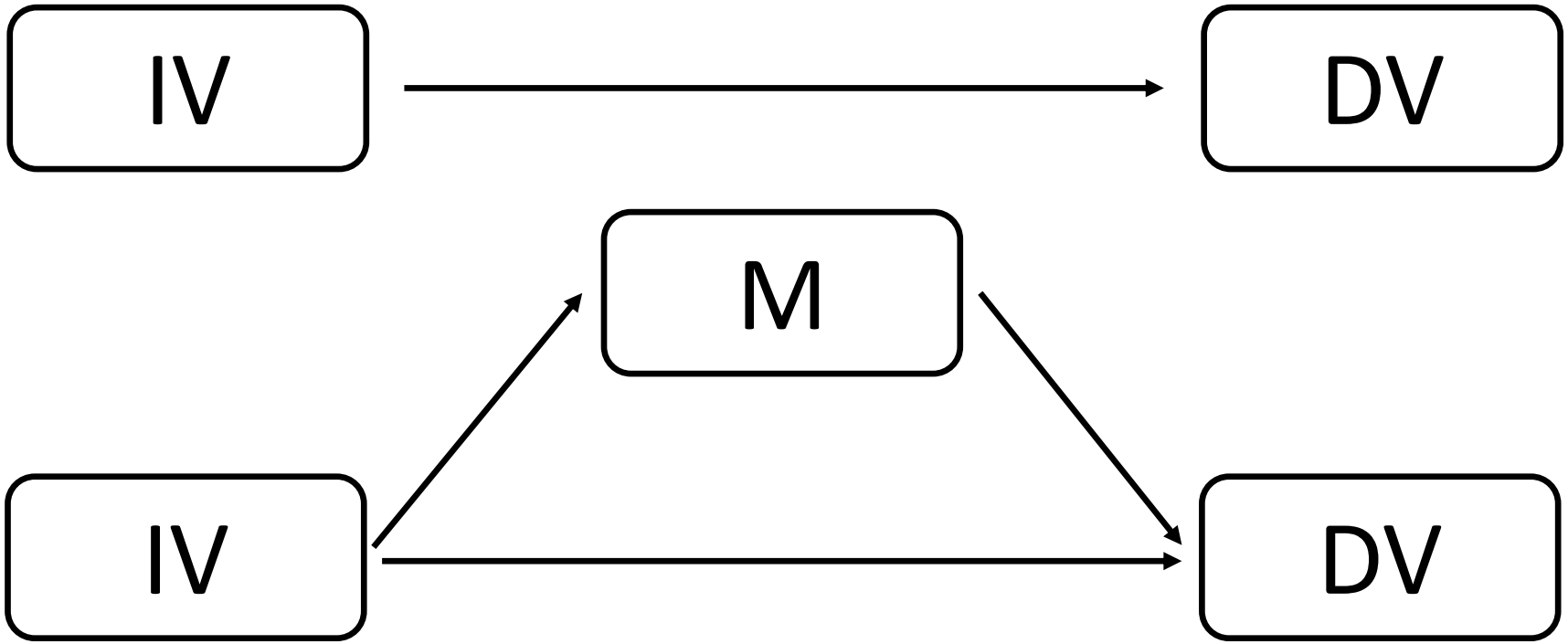
Notes:

(1) This PDF is part of YouTube tutorials. This PDF is for individual, personal usage only. Method 1: https://youtu.be/p2XbncjiA6k Method 2: https://youtu.be/MIIF-ICF52Y

(2) The author accepts no responsibility for the topicality, correctness, completeness or quality of the information provided.

Definition of Mediation

- Mediation:
  - An observed relationship between an independent variable (IV, or X) and a dependent variable (DV, or Y).
  - The name of the mechanism is called mediator (M).

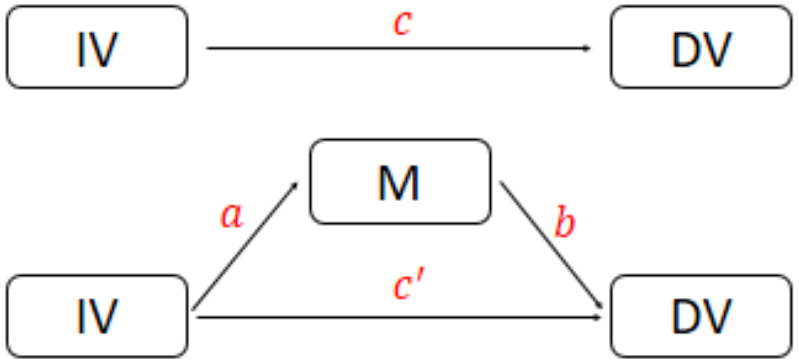


Notes:

Conditions of Mediation

- Based on Baron & Kenny (1986), there are three sets of regression:
- (1) X → Y (c needs to be significant, generally speaking.)

$$Y = cX$$



- (2) X → M (a needs to be significant.)

$$M = aX$$

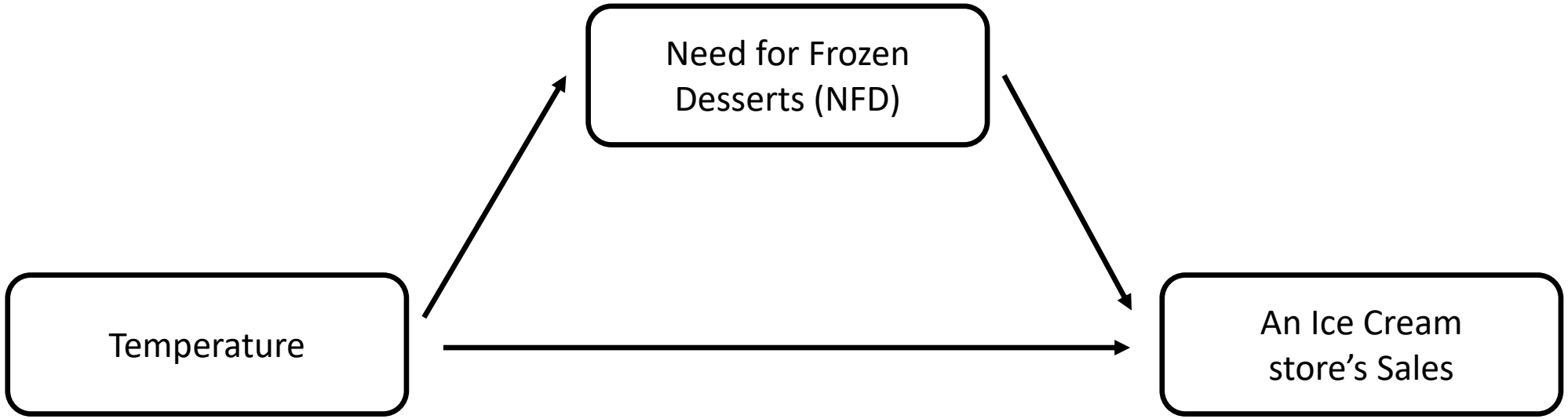
- (3) X + M → Y (b needs to be significant.)

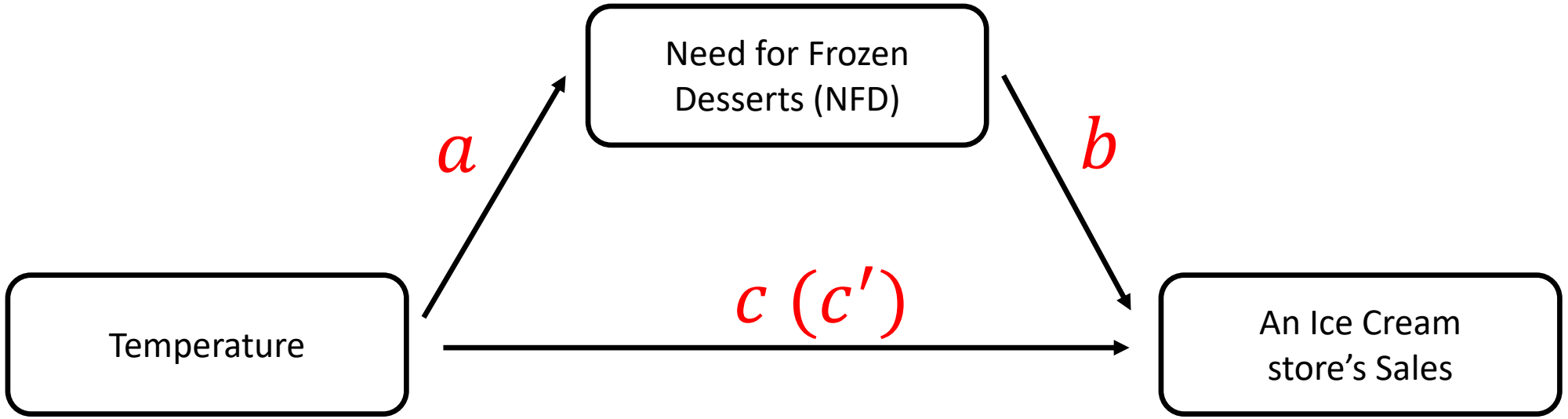
$$Y = c'X + bM$$

Notes:
(1) This PDF is part of YouTube tutorials. This PDF is for individual, personal usage only. Method 1: https://youtu.be/p2XbncjiA6k Method 2: https://youtu.be/MIIF-ICF52Y
(2) The author accepts no responsibility for the topicality, correctness, completeness or quality of the information provided.

Example of Mediation

- The following is a hypothetical study.
  - Higher temperatures increase an ice cream store's sales.
  - This is because higher temperatures make people want to have frozen desserts, making them more likely to buy ice cream from the store





- (1) X → Y (c needs to be significant, generally speaking.)

$$Y = cX$$

$$\text{Sales} = b_0 + c\ \text{Temperature}$$

- (2) X → M (a needs to be significant.)

$$M = a\ X$$

$$\text{NFD} = b_0 + a\ \text{Temperature}$$

- (3) X + M → Y (b needs to be significant.)

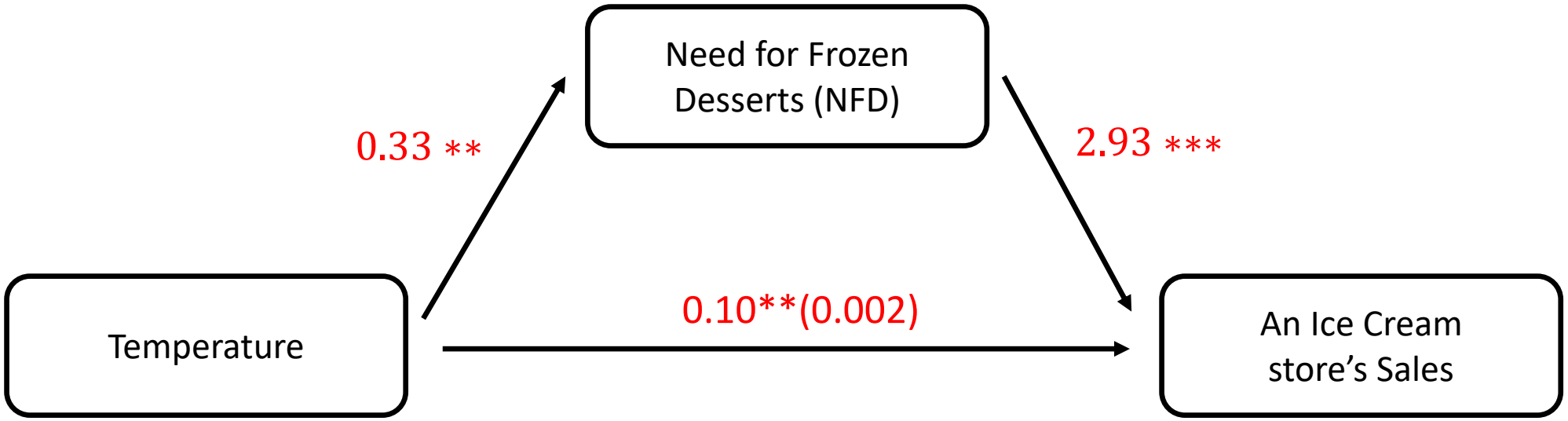
$$Y = c'X + b\ M$$

$$\text{Sales} = c'\text{Temperature} + b\ \text{NFD}$$

Hypothetical Data

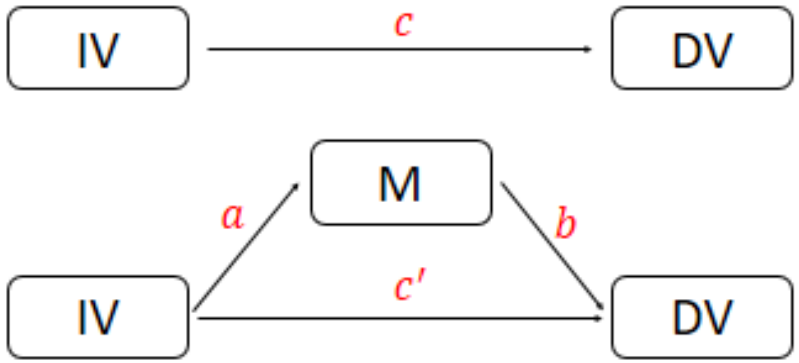
- Note: This data is generated via R programming (i.e., not real data). Please do not interpret the findings from a theoretical perspective.

- IV = Temperature
- DV = Sales
- Mediator = Need for Frozen Desserts (NFD)
  - How much do you want to have frozen desserts?
  - 0 = Not at all, 6 = Very much



* p <0.05; ** p<0.01; *** p< 0.001

Test Indirect effect a*b



- **Bootstrapping**
- Bootstrapping is a non-parametric method based on resampling with replacement which is done many times, e.g., 5000 times.
  - From each of these samples the indirect effect is computed and a sampling distribution can be empirically generated.

Mediation Analysis in R Method - 2

Notes:
(1) This PDF is part of YouTube tutorials. This PDF is for individual, personal usage only. Method 1: https://youtu.be/p2XbncjiA6k Method 2: https://youtu.be/MIIF-ICF52Y
(2) The author accepts no responsibility for the topicality, correctness, completeness or quality of the information provided.

Mediation Analysis in R from Scratch

Notes:

(1) This PDF is part of YouTube tutorials. This PDF is for individual, personal usage only. Method 1: https://youtu.be/p2XbncjiA6k Method 2: https://youtu.be/MIIF-ICF52Y

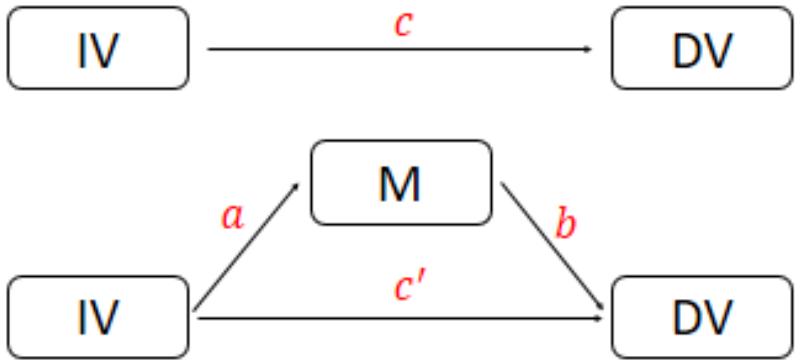
(2) The author accepts no responsibility for the topicality, correctness, completeness or quality of the information provided.

Topics in this video

- 1. What is bootstrapping?
- 2. How to write the R program?
- 3. Standard Error (SE) and Standard Deviation (SD) in bootstrapping
- 4. Bias in bootstrapping

What is bootstrapping?

Test Indirect effect a*b



- **Bootstrapping**
- Bootstrapping is a non-parametric method based on resampling with replacement which is done many times, e.g., 5000 times.
  - From each of these samples the indirect effect is computed and a sampling distribution can be empirically generated.

Use R to test Indirect effect a*b

Sample a data to calculate a and b path and indirect effect (i.e., a*b)

Use boot() to bootstrap it (e.g., 5000 times). Thus, 5000 indirect effects.

95% Confidence Interval for Indirect Effect

- Normal distribution assumption:
  - 95% CI =
  - (indirect effect –bias – 1.69*SE, indirect effect –bias + 1.69*SE)

- Percentile:
  - 95% CI =(2.5% pt, 97.5% pt)