Topic materials:
Dr Raquel Iniesta
Department of Biostatistics and Health Informatics

Narration and contribution:
Zahra Abdula

Improvements:
Nick Beckley-Hoelscher
Kim Goldsmith
Sabine Landau

Institute of Psychiatry, Psychology and Neuroscience

Module Title: Introduction to Statistics

Session Title: Path Diagrams

Topic title: Mediation

After working through this session you should be able to:

- To understand the concept of **mediation**
- To understand what a **path diagram** is and how to build it
- To understand the concept of **direct, indirect and total effect** and how to compute them

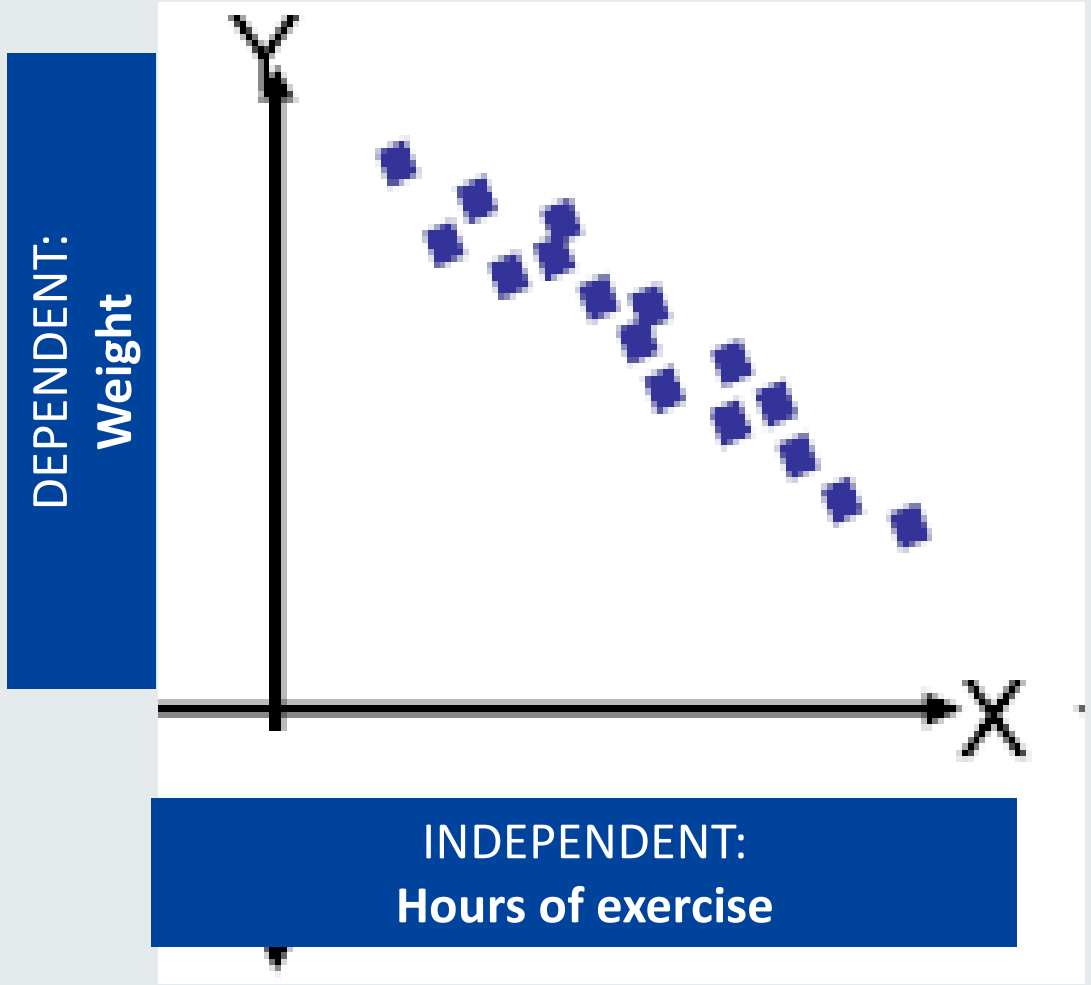
Previously on ‘Introduction to Statistics’

16 people were observed to see if the weight of a person, related to the hours of exercise they conducted. The following hypothesis was investigated:

Hypothesis ‘The higher the number of hours of exercise the lower the weight’.

Plotting the data is essential to understand and visually assess the relationship between pairs of continuous variables

The plot of data points (x,y) with x = **hours of exercise** and y = **weight** of a person where the data is continuous is called a **scatterplot.**



Previously on ‘Introduction to Statistics’

Questions:

Q1: How strong is the linear relationship? Understand the direction and magnitude of the linear relationship

A1: Correlation Coefficient (Pearson) ***r*=-0.85**

There is strong, negative, linear association between hours of exercise and weight loss (*r*=-0.85)

Q2: Can the relationship between variables be described by fitting a line to the observed data?

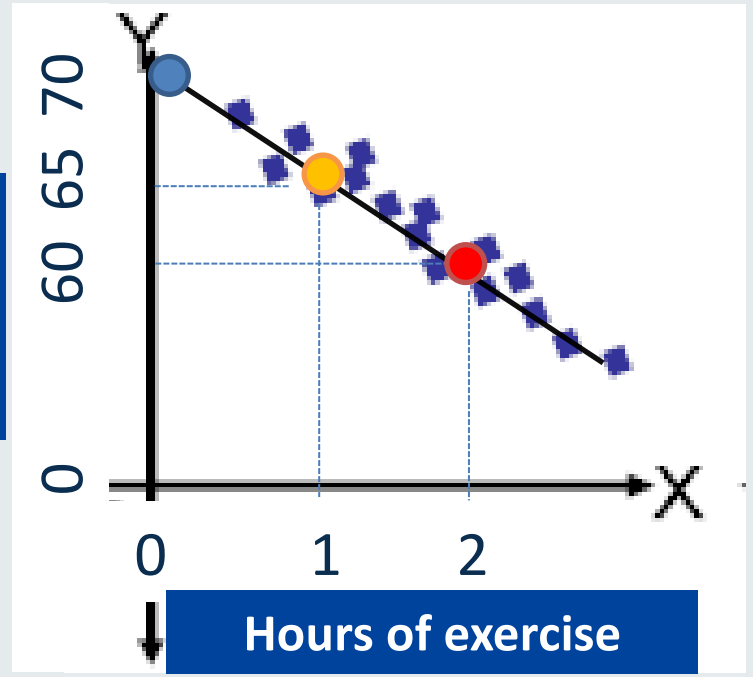
A2: Yes, because there is a **linear relationship**. The relationship is expressed as an equation

$$y = \beta_0 + \beta_1 x$$

where β_0 is the y intercept = 70
where β_1 is the slope of the line = -5

$$y = 70 - 5x$$

X	Y
0	70
1	65
2	60



β_0=70; β_1=-5;

Previously on ‘Introduction to Statistics’

Interpretation:

- β_0 = 70, When hours of exercise = 0, weight is 70kg.
- β_1 = -5, Each additional hour of exercise decreases weight by 5kg.

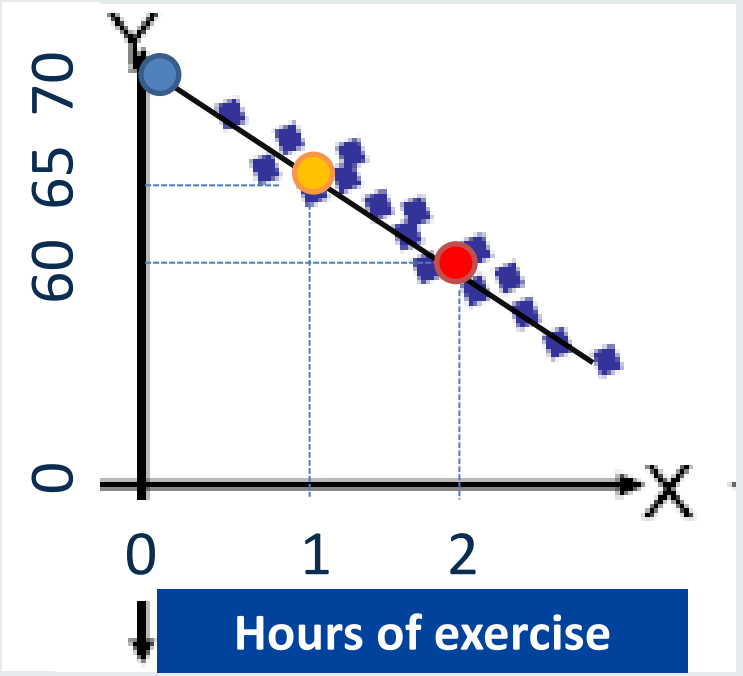
Linear regression model:

- To measure to what extent there is a linear relationship between two variables
- A rule that predicts weight given the hours of exercise.

β_0=70; β_1=-5;

$y = 70 - 5x$

X	Y
0	70
1	65
2	60



Previously on ‘Introduction to Statistics’

Simple linear regression

$$y = 70 - 5x + \varepsilon$$

Where: ***y*=weight; *x*=exercise;**

Multiple linear regression

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

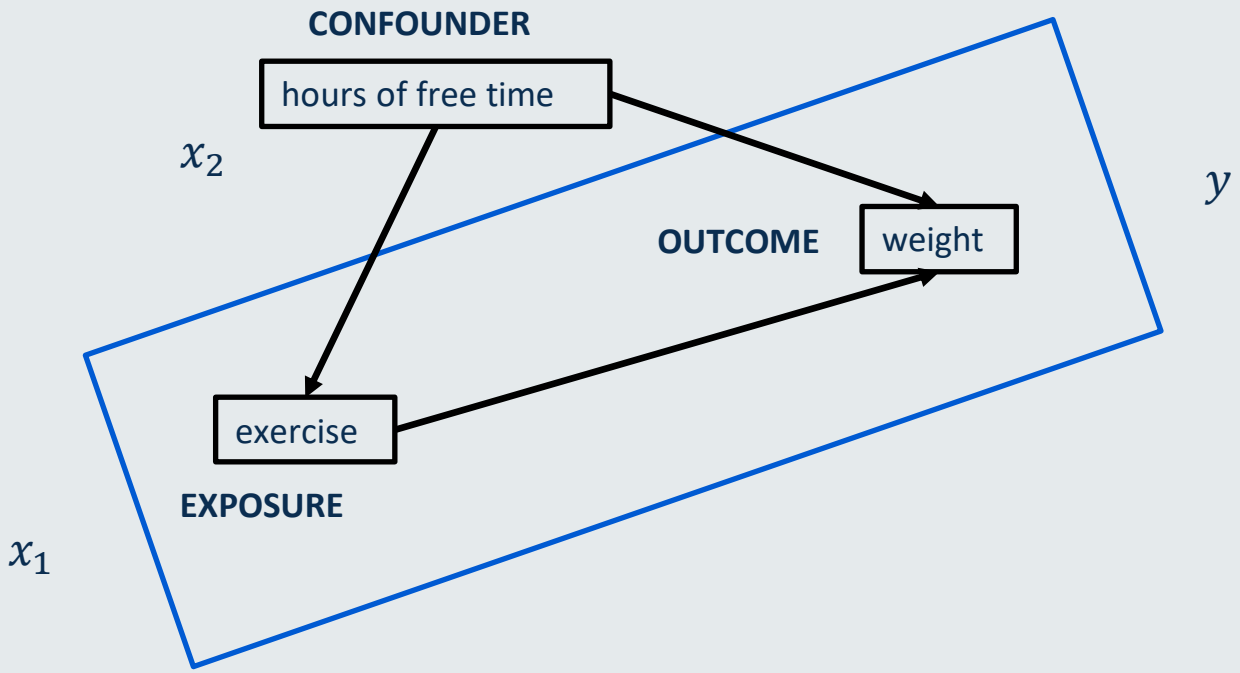
Where: ***y*=weight; x_1=exercise; x_2=diet;**

A **simple regression model** (one independent variable) fits a regression **line**
$y = \beta_0 + \beta_1 x_1$

A **multiple regression model** with two explanatory variables fits a **regression plane**
$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

Previously on ‘Introduction to Statistics’

x_2 is a **confounding variable** when it has an effect both on the dependent Y and independent x_1 variable.



Using multiple linear regression allows us to hold all other independent variables constant, allowing us to get an estimate of the effect of the independent variable of interest while adjusting for other variables in the model which are hypothesized to be confounders.

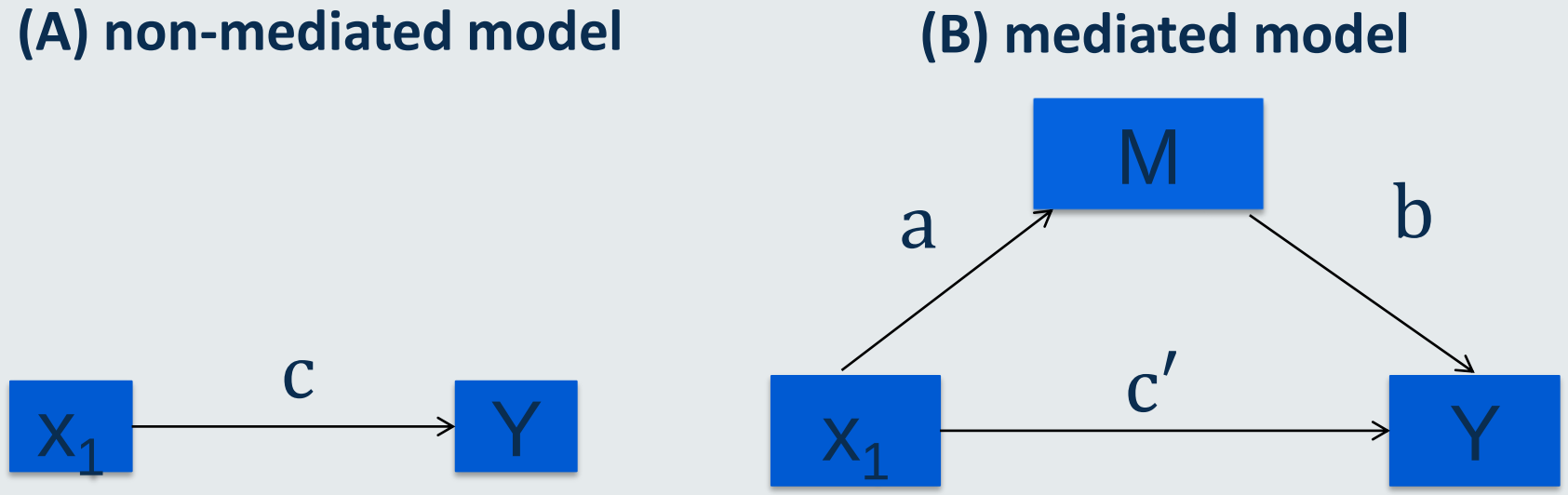
Mediation

The third variable x_2 can take a role of **mediator**. A mediator explains **a portion of the association** between Y and x1. When x_2 is a mediator will denote it "**M**".

Mediation is a hypothesised causal mechanism by which one variable affects another variable.

A **mediator** (**M**) of the causal effect of independent variable (x_1) on dependent variable (Y) is a variable x_2 on the causal pathway from x_1 to Y.

Mediation



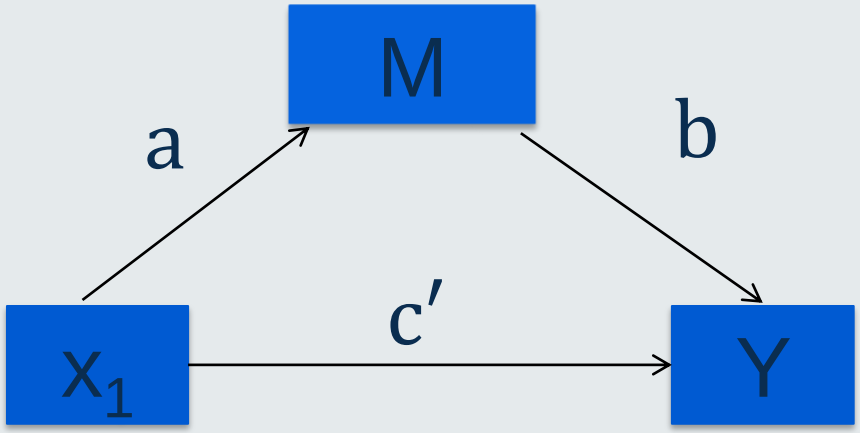
- In a **non-mediated model** (A), the **total effect** of the independent variable x_1 on the dependent Y is denoted by the path c

- Under a **mediated model** (B), the total causal effect c **can be split** into an indirect (or mediated) part with paths a and b and a direct (non-mediated) path c'

Mediation

(a) non-mediated model

(b) mediated model

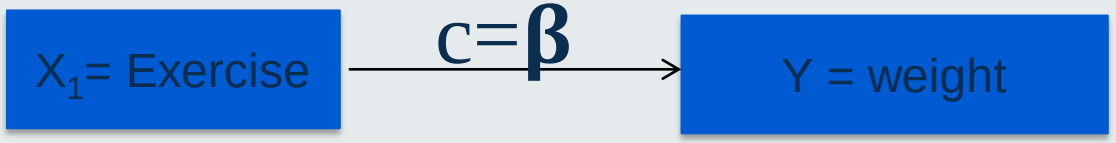


- **Direct** effect = c'
- **Indirect** effect (or "mediated" effect) = $a*b$
- c= **Total** effect = direct + indirect effect =c' + a * b

Investigating a Mediation Effect: Computing a, b and c

We want to look at the relationship between exercise and weight and consider the calories burned as a mediator of the exercise – weight relationship.

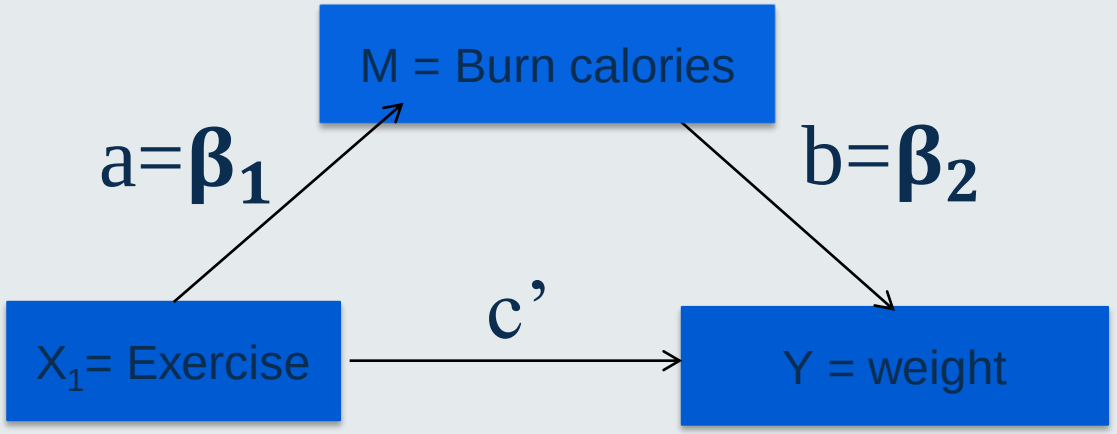
(a)



1. Estimate of **path c**:

$Y = \beta_0 + \beta X_1 + \varepsilon$

(b)



2. Estimate of **path a:**

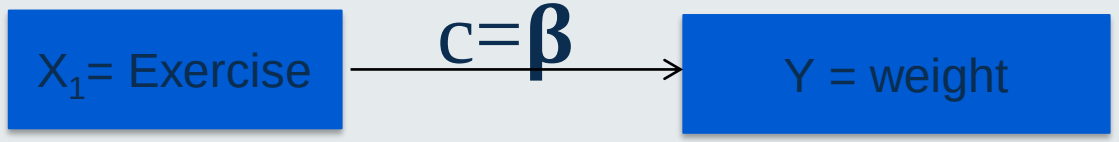
$\mathbf{M} = \beta_0 + \beta_1 X_1 + \varepsilon$

3. Estimate of **path b:**

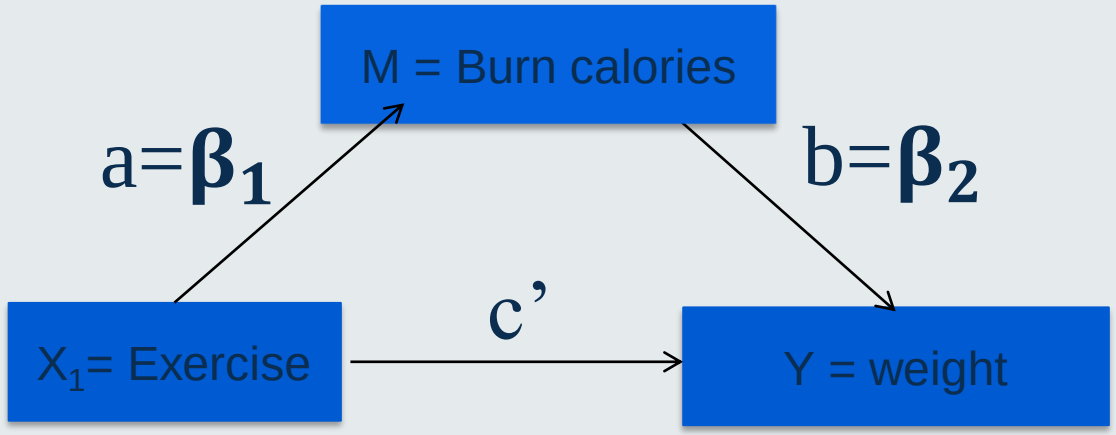
$Y = \beta_0 + \beta_2 M + \beta_3 X_1 + \varepsilon$

Investigating a Mediation Effect: Computing a, b and c

(a)



(b)



4: Estimate of **path c'**:
2 different ways:

i. $c = c' + a * b$

$\beta = c' + \beta_1 * \beta_2$

ii. From step 3 model:

$Y = \beta_0 + \beta_2 M + \beta_3 X_1 + \varepsilon$

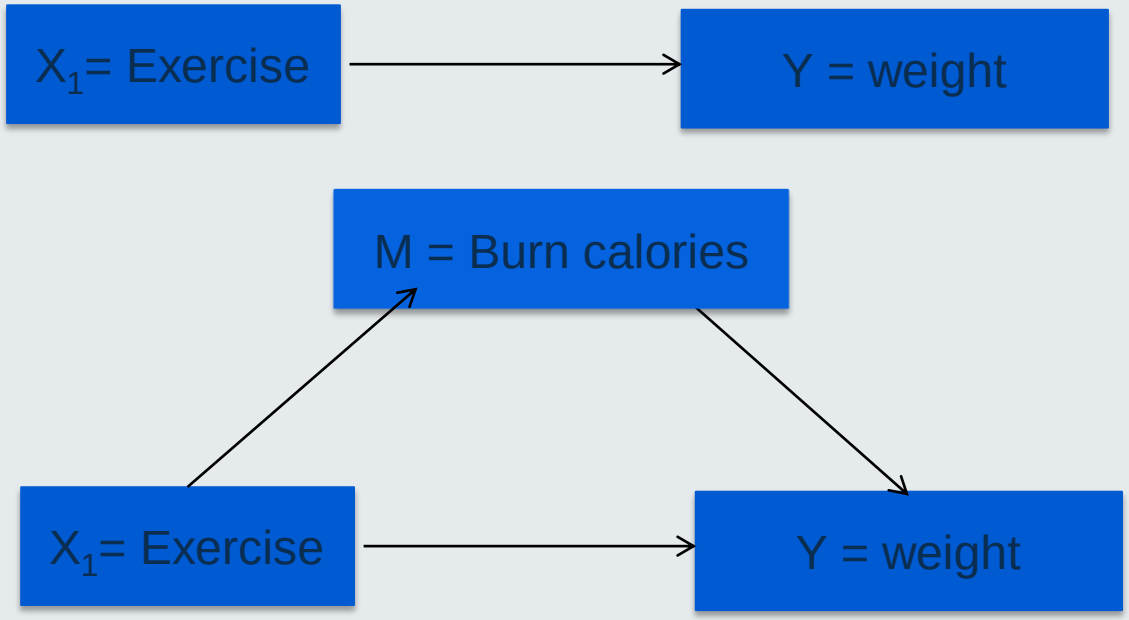
$c' = \beta_3$

Knowledge Check

Q1: In a mediated model, which of the next sentences MUST be TRUE?

a) The independent variable (X) causes the outcome variable (Y)
b) The independent variable (X) causes the mediator variable (M)
c) The mediator (M) causes the outcome variable (Y) when controlling for the independent variable (X).

Q2: Given the two path diagrams below and the set of models, compute a, b, c and c'



$Y = 70 - 5\,X_1 + \varepsilon$

$M = 0.5 + 2X_1 + \varepsilon$

$Y = 69 - 1.5\,M - 2\,X_1 + \varepsilon$

Knowledge Check Solutions

Q1: In a mediated model, which of the next sentences MUST be TRUE?

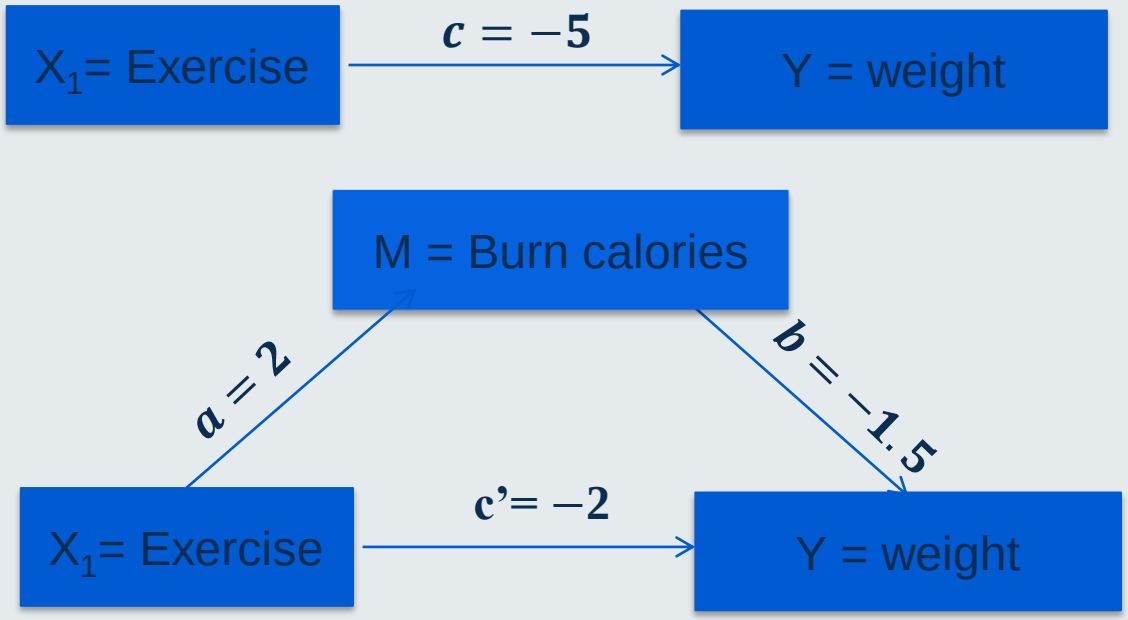
a) The independent variable (X) causes the outcome variable (Y)

b) The independent variable (X) causes the mediator variable (M)

c) The mediator (M) causes the outcome variable (Y) when controlling for the independent variable (X).

Answer: **c) must be TRUE.**

Q2: Given the two path diagrams below and the set of models, compute a, b, c and c'



$Y = 70 - \mathbf{5}\, X_1 + \varepsilon$

$M = 0.5 + \mathbf{2} X_1 + \varepsilon$

$Y = 69 - \mathbf{1.5}\ M - 2\, X_1 + \varepsilon$

$\beta = c' + \beta_1 * \beta_2$

$\mathbf{-5} = c' + \mathbf{2} * \mathbf{-1.5};$

$\mathbf{-5} = c' - 3;$

$c' = -2 = \beta_3$

Answer: a=2; b=-1.5; c=-5; c'=-2

References

MacKinnon, D. P., Fairchild, A. J. and Fritz, M.S (2007). Mediation analysis, Annual Review of Psychology, 58, 593–614

David Kenny's Website on mediation: http://davidakenny.net/cm/mediate.htm

Hayes, A .F. (2013). Introduction to Mediation, Moderation, and Conditional Process Analysis, Guildford Press.

MacKinnon, D. P., Fairchild, A. J. and Fritz, M.S (2007). Mediation analysis, Annual Review of Psychology, 58, 593–614

Thank you

Please contact your module leader or the course lecturer of your programme, or visit the module's forum for any questions you may have.

If you have comments on the materials (spotted typos or missing points) please contact Dr Iniesta:

Raquel Iniesta, PhD
Department of Biostatistics and Health Informatics
IoPPN, King's College London, SE5 8AF, London, UK
raquel.iniesta@kcl.ac.uk

For any other comments or remarks on the module structure, please contact one of the three module leaders of the Biostatistics and Health Informatics department:

Zahra Abdula: **zahra.abdulla@kcl.ac.uk**

Raquel Iniesta: **raquel.iniesta@kcl.ac.uk**

Silia Vitoratou: **silia.vitoratou@kcl.ac.uk**