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

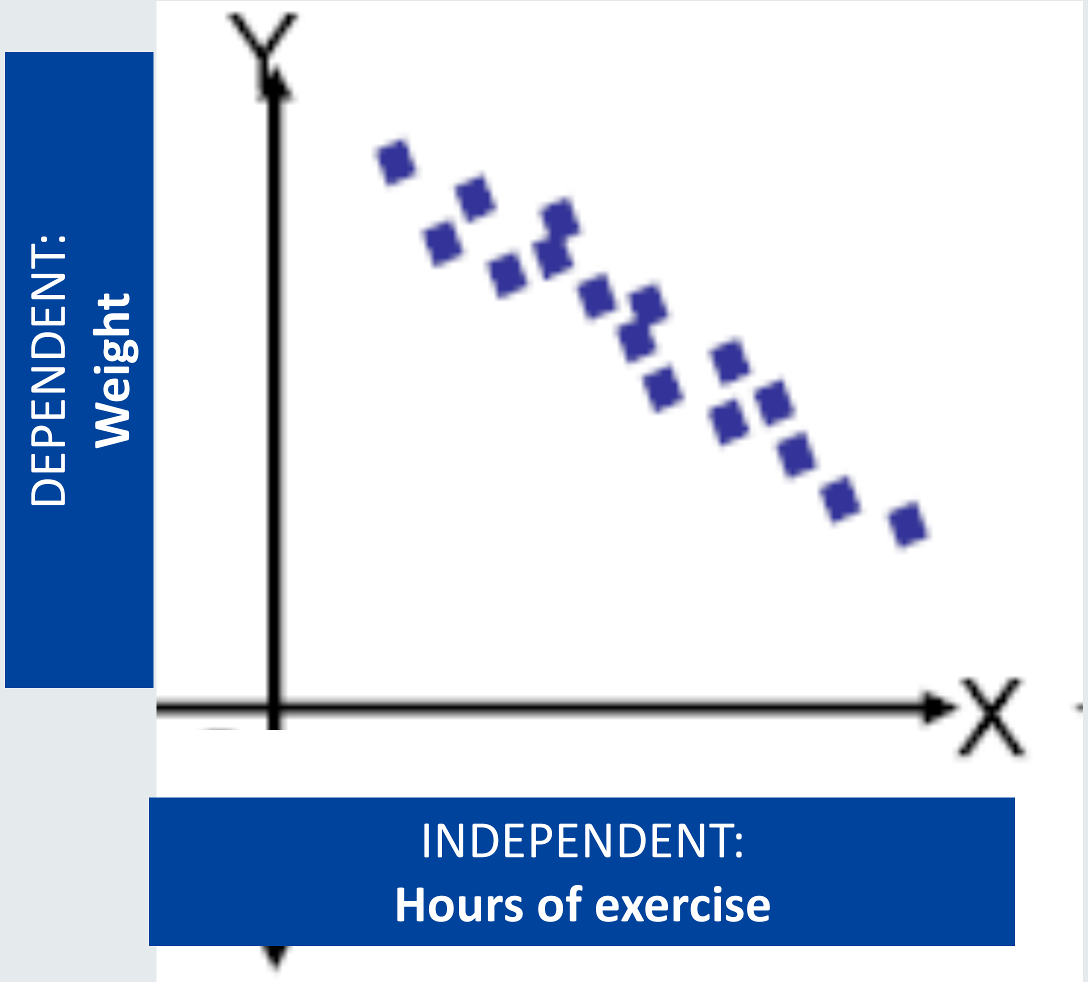
Previously on ‘Introduction to Statistics’

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'

<u>Questions:</u>

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?

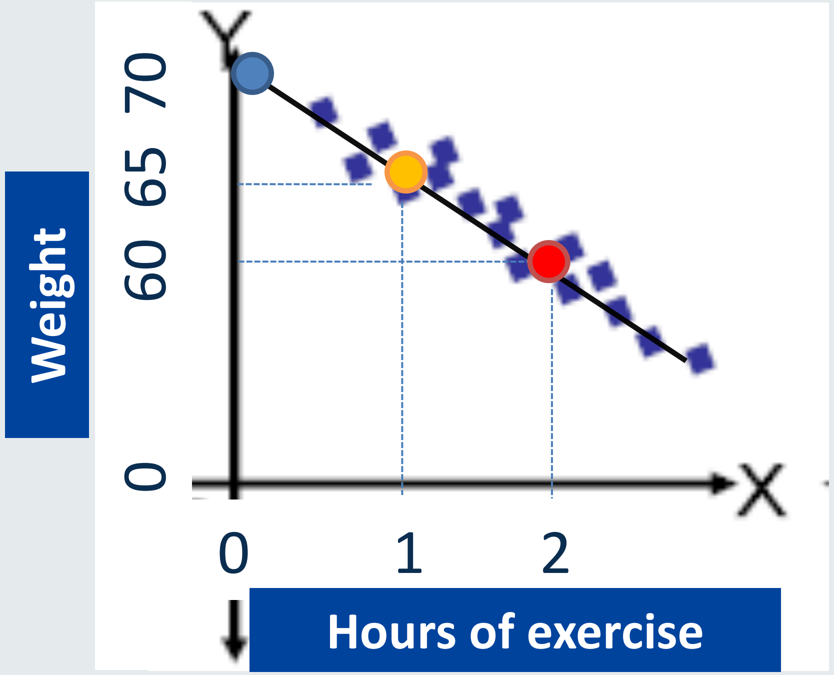
A2: Yes, because there is a **linear relationship**. The relationship is expressed as an equation

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

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.

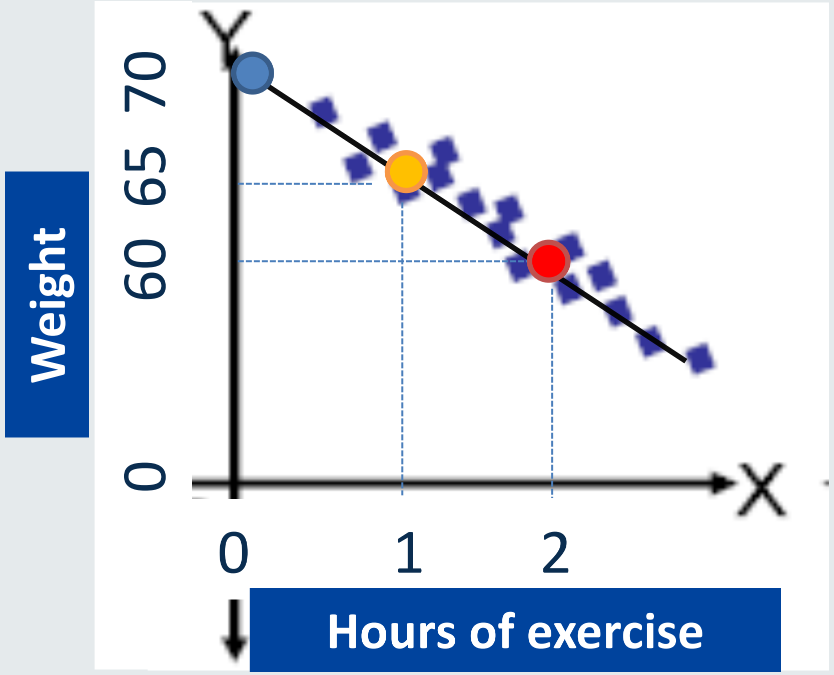
Linear regression model:

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

$$\boldsymbol{\beta_0}\textbf{=70; }\boldsymbol{\beta_1}\textbf{=-5;}$$

$$\boldsymbol{y = 70 - 5x}$$

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

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

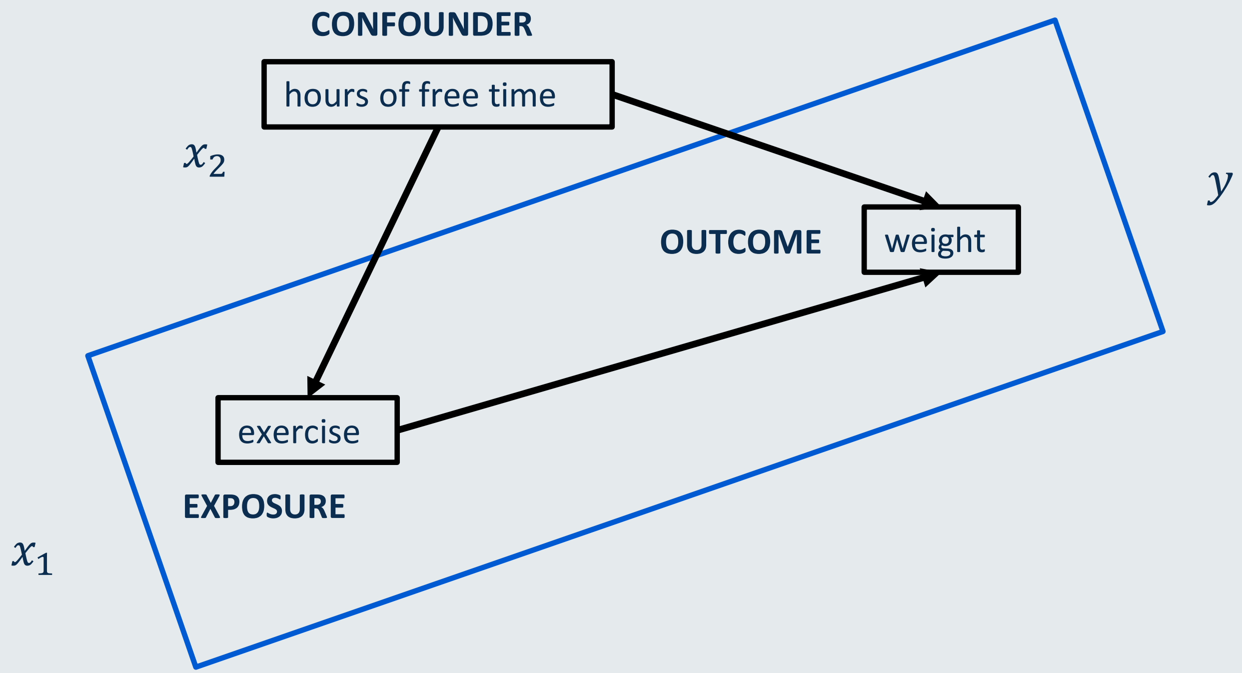
Where: ***y*=weight; x_1=exercise; x_2=diet;**

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

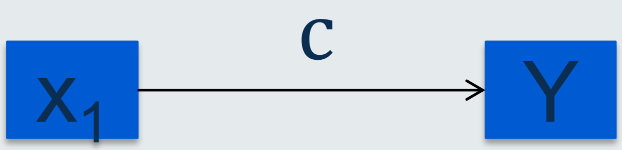
The third variable x_2 can take a role of **mediator**. A mediator explains **a portion of the association** between Y and x1. When x_2 is a mediator will denote it "**M**".

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

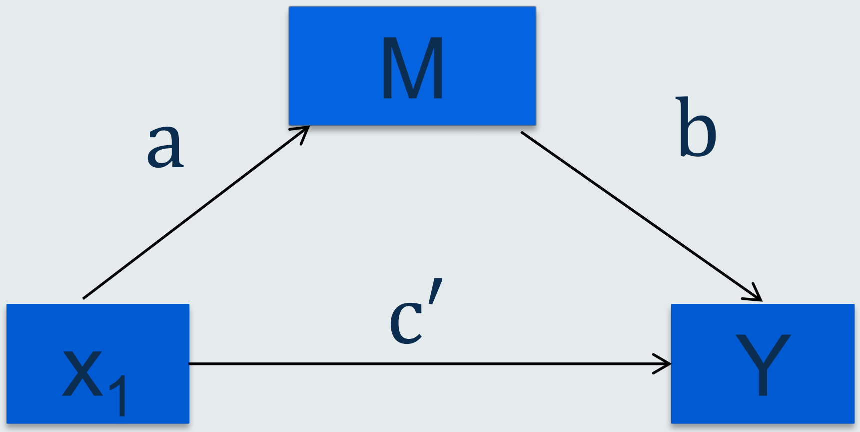
A **mediator** (**M**) of the causal effect of independent variable (x_1) on dependent variable (Y) is a variable x_2 on the causal pathway from x_1 to Y.

Mediation

(A) non-mediated model



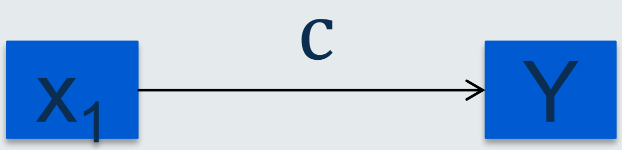
(B) mediated model



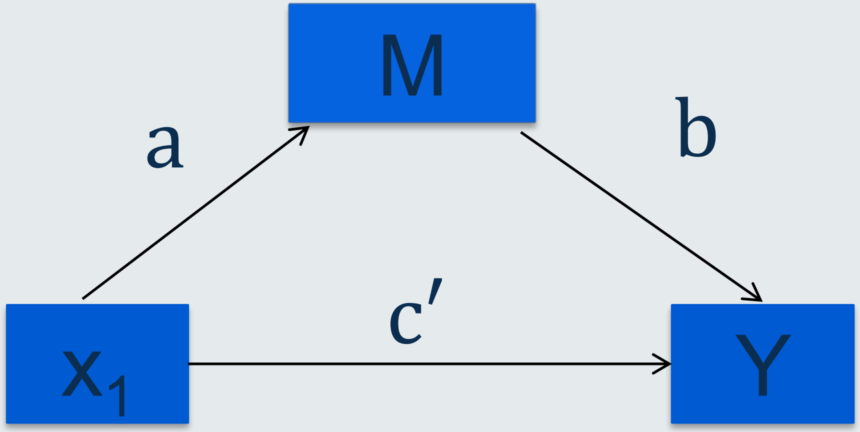
- In a **non-mediated model** (A), the **total effect** of the independent variable x_1 on the dependent Y is denoted by the path c
- Under a **mediated model** (B), the total causal effect c **can be split** into an indirect (or mediated) part with paths a and b and a direct (non-mediated) path c'

Mediation

(a) non-mediated model



(b) mediated model

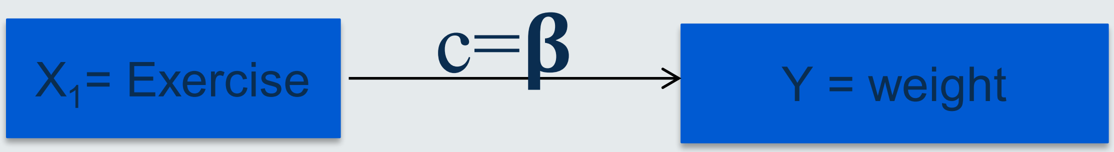


- **Direct** effect = c'
- **Indirect** effect (or "mediated" effect) = $a*b$
- c= **Total** effect = direct + indirect effect =$c' + a * b$

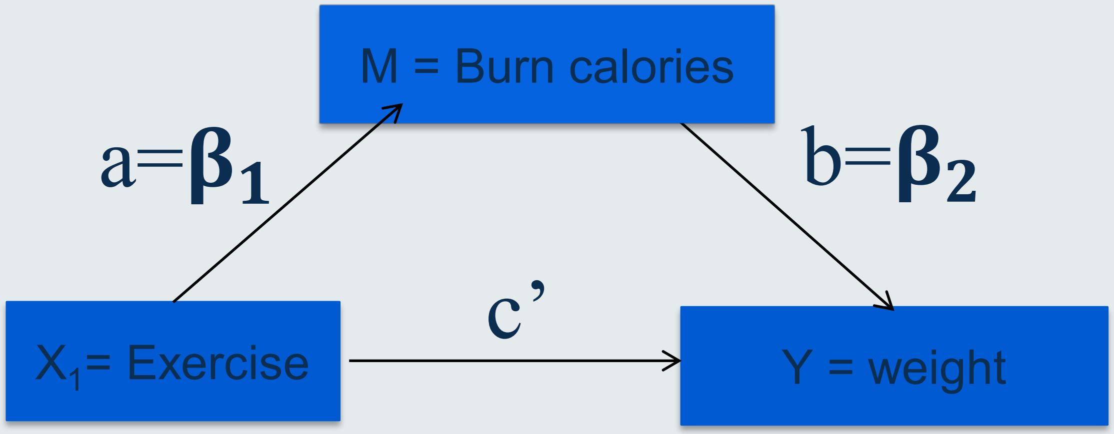
Investigating a Mediation Effect: Computing a, b and c

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)



(b)



1. Estimate of **path c**:

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

2. Estimate of **path a**:

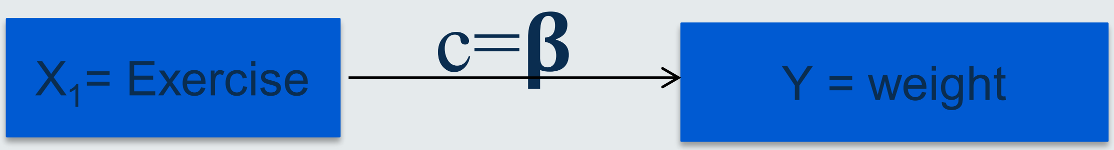
$M = \beta_0 + \beta_1 X_1 + \varepsilon$

3. Estimate of **path b**:

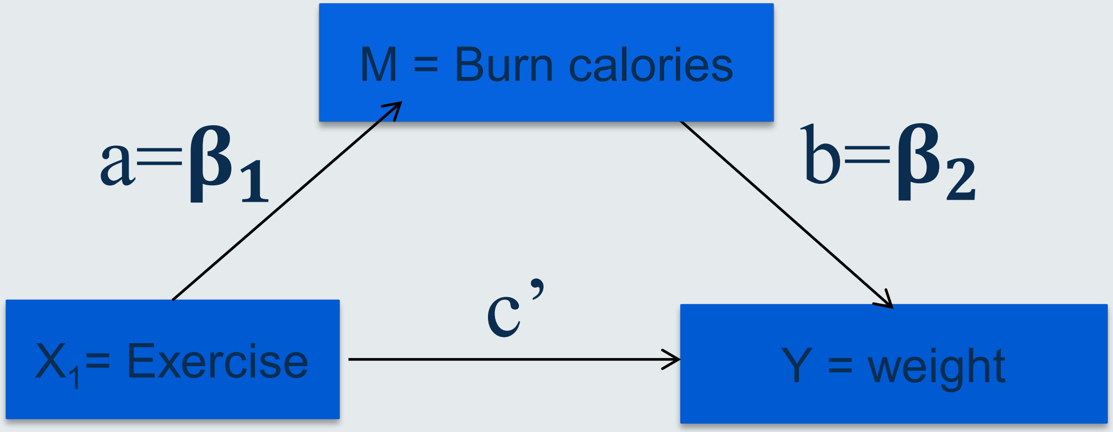
$Y = \beta_0 + \beta_2 M + \beta_3 X_1 + \varepsilon$

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

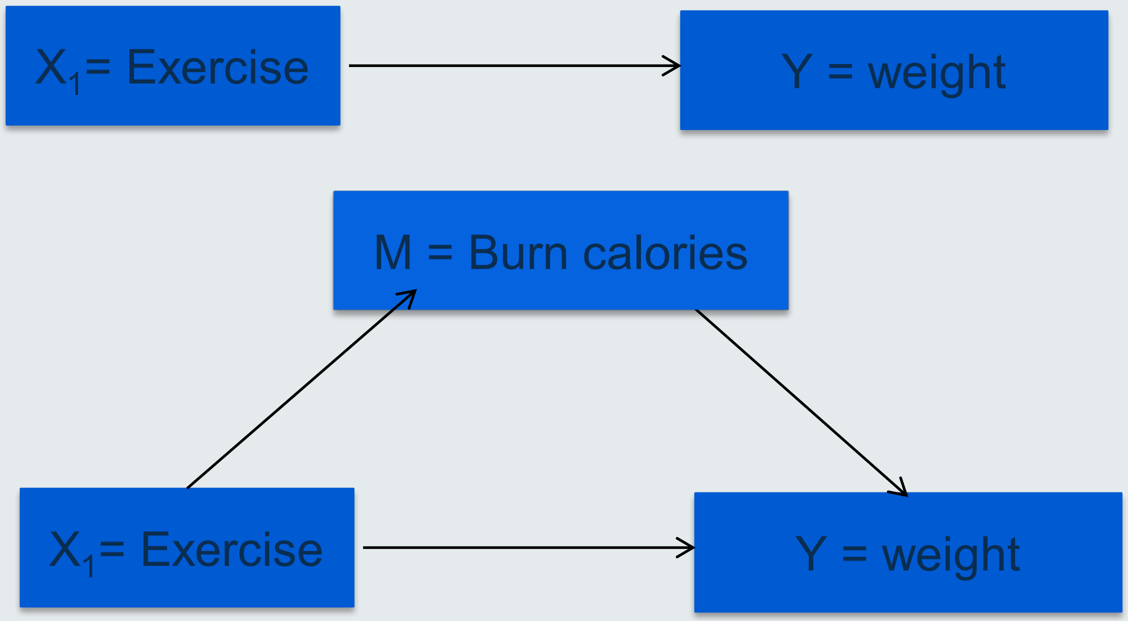
Q1: In a mediated model, which of the next sentences MUST be TRUE?

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?

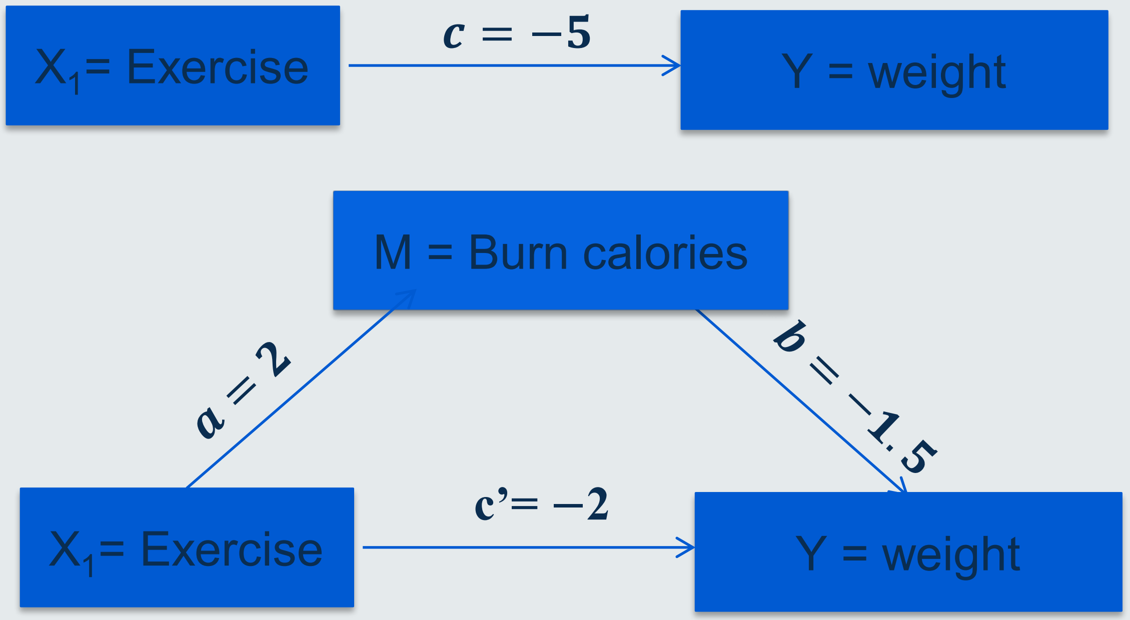
a) The independent variable (X) causes the outcome variable (Y)

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'



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

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

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

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

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

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

$c' = -2 = \beta_3$

<u>Answer</u>: 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 <u>your module leader</u> or <u>the course lecturer of your programme</u>, or visit the module's <u>forum</u> 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**

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: Baron and Kenny Steps

Topic title: Mediation

After working through this session you should be able to:

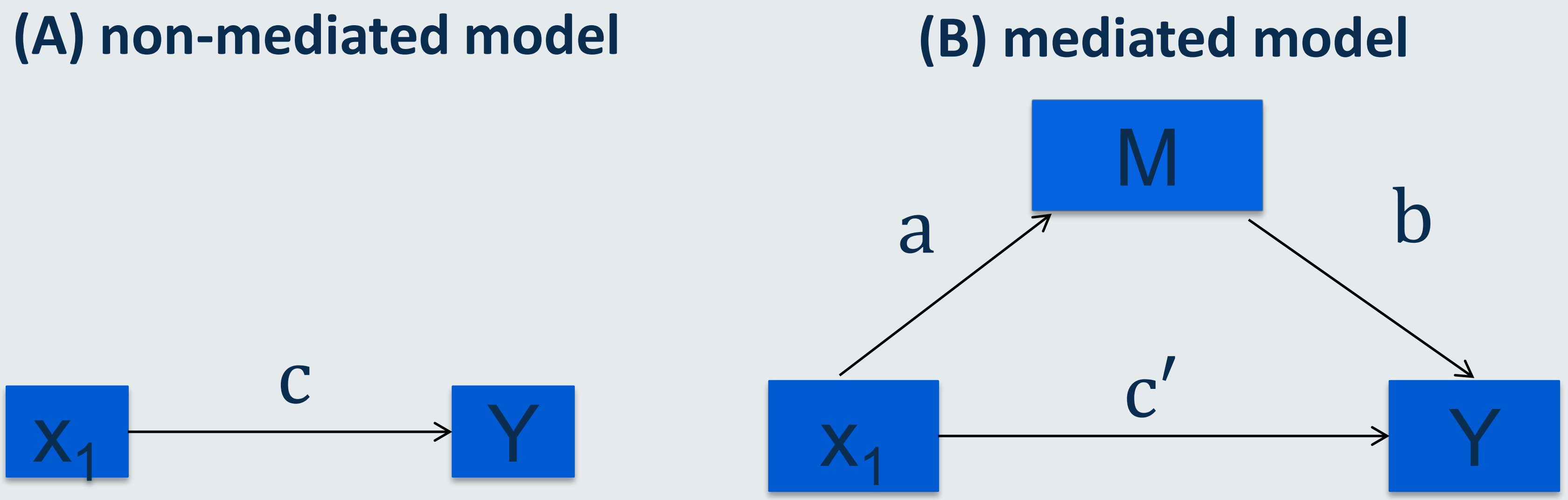
- To understand how to establish mediation using the **Baron and Kenny steps**
- To understand the difference between **partial** and **complete mediation**

Previously on 'Introduction to Statistics'

Before, we focused on the 3 variables (Y, X_1 and X_2) case.

We discussed the different roles that a third variable $\mathbf{X_2}$ can have while investigating the association between an independent X_1 and a dependent variable Y.
X_2 could be a **confounder** (C) or a **mediator** (M)

In the last session, we learnt what a mediator is, and how to estimate a, b, c and c' using simple and multiple linear regression models.



a

Testing a Mediation Effect

How to **test** ("establish") that there is a **mediation effect**?

Baron and Kenny (1986) discussed **four steps** to establish mediation:

Step 1: Test path c ($X_1 \rightarrow Y$): $Y = \beta_0 + \beta X_1 + \varepsilon$

- Establish that the causal variable (X_1) is associated with the dependent (Y).
- Test β : $\begin{cases} H_0: \beta = 0 \\ H_1: \beta \neq 0 \end{cases}$

Testing a Mediation Effect

How to **test** ("establish") that there is a **mediation effect**?

Baron and Kenny (1986) discussed **four steps** to establish mediation:

Step 2: Test path a ($X_1 \rightarrow M$): $M = \beta_0 + \beta_1 X_1 + \varepsilon$

- Show that the causal variable X_1 is associated with the mediator (M).
- Test β_1 : $\begin{cases} H_0: \beta_1 = 0 \\ H_1: \beta_1 \neq 0 \end{cases}$

Testing a Mediation Effect

How to **test** ("establish") that there is a **mediation effect**?

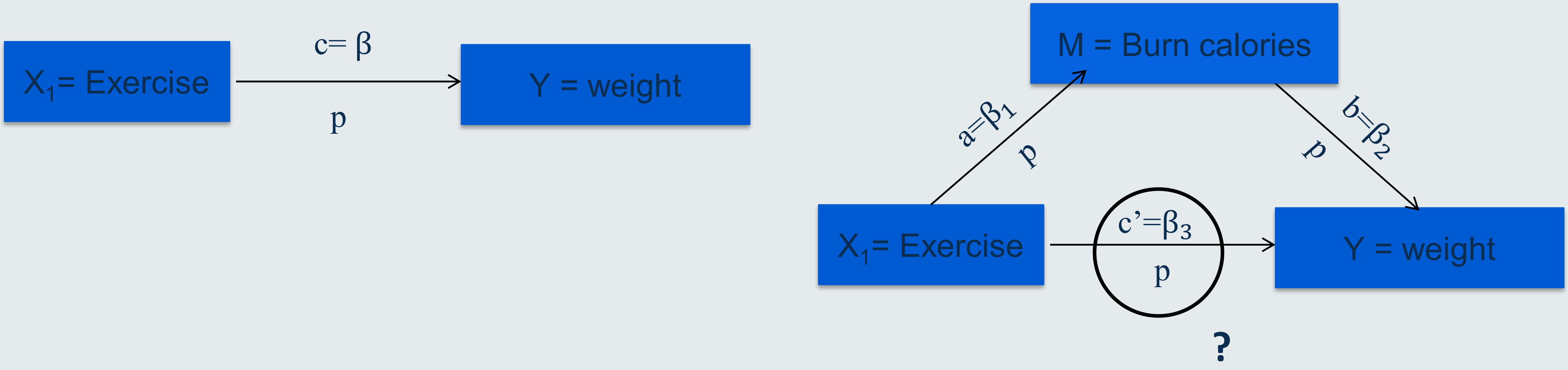
Baron and Kenny (1986) discussed **four steps** to establish mediation:

Step 3: Test path b ($M \rightarrow Y$, controlling for X_1): $Y = \beta_0 + \beta_2 M + \beta_3 X_1 + \varepsilon$

- Show that the mediator M is associated with the dependent Y, adjusting for the causal variable X_1.
- Test β_2 : $\begin{cases} H_0: \beta_2 = 0 \\ H_1: \beta_2 \neq 0 \end{cases}$

Testing a Mediation Effect

P values resulting from each step (1-3) are added to the arrows in the path diagram:



Baron and Kenny Extra Step

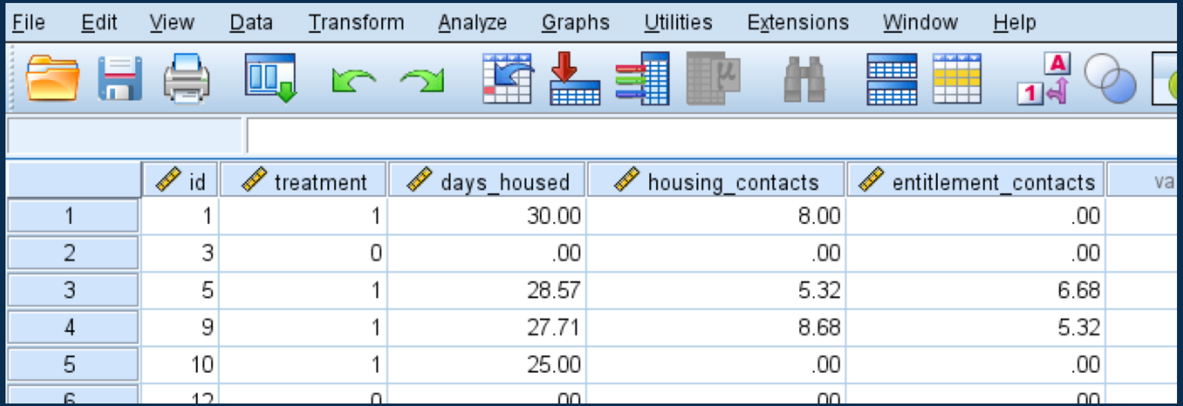
Step 4: **Test path c′ ($X_1 \rightarrow Y$, controlling for M):** $Y = \beta_0 + \beta_2 M + \beta_3 X_1 + \varepsilon$;

$$\begin{cases} H_0: \beta_3 = 0 \\ H_1: \beta_3 \neq 0 \end{cases}$$

- If β_3 is not significantly different from 0 ($p>0.05$) there is **complete mediation.**
  - There is no association between X_1 and Y, when we control for M.
  - This will be the case if the direct effect (path c′) drops to zero after controlling for M.
- If β_3 is significantly different from 0 ($p<0.05$) there is **partial mediation**.
  - c′ is smaller than c (in absolute value).
  - There is association between X_1 and Y when we control for M.

SPSS Slide

Download the data that we are going to use during the lecture. The dataset is the **lecture_8_data.sav**.



	id	treatment	days_housed	housing_contacts	entitlement_contacts
1	1	1	30.00	8.00	.00
2	3	0	.00	.00	.00
3	5	1	28.57	5.32	6.68
4	9	1	27.71	8.68	5.32
5	10	1	25.00	.00	.00

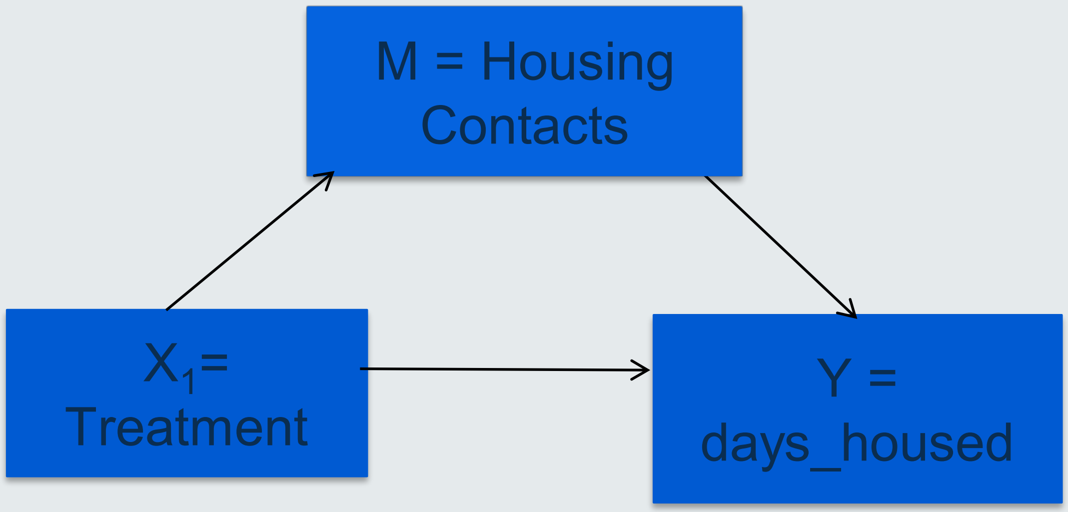
The dataset contains data from 109 subjects, measuring days in stable housing after receiving continuous treatment programme versus the standard treatment. Collecting information in respect to

- **Treatment: '1' = received continuous support from their assigned team, '0' = control (received standard treatment)**
- **Days_housed** : Average number of days per month in stable housing
- **Housing_contacts**: Average number of days per month that the respondent was in contact with their assigned housing services team
- **Entitlements contacts**: Average contact regarding eight specific services

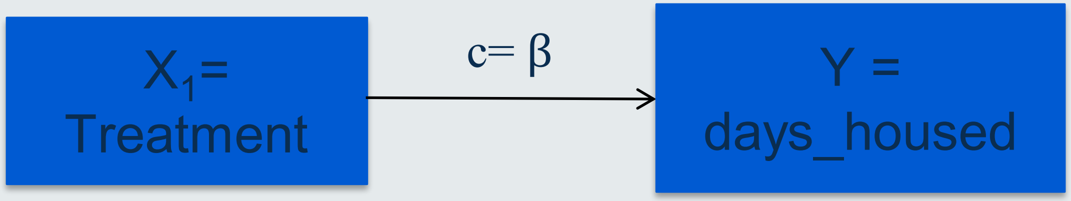
Example: Effect of Treatment on Stable Housing

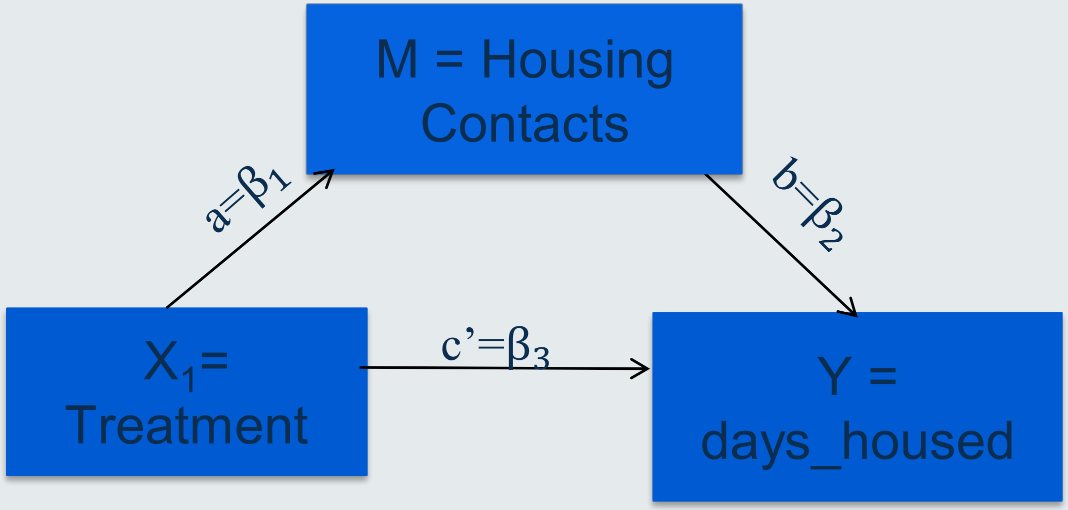
Homeless people are more likely to have serious mental illness. Morse et al. (1994) found that a treatment program which gives continuous support can be effective in increasing the average days they spend in stable housing (‘Treatment’ → ‘days_housed’).
The study also looked at the treatment effects on ‘contact for housing’ Average number of days per month that the respondent was in contact with their assigned treatment programme (‘Treatment’ → ‘housing_contacts’)
Does housing contacts mediate the treatment effect?

Path diagram:



Testing Mediation: Does Housing Contacts Mediate the Treatment Effect?





Step 1: **path c**

X_1 is associated with Y

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

$$\begin{cases} H_0: \beta = 0 \\ H_1: \beta \neq 0 \end{cases}$$

Step 2: **path a**

X_1 is associated with M

$$M = \beta_0 + \beta_1 X_1 + \varepsilon$$

$$\begin{cases} H_0: \beta_1 = 0 \\ H_1: \beta_1 \neq 0 \end{cases}$$

Step 3: **path b**

M is associated with Y, regardless X_1

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

$$\begin{cases} H_0: \boldsymbol{\beta_2} = 0 \\ H_1: \boldsymbol{\beta_2} \neq 0 \end{cases}$$

Step 4: **path c'**

X_1 is associated with Y, regardless M

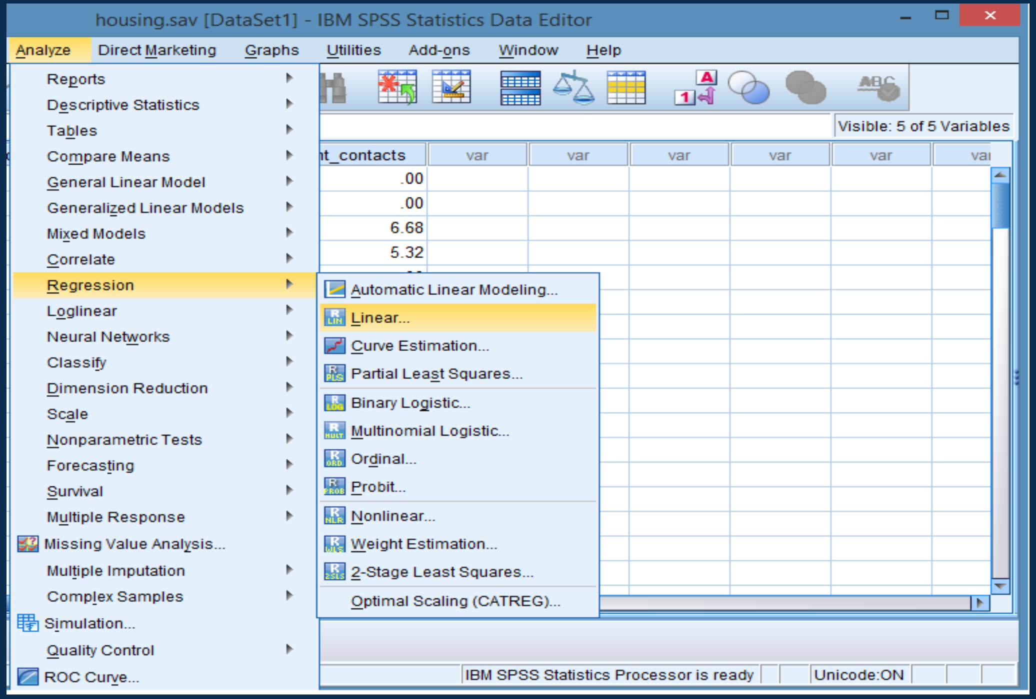
$$Y = \beta_0 + \beta_2 M + \beta_3 X_1 + \varepsilon$$

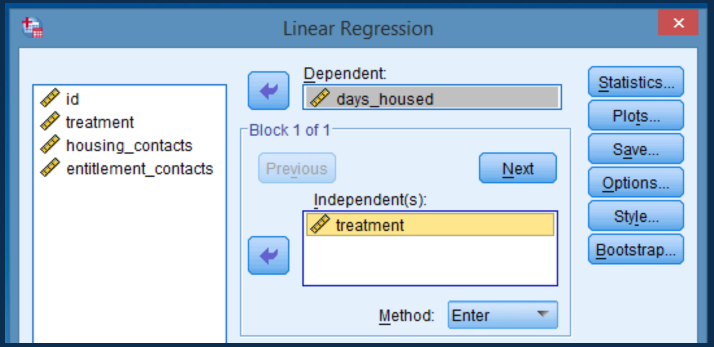
$$\begin{cases} H_0: \boldsymbol{\beta_3} = 0 \\ H_1: \boldsymbol{\beta_3} \neq 0 \end{cases}$$

SPSS Slide: ‘How to’ Steps

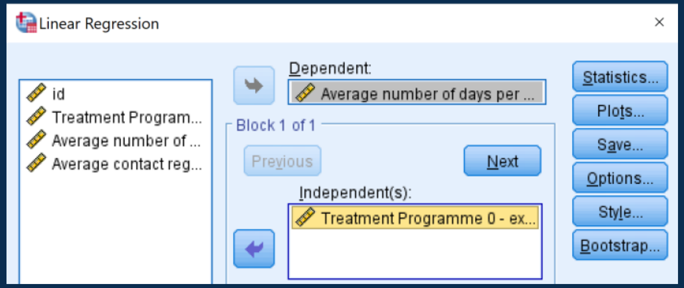
Computing three linear regression models from ‘housing.sav’ data:

1) Use ‘Analyse’ -> ‘Regression’ -> ‘Linear’
2) Drag and drop dependent, and independent variables.

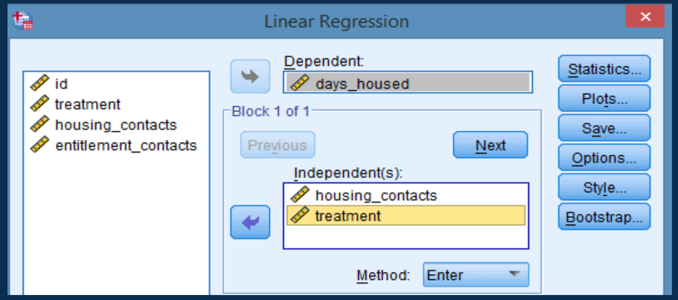




Step 1: simple linear regression for path c



Step 2: simple linear regression for path a



Step 3/4: multiple linear regression for paths b and c’

Output and Interpretation

Total effect-path c

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	12.784	1.607		7.955	.000	9.598	15.970
	Treatment	6.558	2.474	.248	2.651	.009	1.654	11.462

a. Dependent Variable: Average number of days per month in stable housing

Path a

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	2.689	.473		5.688	.000	1.752	3.626
	Treatment	1.831	.728	.236	2.517	.013	.389	3.274

a. Dependent Variable: Average number of days per month that the respondent was in contact with their assigned treatment programme

Path b

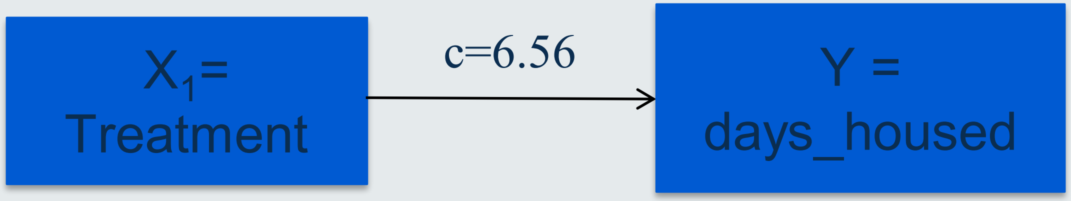
Path c'

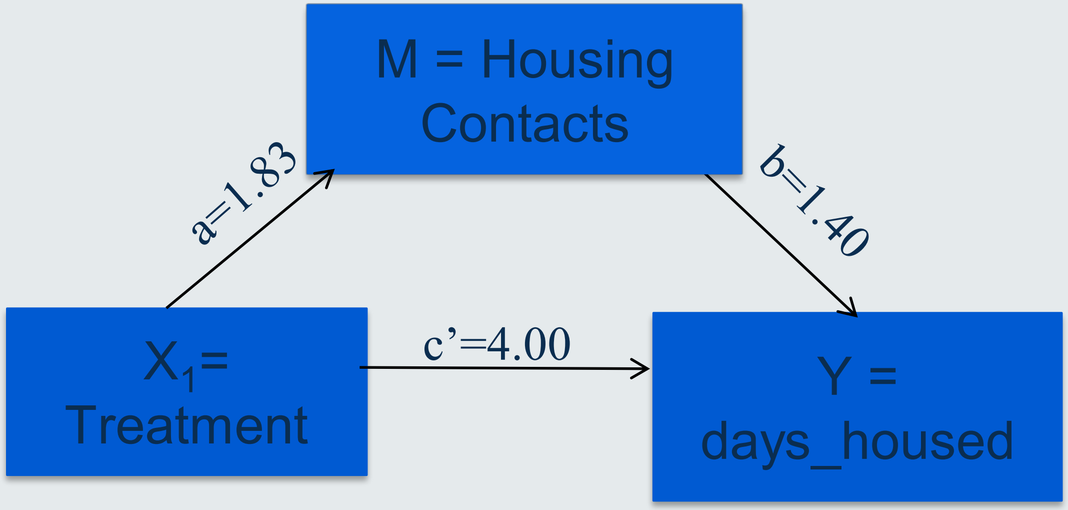
Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	9.025	1.680		5.373	.000	5.695	12.355
	housing_contacts	1.398	.301	.410	4.645	.000	.801	1.995
	Treatment	3.998	2.332	.151	1.715	.089	-.625	8.621

a. Dependent Variable: Average number of days per month in stable housing

Testing Mediation: Does Housing Contacts Mediate the Treatment Effect?





Step 1: **path c**

X_1 is associated with Y

days_housed =
12.78 + **6.56** x **treatment** $+\varepsilon$

Step 2: **path a**

X_1 is associated with M

housing_contacts =
2.69 + **1.83** x **treatment**$+\varepsilon$

Step 3: **path b**

M is associated with Y, regardless X_1

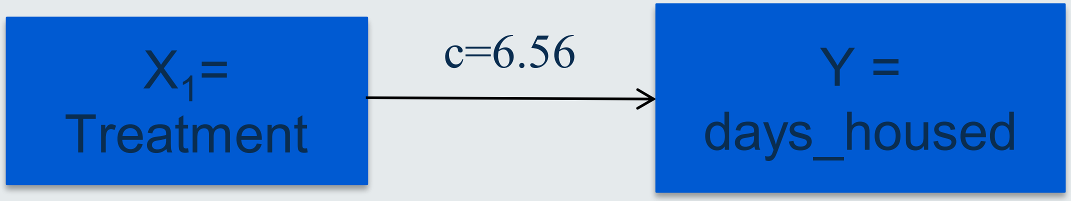
days_housed =
9.03+**1.40** x **housing_contacts**
+ 4.00 x treatment$+\varepsilon$

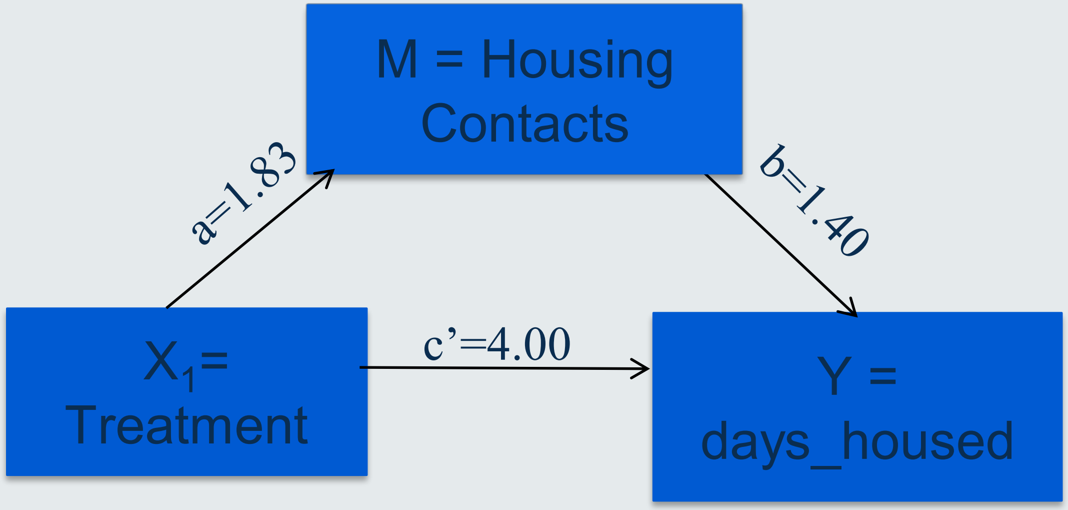
Step 4: **path c’**

X_1 is associated with Y, regardless M

days_housed =
9.03+1.40 x housing_contacts
+ 4.00 x **treatment**$+\varepsilon$

Testing Mediation: Does Housing Contacts Mediate the Treatment Effect?





Step 1: **path c**

X_1 is associated with Y

days_housed =
12.78 + **6.56** x **treatment** $+\varepsilon$

Step 2: **path a**

X_1 is associated with M

housing_contacts =
2.69 + **1.83** x **treatment**$+\varepsilon$

Step 3: **path b**

M is associated with Y, regardless X_1

days_housed =
9.03+**1.40** x **housing_contacts**
+ 4.00 x treatment$+\varepsilon$

Step 4: **path c'**
(Alternative way)

Indirect effect:
a*b = 1.83 x 1.40 = 2.56

Direct effect:

c = c' + a*b
6.56 = c' + 2.56
c'= 4.00

Output and Interpretation

<u>Step 1</u>: Test path c ($X_1 \rightarrow Y$): $Y = \beta_0 + \beta X_1 + \varepsilon$

(x_1 = treatment $\rightarrow$ days_housed = y):

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	12.784	1.607		7.955	.000	9.598	15.970
	Treatment	6.558	2.474	.248	2.651	.009	1.654	11.462

a. Dependent Variable: Average number of days per month in stable housing

Path c (effect of treatment on stable housing) is equal to 6.558 (p value = 0.009), with a 95% confidence interval of [1.65 to 11.46]

Treatment has a significant effect on the outcome – Step 1 passed

Output and Interpretation

<u>Step 2</u>: Test path a ($X_1 \rightarrow M$): $M = \beta_0 + \beta_1 X1 + \varepsilon$

(x_1 = treatment $\rightarrow$ housing_contacts = M):

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	2.689	.473		5.688	.000	1.752	3.626
	Treatment	1.831	.728	.236	2.517	.013	.389	3.274

a. Dependent Variable: Average number of days per month that the respondent was in contact with their assigned treatment programme

Path a (effect of treatment on housing contact) is equal to 1.83 (p = 0.013), with a 95% confidence interval of [0.39 to 3.27]

Treatment has a significant effect on the hypothesised mediator – Step 2 passed

Output and Interpretation

<u>Step 3</u>: Test path b (M →Y, controlling for X_1): $Y = \beta_0 + \beta_2 M + \beta_3 X_1 + \varepsilon$

(x_1 = treatment, M = housing_contacts → days_housed = y):

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	9.025	1.680		5.373	.000	5.695	12.355
	housing_contacts	1.398	.301	.410	4.645	.000	.801	1.995
	Treatment	3.998	2.332	.151	1.715	.089	-.625	8.621

a. Dependent Variable: Average number of days per month in stable housing

Path b (effect of housing contacts on stable housing controlling for treatment) is equal to 1.398 ($p < 0.001$), with a 95% confidence interval of [0.801 to 1.995]

Mediator has a significant effect on the outcome – Step 3 passed

Output and Interpretation

Step 4: Test path c' : there is complete or partial mediation?

(x_1 = treatment, M = housing_contacts → days_housed = y):

Coefficients[a]

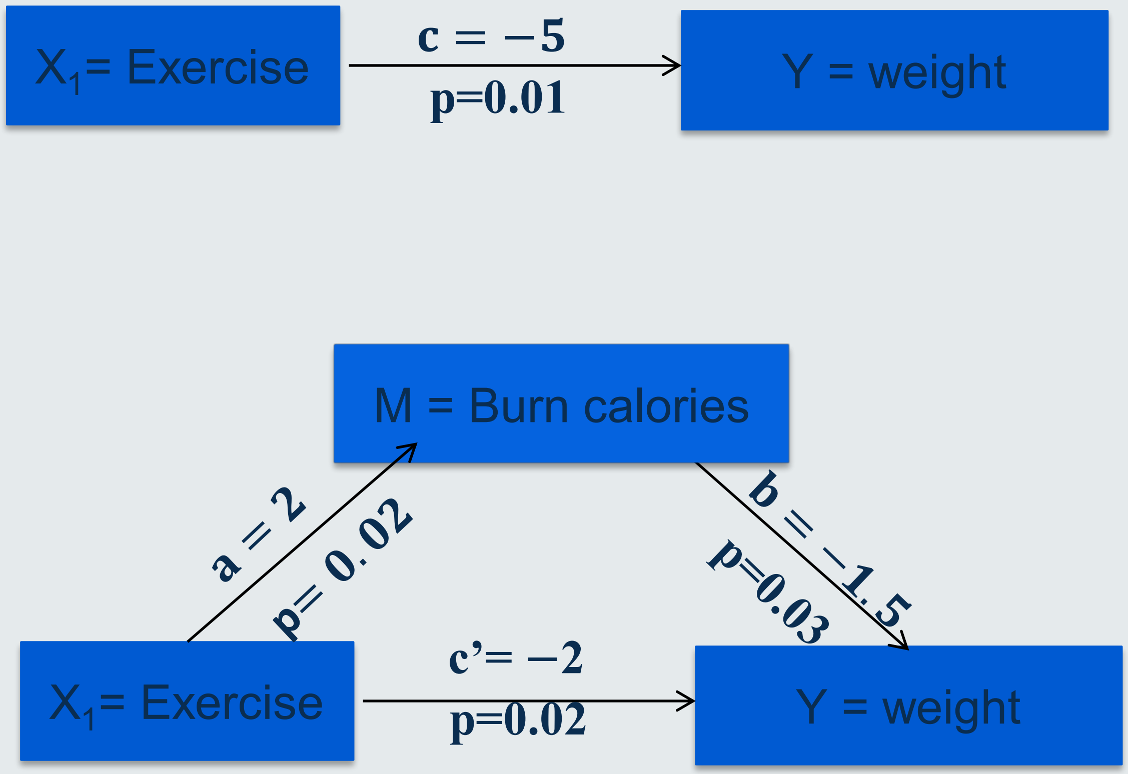
Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	9.025	1.680		5.373	.000	5.695	12.355
	housing_contacts	1.398	.301	.410	4.645	.000	.801	1.995
	Treatment	3.998	2.332	.151	1.715	.089	-.625	8.621

a. Dependent Variable: Average number of days per month in stable housing

- Path c′ is the **direct effect** of treatment on the outcome
- This is estimated from the same regression model fitted in Step 3
- **Path c′ (effect of treatment on stable housing controlling for the mediator) is equal to 4.00 (p =0.09), with a 95% confidence interval of -0.63 to 8.62**.
- Controlling for the mediator **substantially reduces** the effect of treatment (c′= 4.00 < c=6.56)
- **Step 4 passed**. We conclude: **There is complete mediation, as the direct effect is not significantly different from 0.**

Knowledge Check

Q1: Given the two path diagrams below, is there a complete or partial mediation?



Knowledge Check

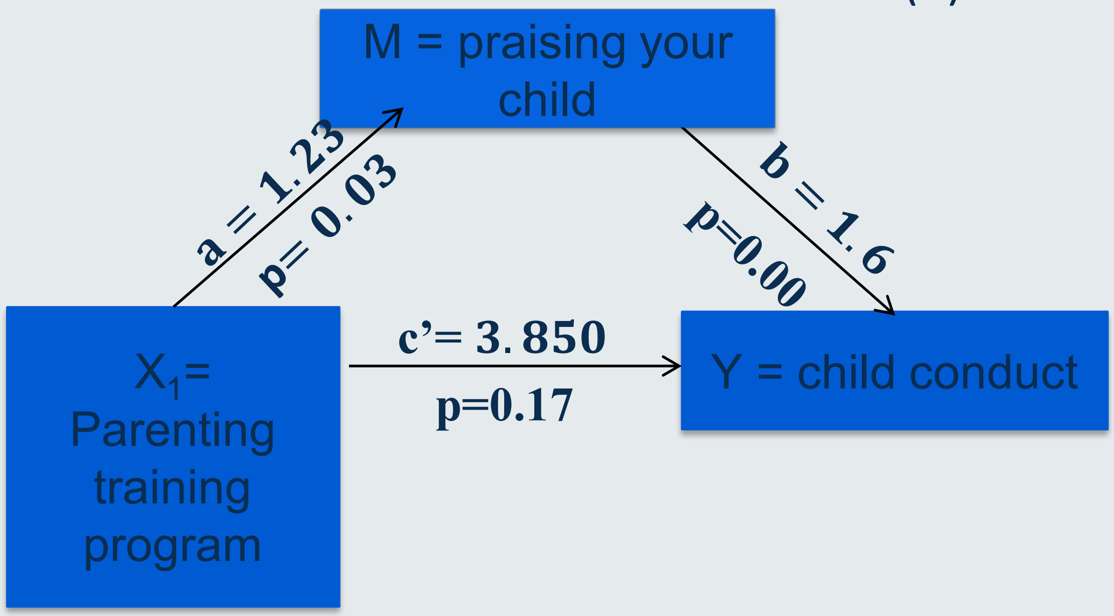
Q2: Given the path diagram below, Report:

(a) the mediated indirect effect

(b) the non-mediated direct effect

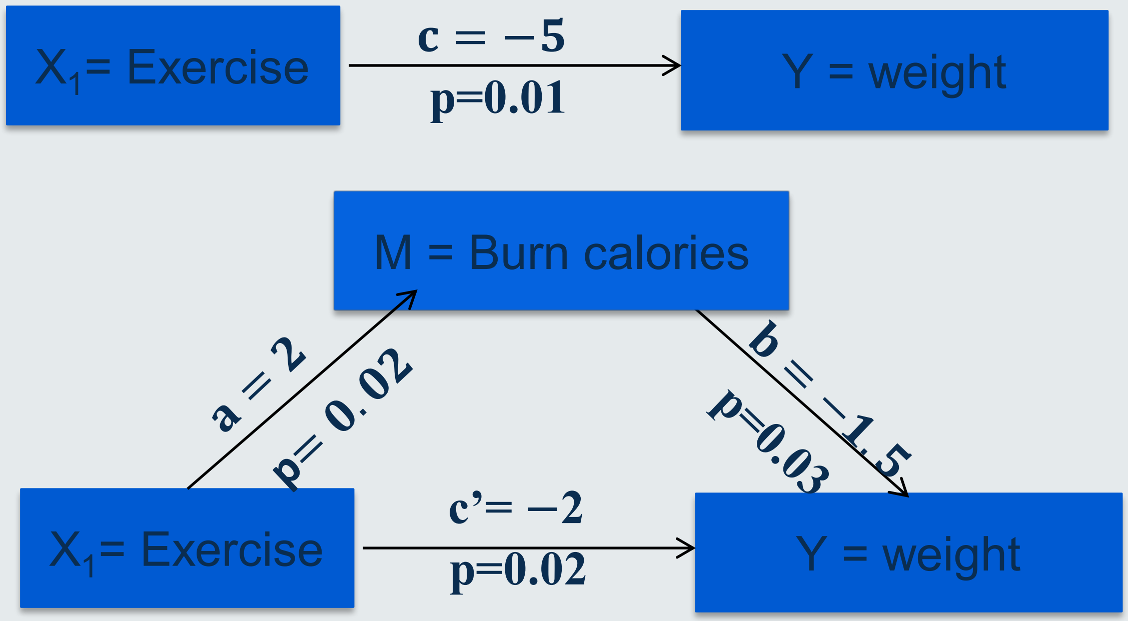
(c) the total effect

(d) can we establish mediation? is it complete or partial?



Knowledge Check Solutions

Q1: In the example below, is there a complete or partial mediation?



- Answer: As **c' is significantly different from 0**, of the total effect -5, there is a direct effect of -2 and a **partially mediated** indirect effect of -3.

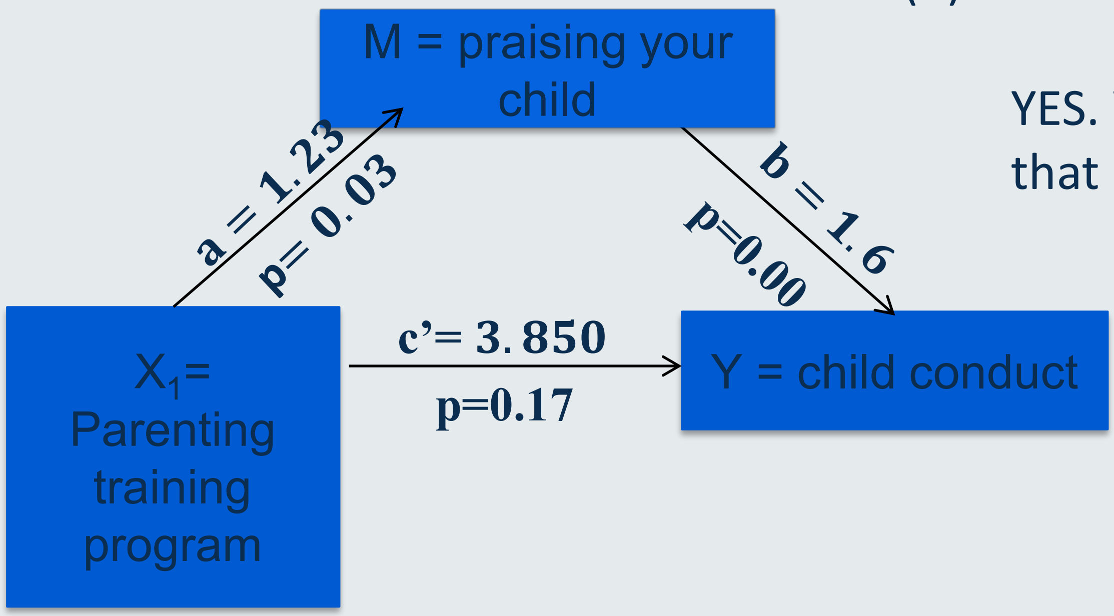
Knowledge Check Solutions

Q2: Given the path diagram below, Report: (a) the mediated indirect effect : a*b=1.23*1.60= **1.97**

(b) the non-mediated direct effect: c'= **3.850**

(c) the total effect: c= c' + a*b = 3.850 + 1.97 = **5.82**

(d) can we establish mediation? is it complete or partial?



YES. We can establish **complete** mediation given that paths a and b are significant, and c' is not significant.

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.

An extension to Baron and Kenny: Andrew F. Hayes (2009) Beyond Baron and Kenny: Statistical Mediation Analysis in the New Millennium, Communication Monographs, 76:4, 408-420, DOI:10.1080/03637750903310360. To link to this article: http://dx.doi.org/10.1080/03637750903310360

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**

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

08/2020

Module Title: Introduction to Statistics

Session Title: Testing Indirect Effects

Topic title: Mediation

After working through this session you should be able to:

- To understand how to test the indirect effect to establish mediation
- To use parametric and non-parametric tests for testing the indirect effects

Baron and Kenny Steps

Before, we focused on understanding the four steps from Baron and Kenny to establish mediation.

Are all four steps essential?

- Step 1 establishes that there is an effect (path c) that may be mediated, but is **not essential** for establishing mediation. (see **http://davidakenny.net/cm/mediate.htm**)
- Steps 2 and 3 are **essential** for establishing mediation
  - These steps (2 & 3) establish paths a and b (and also c′) which lead to an estimate of the indirect effect (ab). Existence of an indirect effect is **sufficient** to justify mediation
- Newer methods (e.g. Sobel test) recommend testing **only the indirect effect** (paths a and b) to establish mediation

Testing the Indirect Effect "ab"

- There are several methods for testing the indirect effect:

$$\begin{cases} H_0: ab = 0 \\ H_1: ab \neq 0 \end{cases}$$

- Two of the commonly used tests are:

- Sobel test (Normal Theory Approach)
- Nonparametric Sobel test (bootstrapping)

Sobel Test of Indirect Effect

$$\begin{cases} H_0: ab = 0 \\ H_1: ab \neq 0 \end{cases}$$

- Sobel statistic test is based on an approximate z-statistic, given by: $z = \frac{ab}{SE(ab)}$
- SE(ab) denotes the standard error of the estimated indirect effect, given by:

$$\mathbf{SE(ab)} = \sqrt{\mathbf{a^2 S_b^2 + b^2 S_a^2}}$$

Where $\mathbf{S_a}$ and $\mathbf{S_b}$
are **SE of the coefficients for a and b**
(Taken from the multiple linear regression model)

- <u>Decision rule</u>: if Z in absolute value is greater than 1.96, reject the hypothesis that the indirect effect is zero.

Software, Output and Interpretation Slide

- The test can be done using online calculator
- **http://quantpsy.org/sobel/sobel.htm**

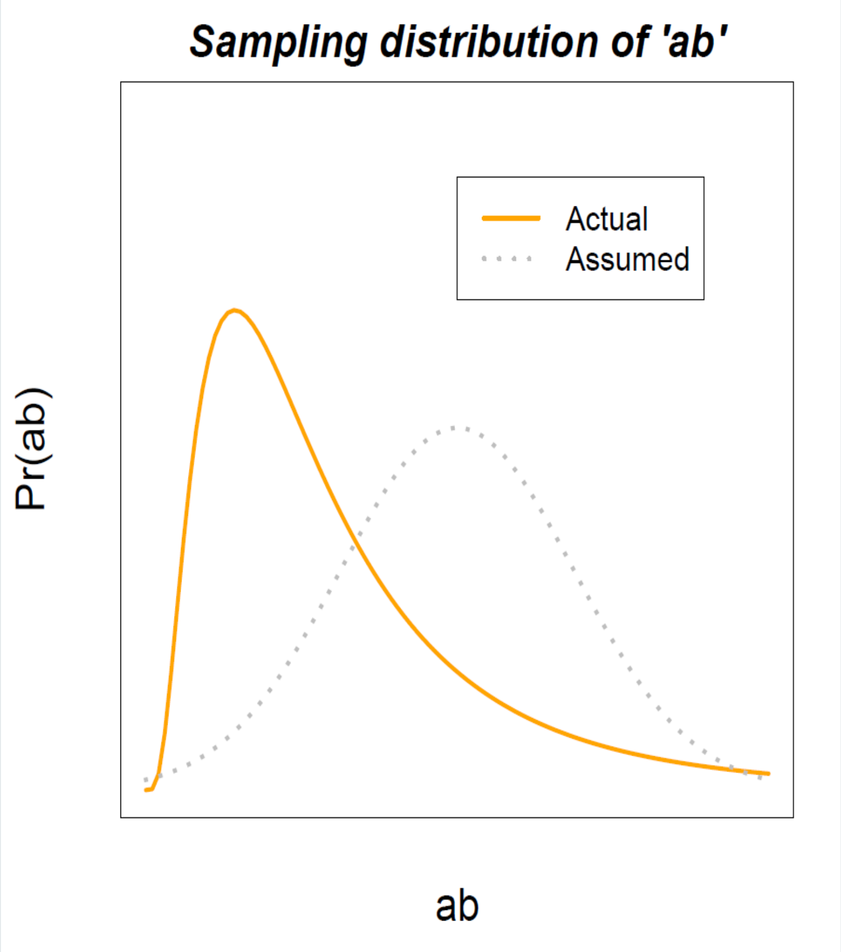
Input:			Test statistic:	Std. Error:	*p*-value:
a	1.83	Sobel test:			
b	1.40	Aroian test:			
s_a	0.30	Goodman test:			
s_b	0.73	Reset all	Calculate		

Input:			Test statistic:	Std. Error:	*p*-value:
a	1.83	Sobel test:	1.8295199	1.40036738	0.06732176
b	1.40	Aroian test:	1.80754975	1.41738838	0.07067661
s_a	0.30	Goodman test:	1.85231116	1.38313695	0.06398115
s_b	0.73	Reset all	Calculate		

As Z (Sobel Test Statistic) in absolute value is less than 1.96, fail to reject the null hypothesis that the indirect effect is zero (p=0.067)

Limitation of Sobel Test

- Sobel test is based on **normal** approximation (z-test)
- Sampling distribution of ‘ab’ is actually highly skewed
- Large values of ‘ab’ are more variable than the smaller values
- This may lower the statistical power of the Sobel test
- Sobel test works well only in **large samples**, because the skewness is reduced.



Non-parametric Sobel Test

- Nonparametric version of Sobel test via bootstrapping offers a **better alternative** that **imposes no distributional assumptions.**
- **Bootstrapping** requires taking a **large number of samples** (with replacement) from the original dataset
- **Indirect effect (ab)** is **estimated for each** of the bootstrap samples
- These bootstrap estimates are used to **form a non-parametric sampling distribution** of the indirect effect
- From the sampling distribution a **confidence interval for ab** is estimated.
- Indirect effect is said to be significant if the **confidence interval does not contain zero.**

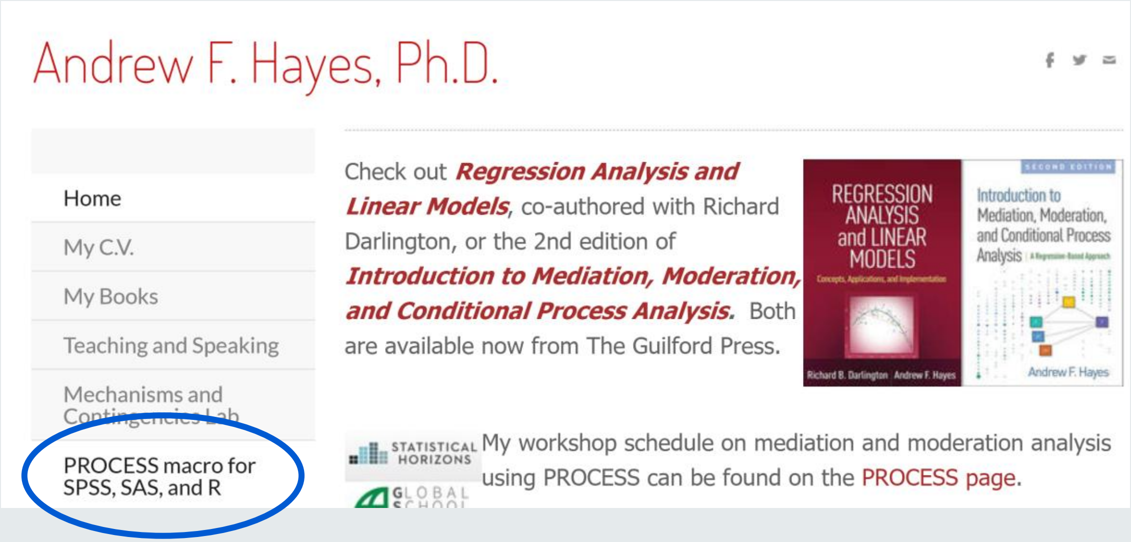
Bias-corrected Bootstrap

It is possible that the mean of the bootstrap estimates differs slightly from the original estimate of indirect effect (ab).

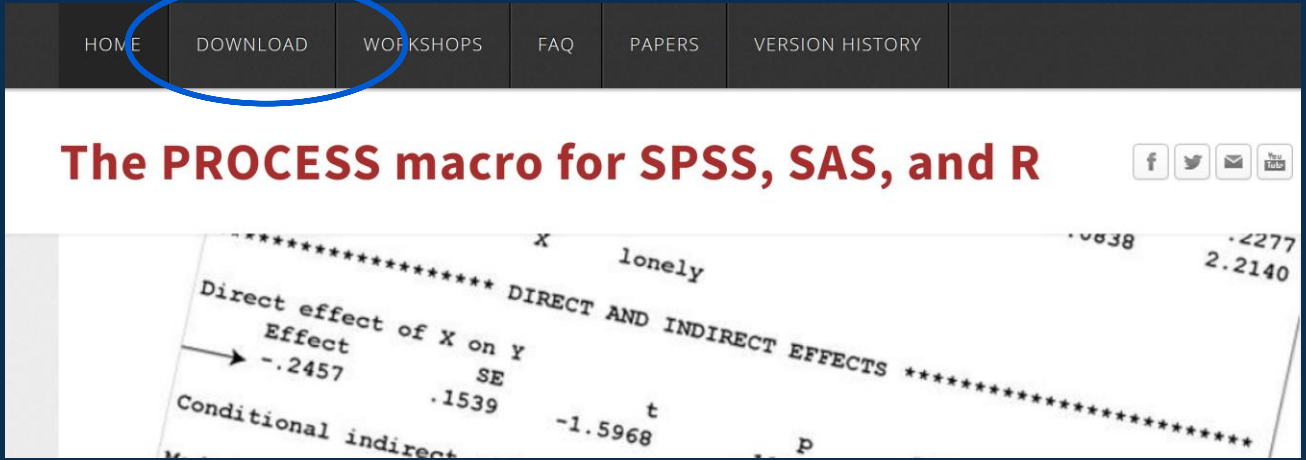
Bias-corrected bootstrap is the recommended method for testing indirect effect

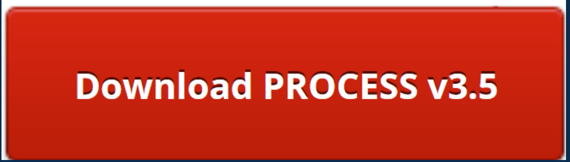
This can be done using the PROCESS macro (see Hayes & Rockwood, 2017), if installed in your SPSS

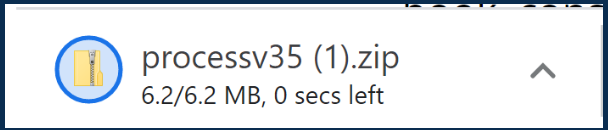
Can be downloaded free from Andrew Hayes' website http://www.afhayes.com



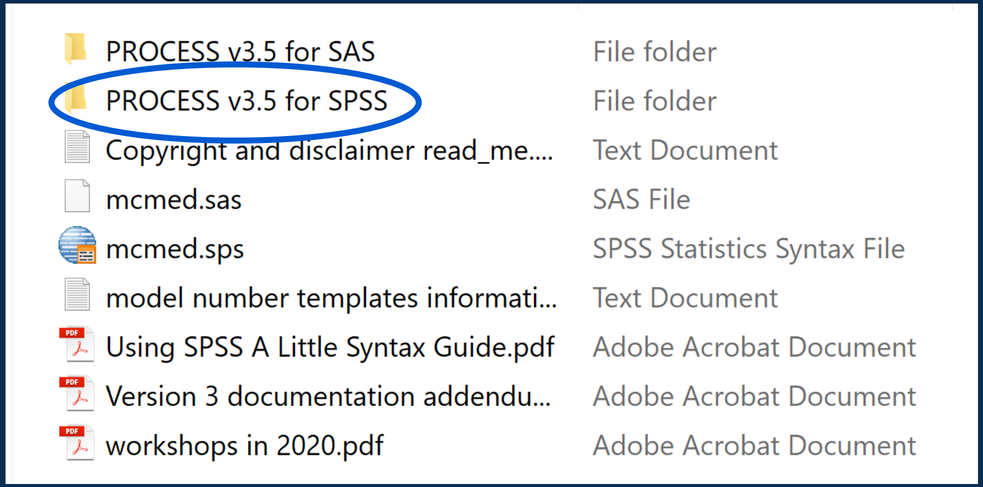
Process Macro 'how to' Option 1

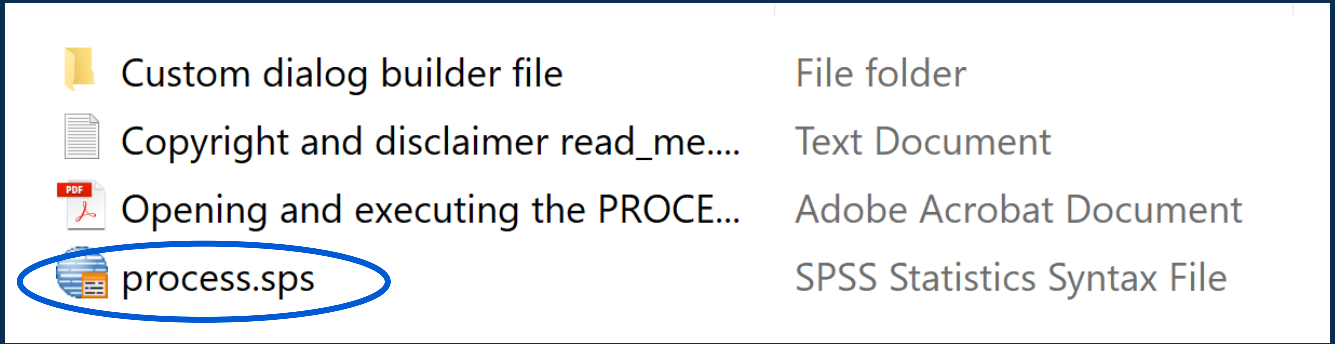






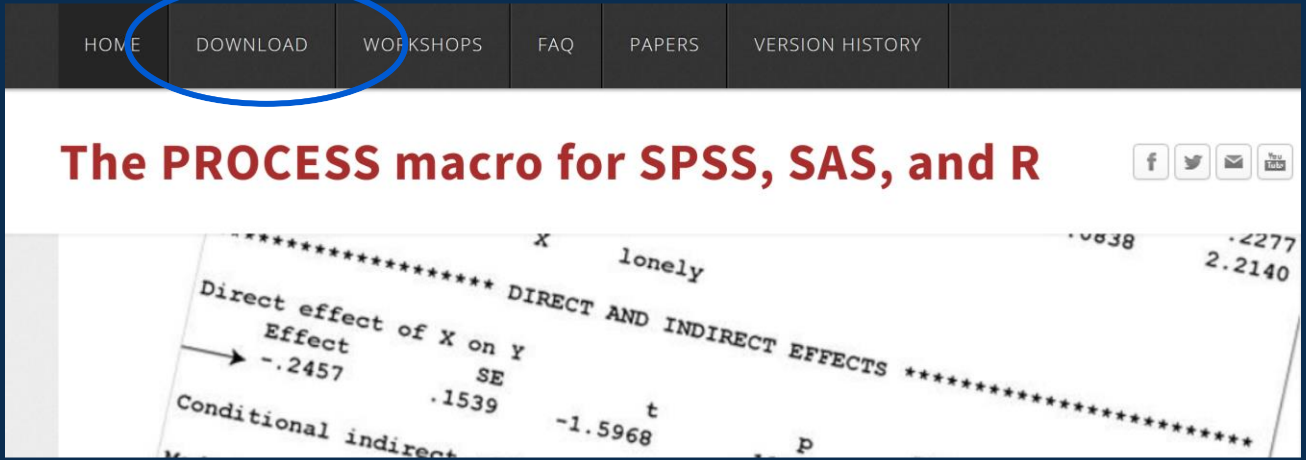
Click 'Download; Scroll down the page and Click 'Download PROCESS v3.5). Open the zip file Open the PROCESS v3,5 for SPSS folder.

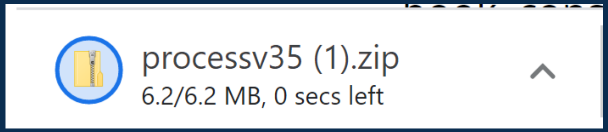




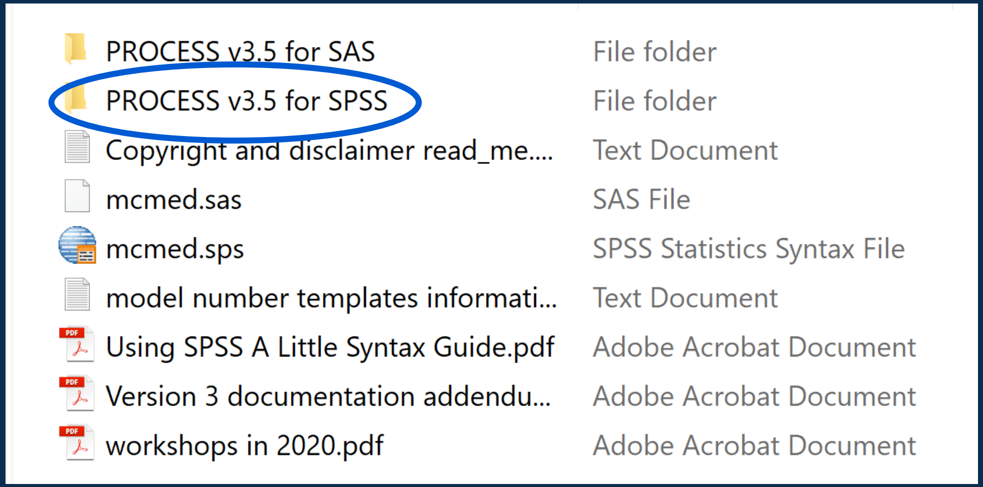
Extract and Open the process.sps file in a new syntax window in SPSS
Run all by , selecting all syntax clicking on the big green triangle

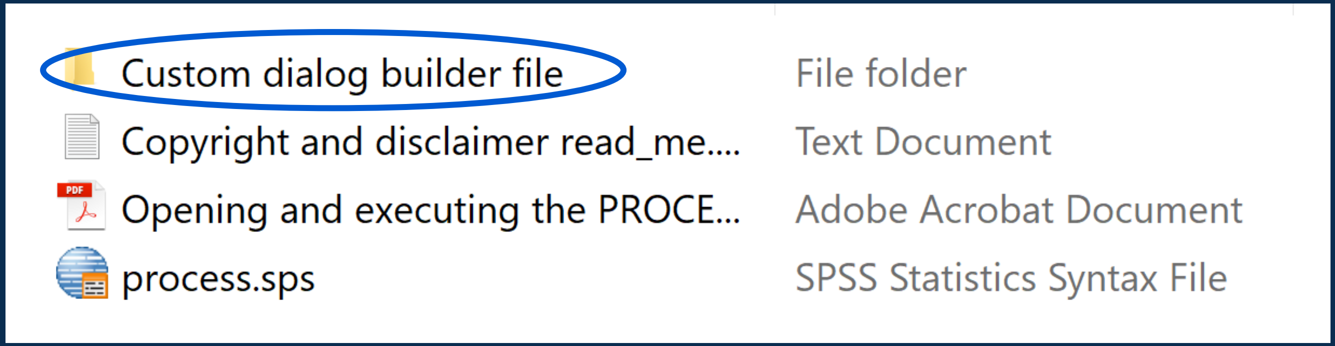
Process Macro 'how to' Option 2



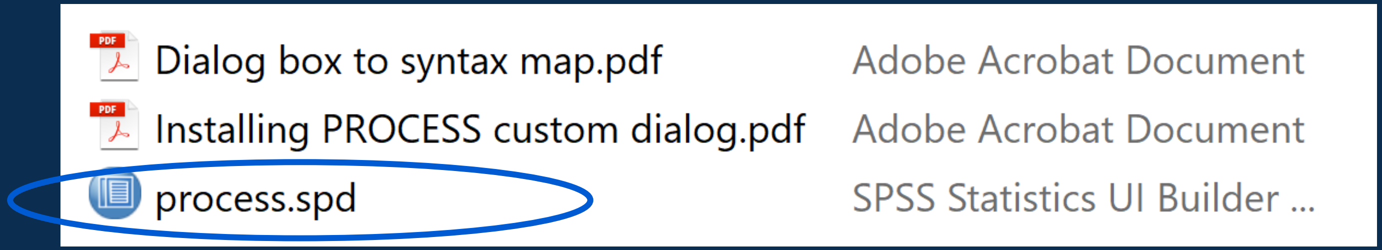


Click 'Download; Scroll down the page and Click 'Download PROCESS v3.5). Open the zip file Open the PROCESS v3,5 for SPSS folder.



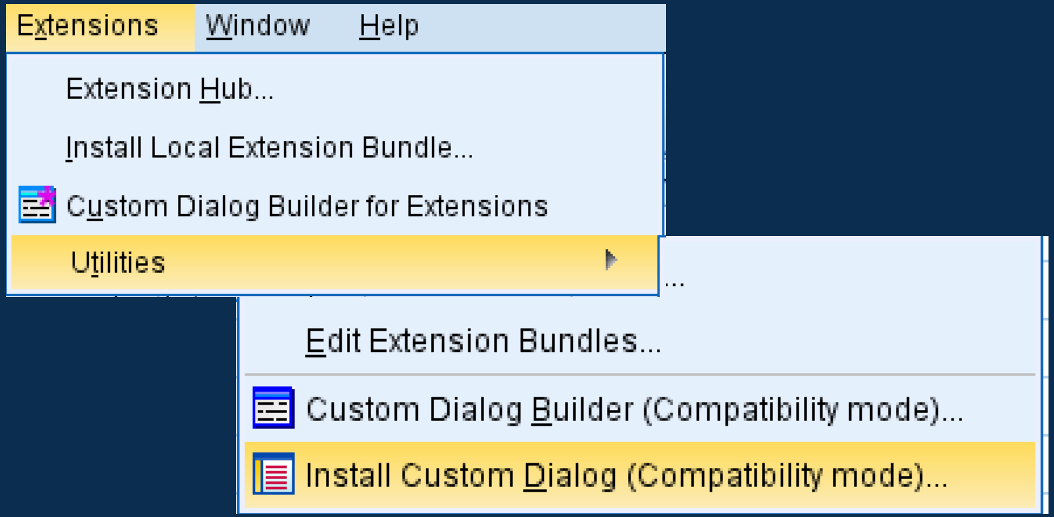


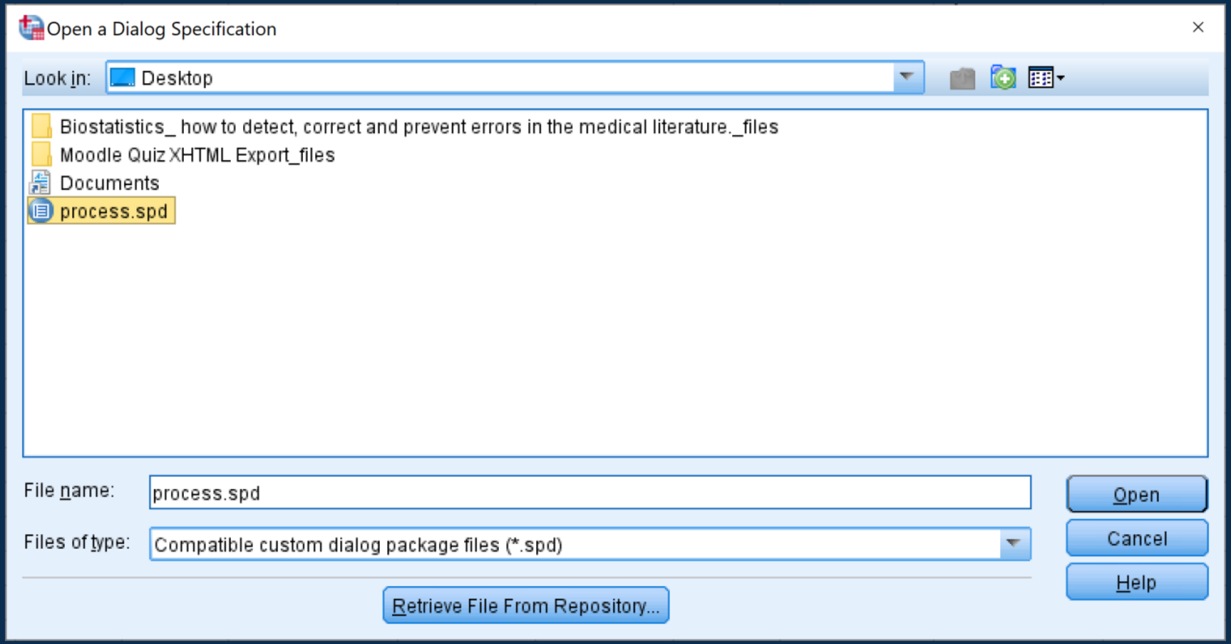
Process Macro ‘how to’ Option 2



Extract the process.spd file

Extensions → Utilities → Install Custom Dialog





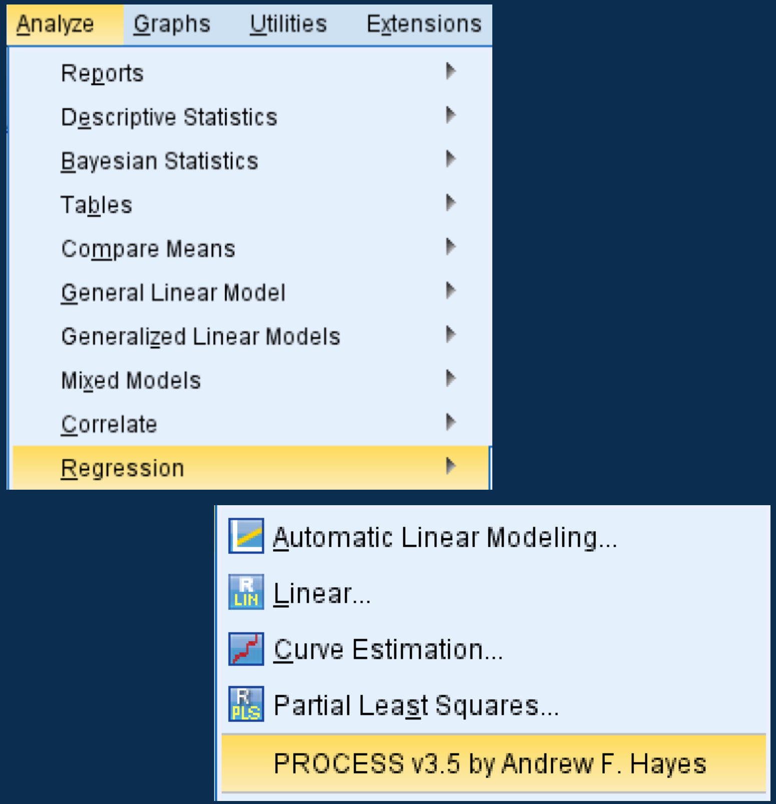
SPSS Slide: 'How to'

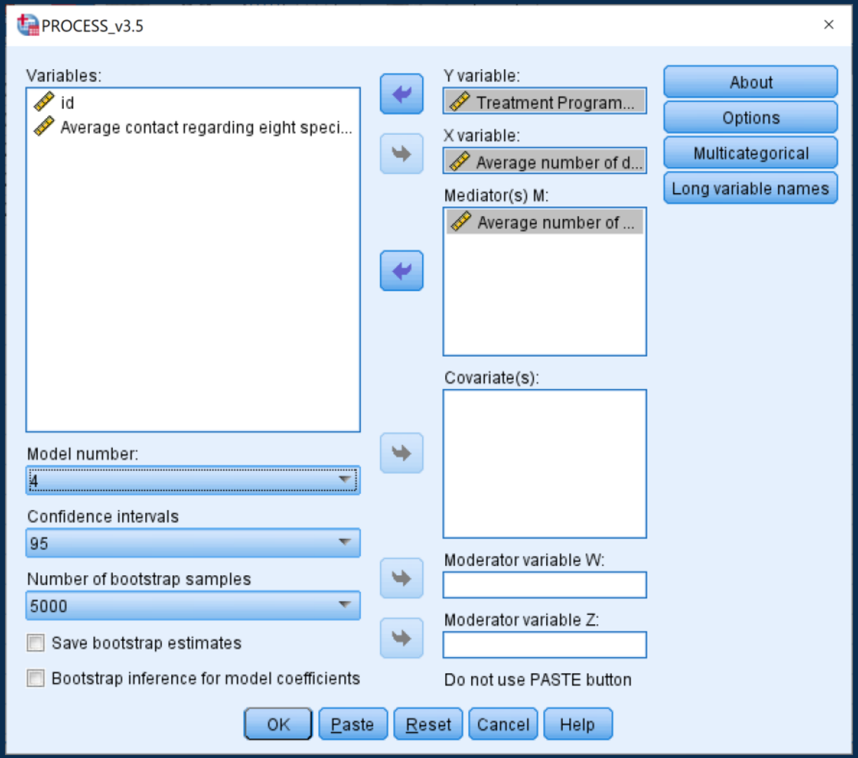
Use Lecture_8_data to test an indirect effect. In the regression menu you will see a new option PROCESS

Computing probit regression models

1) Use **Analyse -> Regression -> PROCESS**

2) Add' days_hous' in '**Outcome**' box, 'treat' in the '**independent variables**' box and the contacts in the '**M Variables**'. Box, choose **'Model 4'**





Note:
PROCESS does not allow variable names to be more than <u>eight</u> characters

Make the names shorter in the 'variable view' of the dataset.

Output and Interpretation Slide

```
Model  : 4
    Y  : days_hou
    X  : treat
    M  : contacts

Sample
Size:  109
```

OUTCOME VARIABLE:
contacts

Model Summary

R	R-sq	MSE	F	df1	df2	p
.2364	.0559	14.0765	6.3329	1.0000	107.0000	.0133

Model

	coeff	se	t	p	LLCI	ULCI
constant	2.6889	.4727	5.6885	.0000	1.7518	3.6259
treat	1.8311	.7276	2.5165	.0133	.3887	3.2736

OUTCOME VARIABLE:
days_hou

Model Summary

R	R-sq	MSE	F	df1	df2	p
.4694	.2203	136.4668	14.9774	2.0000	106.0000	.0000

Model

	coeff	se	t	p	LLCI	ULCI
constant	9.0246	1.6796	5.3729	.0000	5.6946	12.3547
treat	3.9979	2.3317	1.7146	.0893	-.6249	8.6206
contacts	1.3982	.3010	4.6450	.0000	.8014	1.9949

Printed: Baron and Kenny Step 2 and Step 3

Output and Interpretation

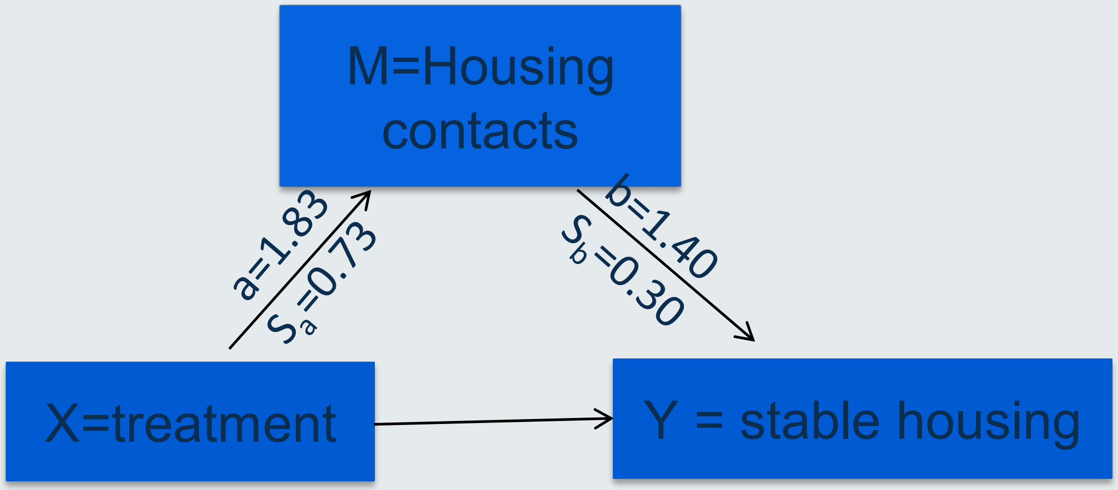
```
****************** DIRECT AND INDIRECT EFFECTS OF X ON Y ******************

Direct effect of X on Y
     Effect          se           t           p        LLCI        ULCI
     3.9979      2.3317      1.7146       .0893      -.6249      8.6206

Indirect effect(s) of X on Y:
             Effect      BootSE    BootLLCI    BootULCI
contacts     2.5602      1.1526       .4928      5.0439
```

Check the 95% Bias-corrected bootstrap confidence interval. As the interval does not contain zero **we can reject the null hypothesis that the indirect effect is zero and say that the indirect effect is significant. Thus, there is significant mediation.**

Knowledge Check

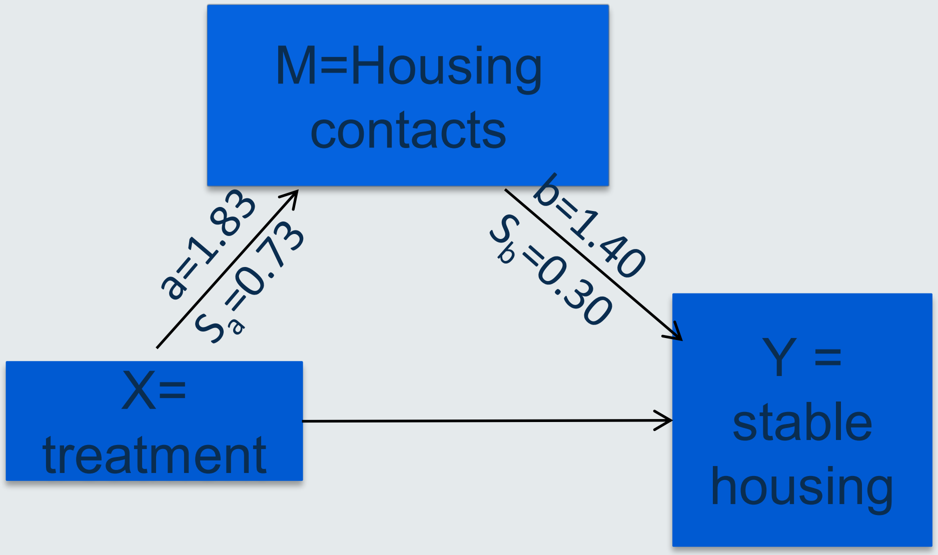
Using the stable housing data and the given path diagram, answer:

Q1. Compute the indirect effect 'ab'.
Q2. Compute the standard error of the indirect effect 'se(ab)'
Q3. Is the indirect effect significantly different from zero?



Knowledge Check Solutions

Using the stable housing data and the given path diagram, answer:



Q1. Compute the indirect effect 'ab'.

Indirect effect: ab = 1.83×1.40 =2.56

Q2. Compute the standard error of the indirect effect 'SE(ab)'

$$\mathbf{SE(ab)} = \sqrt{\mathbf{a^2 S_b^2 + b^2 S_a^2}} = \sqrt{(\mathbf{1.83})^2(\mathbf{0.30})^2 + (\mathbf{1.4})^2(\mathbf{0.73})^2} = \mathbf{1.16}$$

Q3. Is the indirect effect significantly different from zero?

Z-statistic = ab/SE(ab)= 2.56/1.16= 2.21; Z-statistic > 1.96, we reject the hypothesis that ab=0 (at 5% significance level) p-value =0.027 (<0.05; significant)

We conclude that the indirect effect is **statistically different from zero**.

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.

Andrew Hayes' website (www.afhayes.com) offering free downloads of SPSS macros plus data files for the book's examples.

Preacher, Kristopher J.; Hayes, Andrew F (2008). "Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models". Behavior Research Methods. 40 (3): 879–891. doi:10.3758/BRM.40.3.879

Frazer, Baron and Tix (2004) Testing Moderator and Mediator Effects in Counselling Psychology
Journal of Counselling Psychology Copyright 2004 by the American Psychological Association, Inc.
2004, Vol. 51, No. 1, 115–134 0022-0167/04/$12.00 DOI: 10.1037/0022-0167.51.1.115

More advanced book:
MacKinnon, D. P (2007). Introduction to Statistical Mediation Analysis, Lawrence Erlbaum Associates, New York

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**