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

King's College London

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**

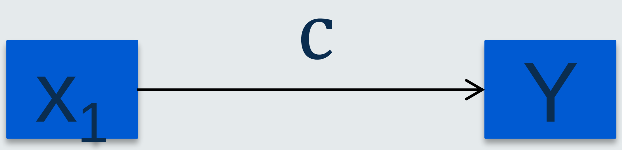
Previously on ‘Introduction to Statistics’

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

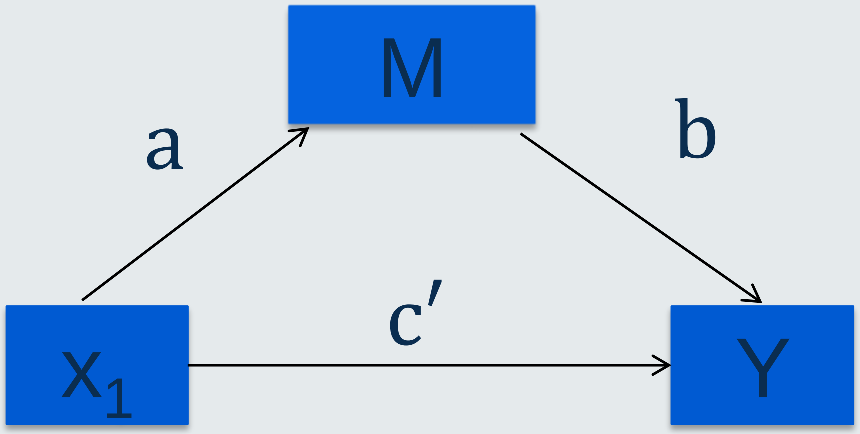
We discussed the different roles that a third variable $\mathbf{X_2}$ can have while investigating the association between an independent X_1 and a dependent variable Y.
X_2 could be a **confounder** (C) or a **mediator** (M)

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) non-mediated model



(B) mediated model



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**?

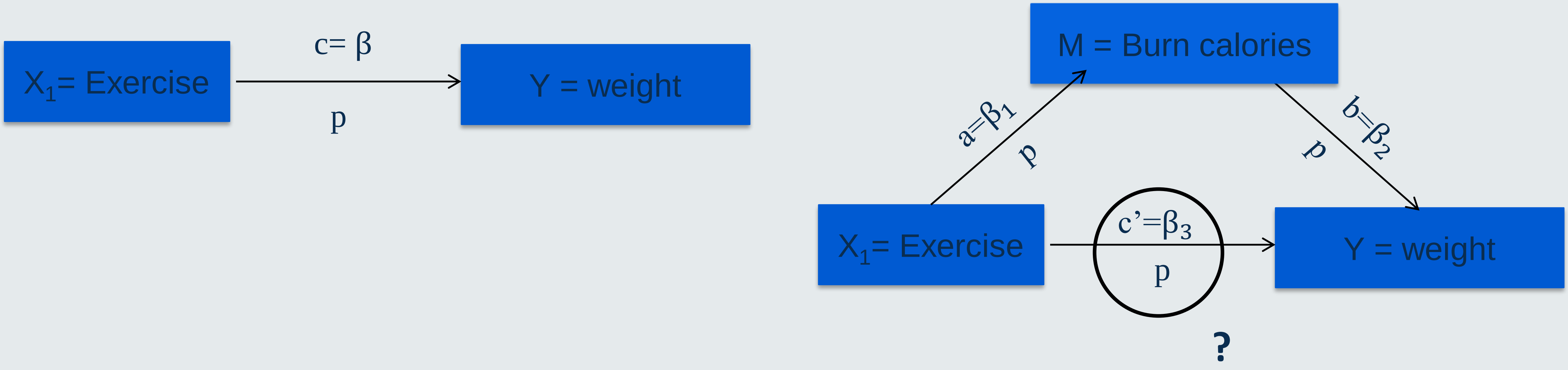
Baron and Kenny (1986) discussed **four steps** to establish mediation:

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

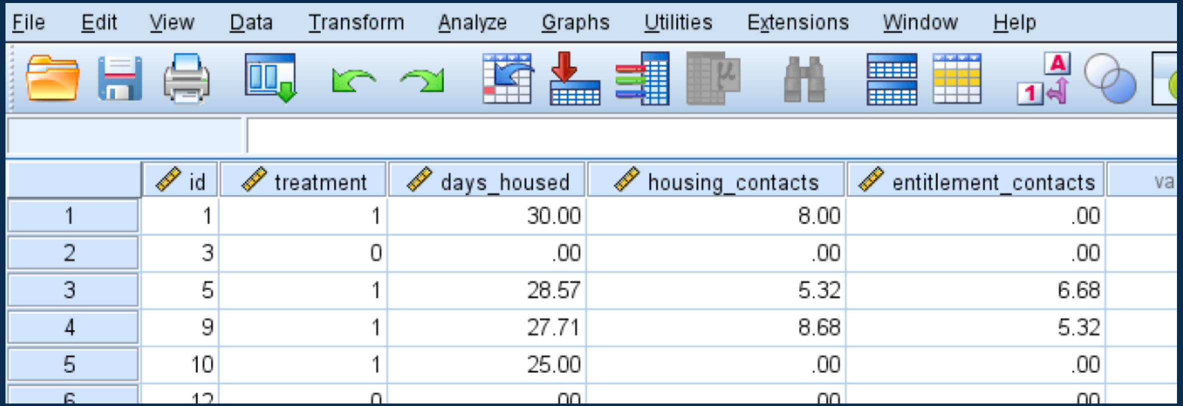
<u>Step 4</u>: **Test path c′ ($X_1 \rightarrow Y$, controlling for M):** $Y = \beta_0 + \beta_2 M + \beta_3 X_1 + \varepsilon$;

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

- If β_3 is <u>not significantly different from 0</u> ($p>0.05$) 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 <u>significantly different from 0</u> ($p<0.05$) 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

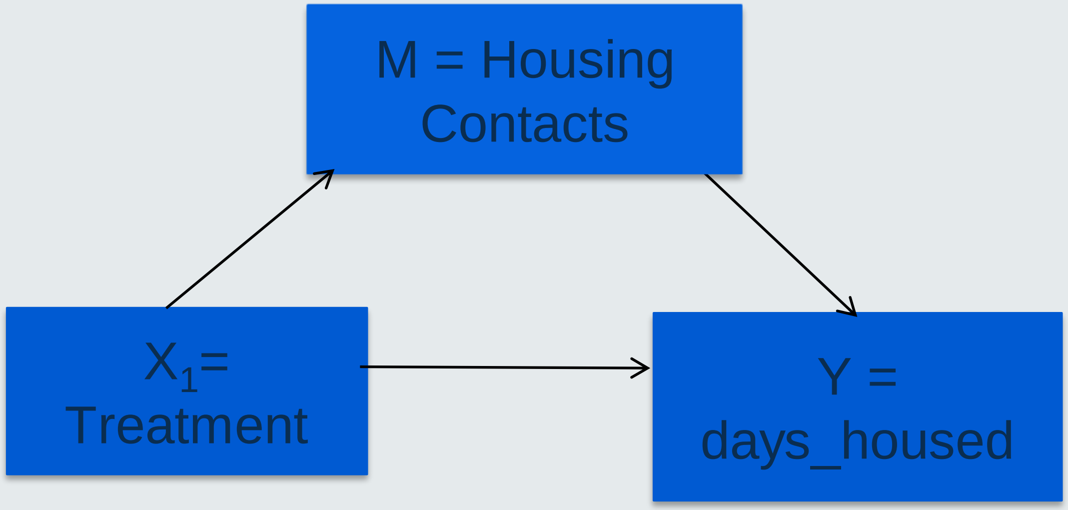
The dataset contains data from 109 subjects, measuring days in stable housing after receiving continuous treatment programme versus the standard treatment. Collecting information in respect to

- **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

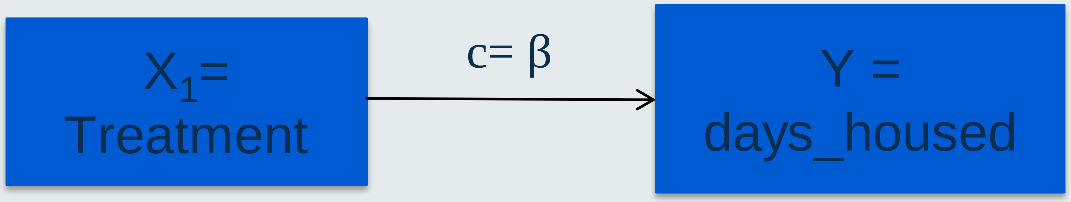
Example: Effect of Treatment on Stable Housing

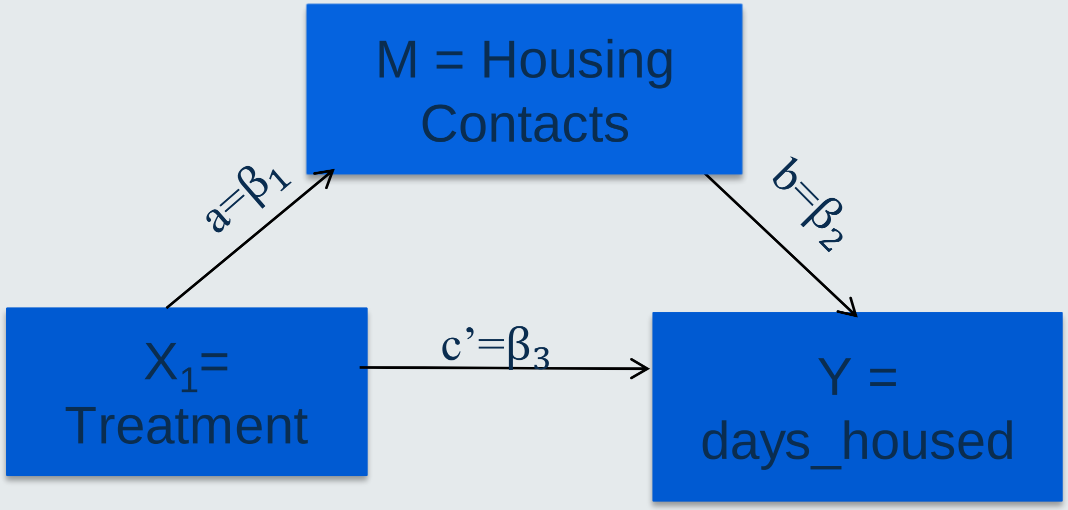
Homeless people are more likely to have serious mental illness. Morse et al. (1994) found that a treatment program which gives continuous support can be effective in increasing the average days they spend in stable housing (‘Treatment’ → ‘days_housed’).
The study also looked at the treatment effects 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

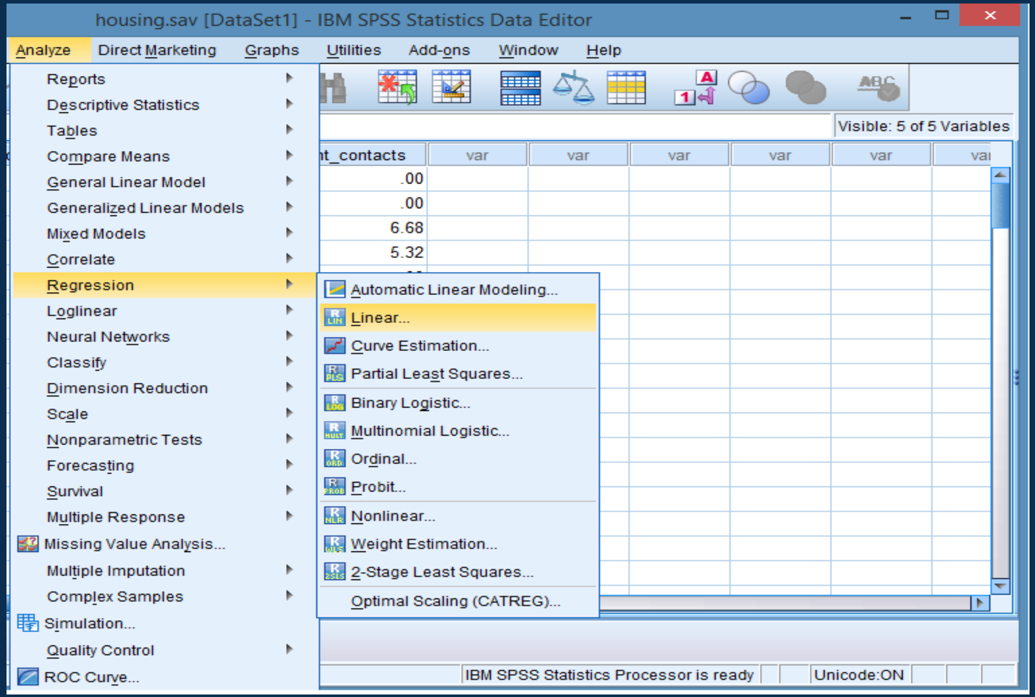
$Y = \beta_0 + \beta_2 M +\beta_3 X_1+\varepsilon$

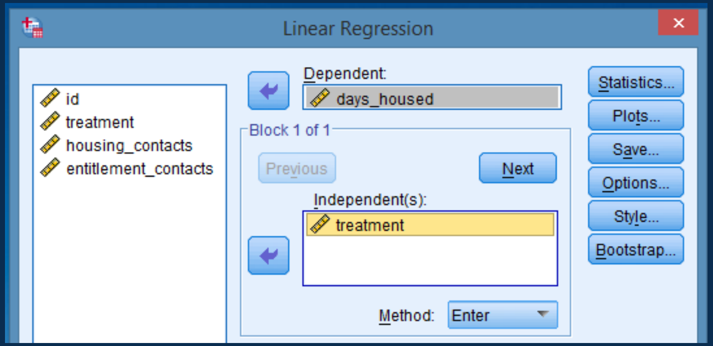
$$\begin{cases} H_0: \boldsymbol{\beta_3} = 0 \\ H_1: \boldsymbol{\beta_3} \neq 0 \end{cases}$$

SPSS Slide: ‘How to’ Steps

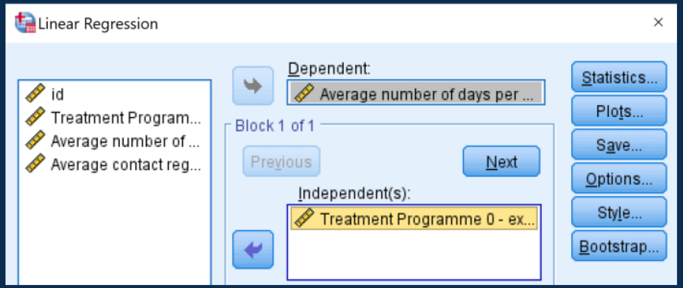
Computing three linear regression models from ‘housing.sav’ data:

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

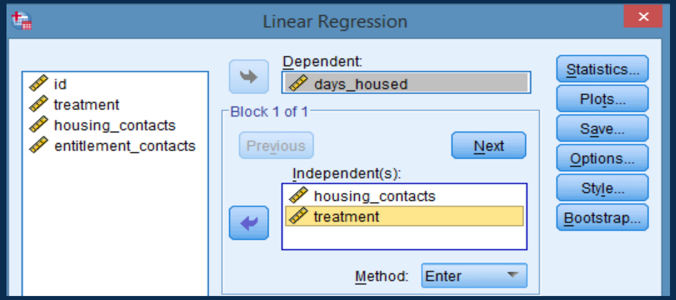




Step 1: simple linear regression for path c



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

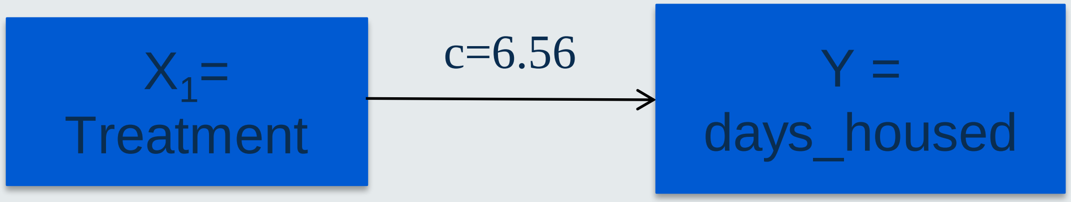
Path c’

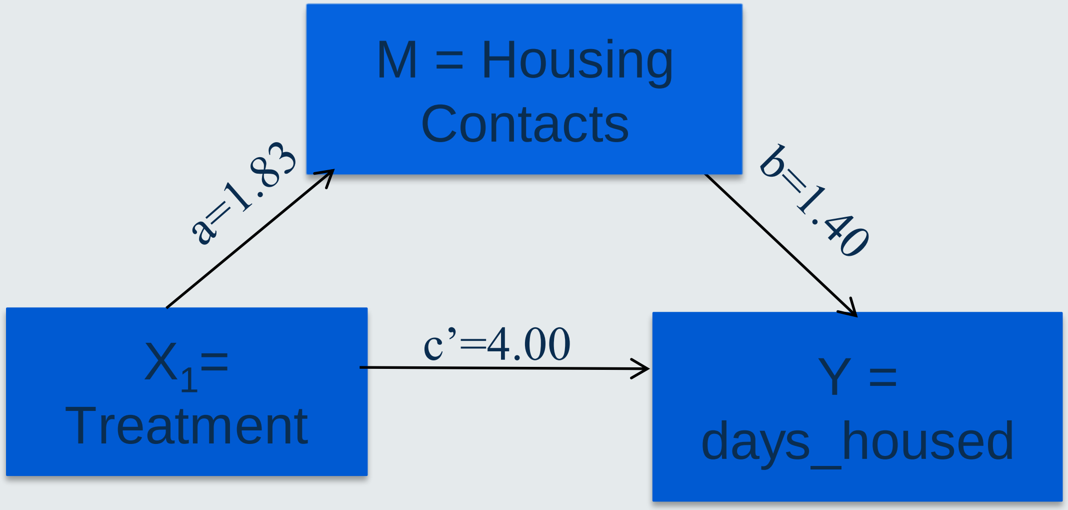
Coefficients[a]

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**$+\varepsilon$

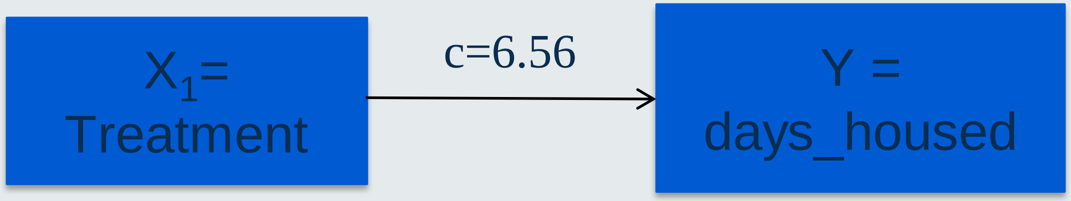
Step 3: **path b**
M is associated with Y, regardless X_1

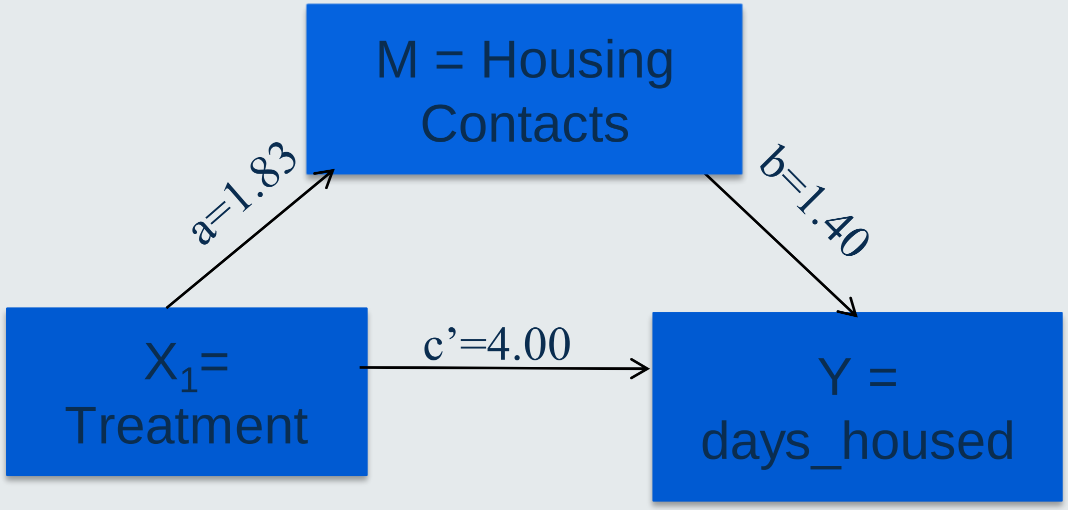
days_housed =
9.03+**1.40** x **housing_contacts**
+ 4.00 x treatment$+\varepsilon$

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 → housing_contacts = M):

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	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]

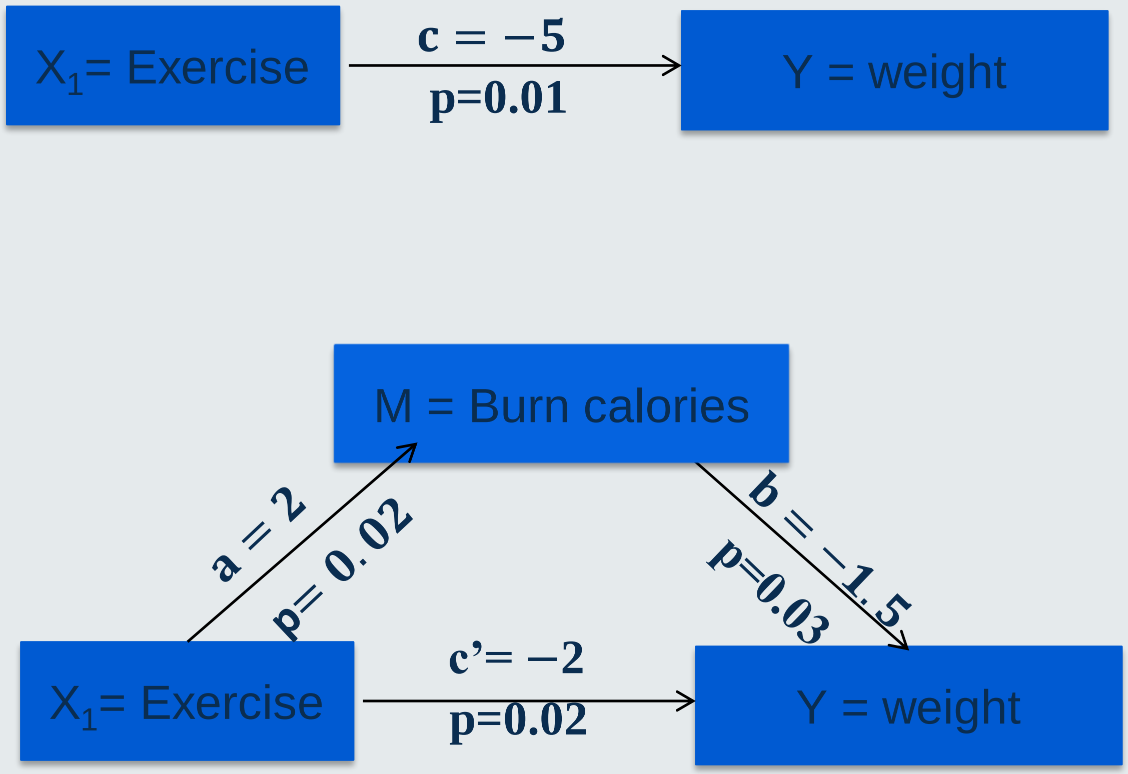
Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	9.025	1.680		5.373	.000	5.695	12.355
	housing_contacts	1.398	.301	.410	4.645	.000	.801	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

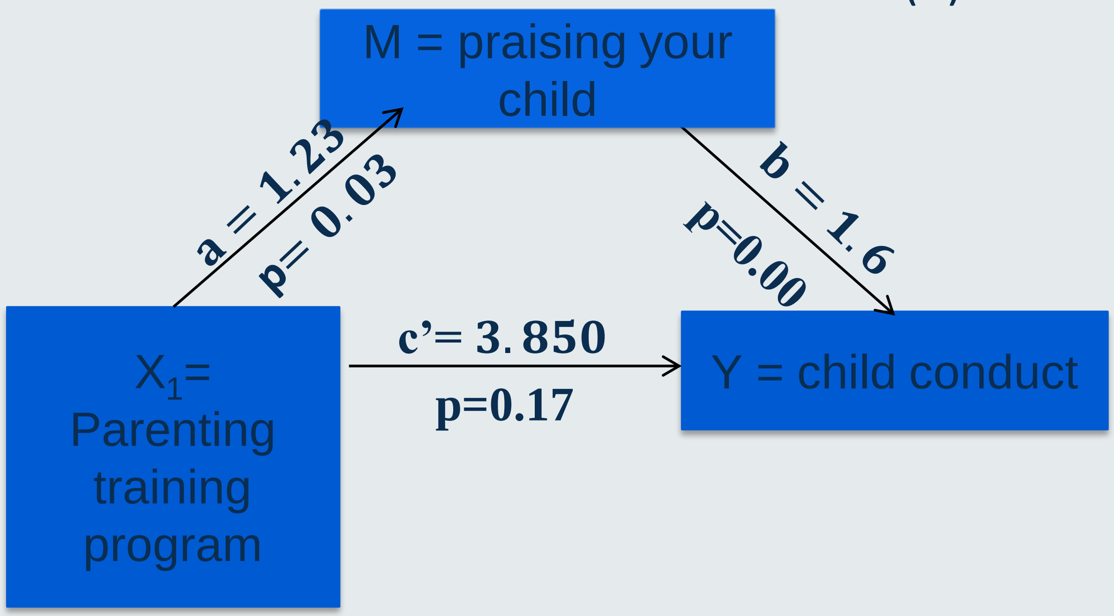
Q1: Given the two path diagrams below, is there a complete or partial mediation?



Knowledge Check

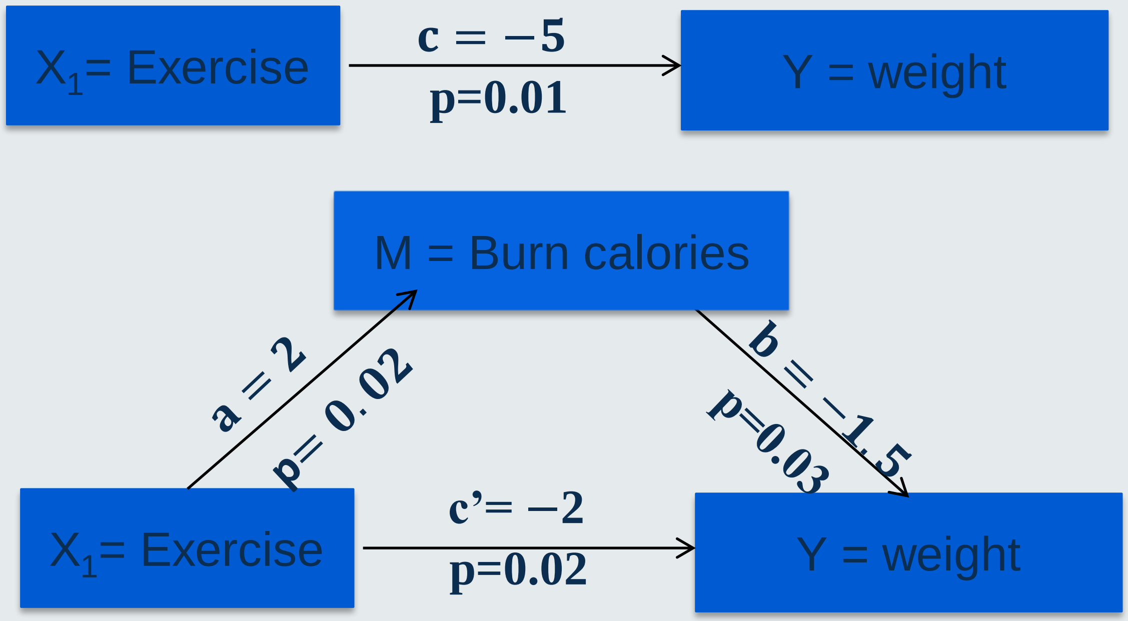
Q2: Given the path diagram below, Report:

(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?



- <u>Answer</u>: 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

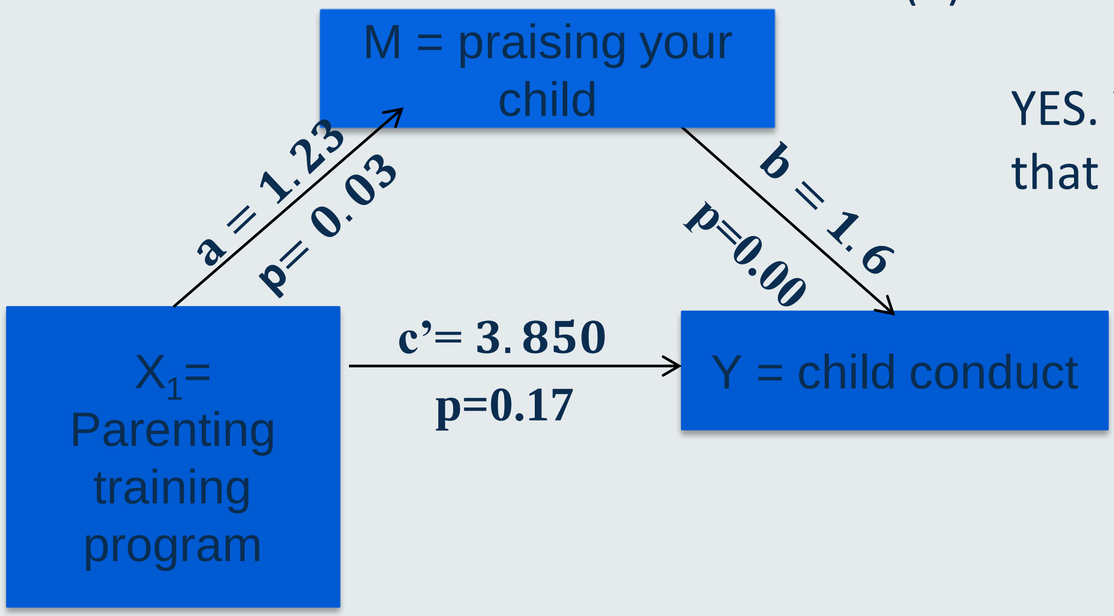
Q2: Given the path diagram below, Report:

(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 <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**