📘 WEEK 8 – Mediation Analysis (Baron & Kenny + Indirect Effects)

🔷 What is Mediation?

A **mediator (M)** is a variable that explains **how** or **why** an independent variable (X) influences a dependent variable (Y).

> Mediation describes **indirect pathways** in a causal model.

🔶 Mediation Model Breakdown:

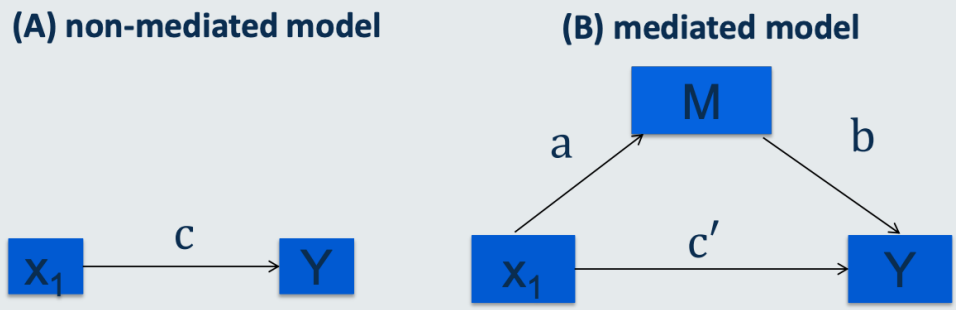
- **X → M → Y**
- Total effect = **c**
- Direct effect (X → Y, controlling for M) = **c′**
- Indirect effect (X → M → Y) = **a × b**

> So:
> **Total effect (c) = Direct effect (c′) + Indirect effect (a × b)**

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.



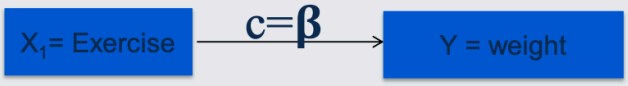
- In a **non-mediated model** (A), the **total effect** of the independent variable x_1 on the dependent Y is denoted by the path c
- Under a **mediated model** (B), the total causal effect c **can be split** into an indirect (or mediated) part with paths a and b and a direct (non-mediated) path c'

- **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.

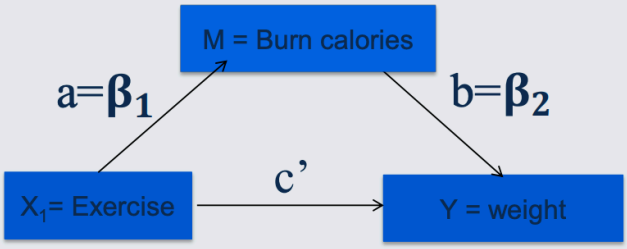
(a)



1. Estimate of **path c**:

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

(b)



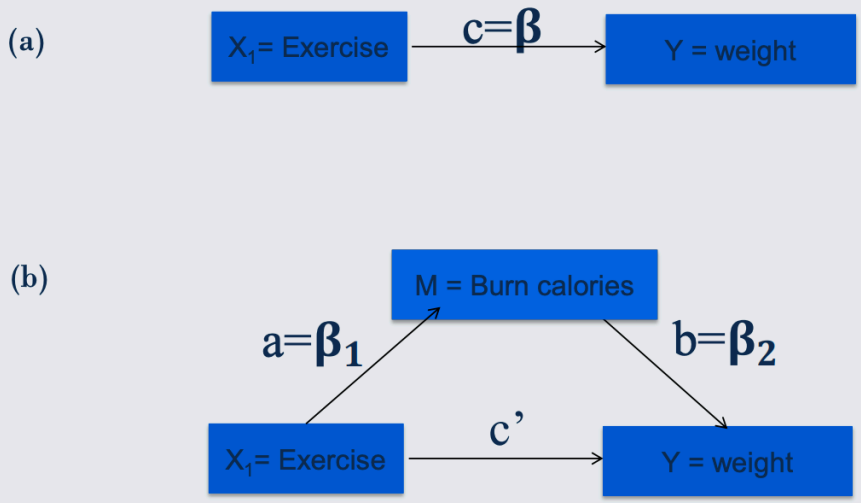
2. Estimate of **path a**:

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

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



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

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

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

ii. From step 3 model:

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

$c' = \beta_3$

🧪 Baron & Kenny's 4 Steps to Establish Mediation

Step	Path	Model Equation	Condition
1	c	$Y = \beta_0 + c \cdot X + \varepsilon$	X predicts Y
2	a	$M = \beta_0 + a \cdot X + \varepsilon$	X predicts M
3	b	$Y = \beta_0 + b \cdot M + c' \cdot X + \varepsilon$	M predicts Y (controlling for X)
4	c′	Same as step 3	c′ smaller than c. If c′ is not significant → complete mediation. If still significant → partial mediation

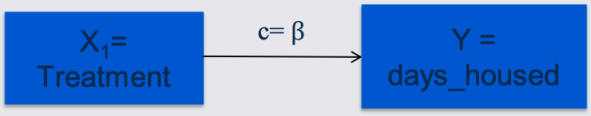
- If β_3 is not significantly different from 0 ($p>0.05$) there is **complete mediation.**
  - There is no association between X_1 and Y, when we control for M.
  - This will be the case if the direct effect (path c') drops to zero after controlling for M.
- If β_3 is significantly different from 0 ($p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

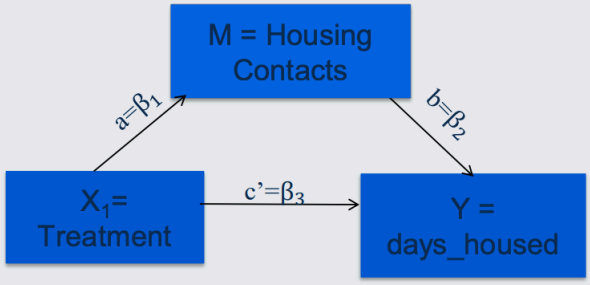
Which Baron and Kenny steps are essential to establish mediation?

Select one:

- a. Steps 2 and 3. ✓
- b. Steps 1 and 4.
- c. Steps 1 and 3.
- d. All steps.

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

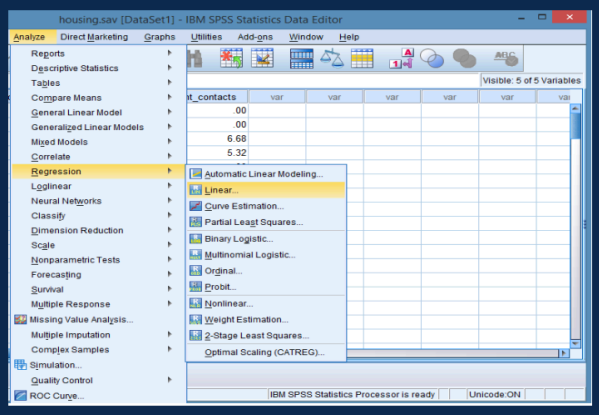
$Y = \beta_0 + \beta_2 M + \beta_3 X_1 + \varepsilon$

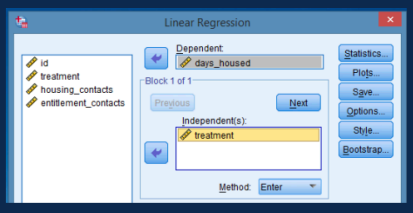
$\begin{cases} H_0: \boldsymbol{\beta_3} = 0 \\ H_1: \boldsymbol{\beta_3} \neq 0 \end{cases}$

SPSS Slide: 'How to' Steps

Computing three linear regression models from '**housing.sav**' data:

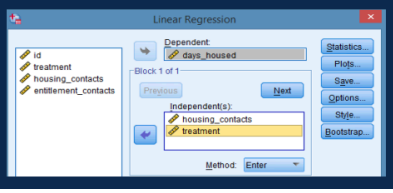
1) Use 'Analyse' -> 'Regression' -> 'Linear'
2) Drag and drop dependent, and independent variables.





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

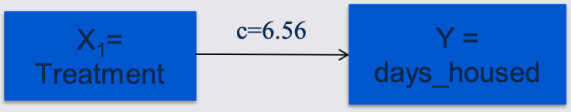
Path b

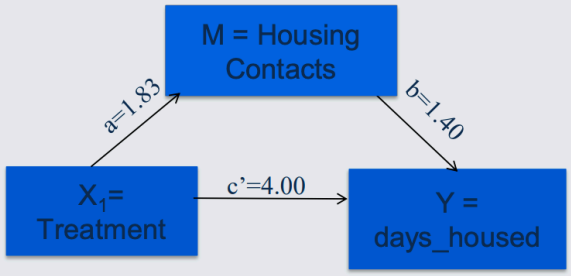
Path c'

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





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 → 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

<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 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

<u>Step 3</u>: Test path b ($M \rightarrow Y$, 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

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

✅ SPSS Commands for Baron & Kenny Approach

Step 1–3: Use Linear Regression

- `Analyze > Regression > Linear`
- Change outcome and predictors for each step.

Step 4: Examine output

- If c′ becomes non-significant → complete mediation
- If c′ remains significant but reduced → partial mediation

🧪 Testing the Indirect Effect

Baron & Kenny's steps are not statistical tests for the indirect effect. To formally test **a × b**, use:

✅ Sobel Test

- Requires standard errors (SE) of a and b
- Z = (a × b) / SE(ab), where:

$$SE(ab) = \sqrt{a^2 \cdot SE_b^2 + b^2 \cdot SE_a^2}$$

✅ Bootstrapping (Recommended)

- Doesn't assume normal distribution of ab
- Use **PROCESS macro** in SPSS

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

http://davidakenny.net/cm/mediate.htm

Newer methods (e.g. Sobel test) recommend testing only the indirect effect (paths a and b) to establish mediation

Term	Meaning
a	Effect of X on M
b	Effect of M on Y (controlling for X)
c	Total effect of X on Y
c′	Direct effect of X on Y (controlling for M)
a × b	Indirect (mediated) effect

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)

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

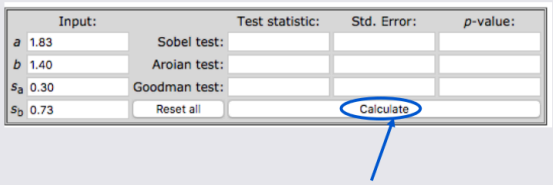
- Sobel statistic test is based on an approximate z-statistic, given by: $z = \frac{ab}{SE(ab)}$
- SE(ab) denotes the standard error of the estimated indirect effect, given by:

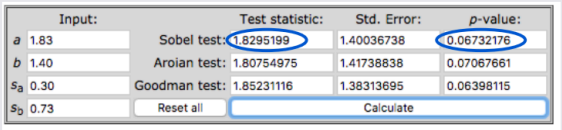
$$SE(ab) = \sqrt{a^2 S_b^2 + b^2 S_a^2}$$

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

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

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



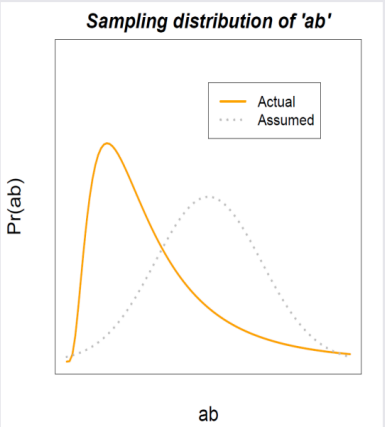


As Z (Sobel Test Statistic) in absolute value is less than 1.96, fail to reject the null hypothesis that the indirect effect is zero (p=0.067)

http://quantpsy.org/sobel/sobel.htm

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.**

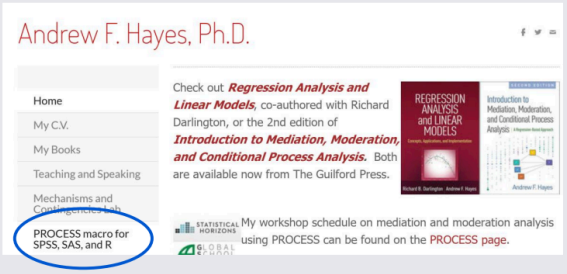
Bias-corrected Bootstrap

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

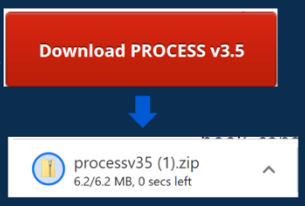
This can be done using the PROCESS macro (see Hayes & Rockwood, 2017), if installed in your SPSS

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

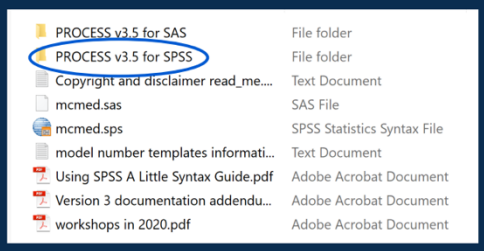


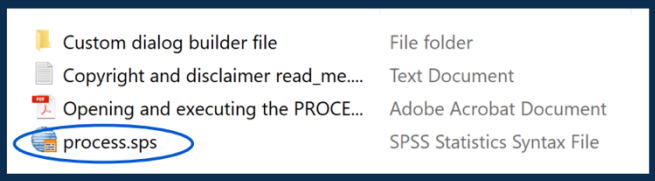
http://www.afhayes.com/

Process Macro 'how to' Option 1



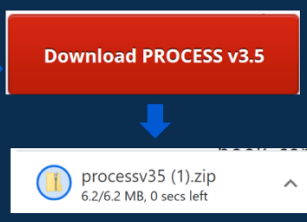
Click 'Download; Scroll down the page and Click 'Download PROCESS v3.5). Open the zip file Open the PROCESS v3,5 for SPSS folder.



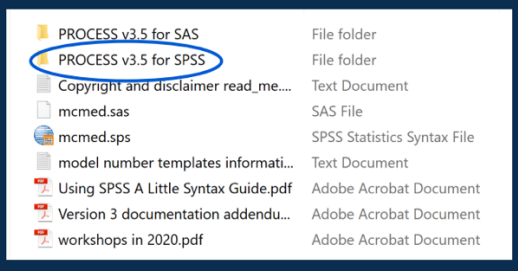


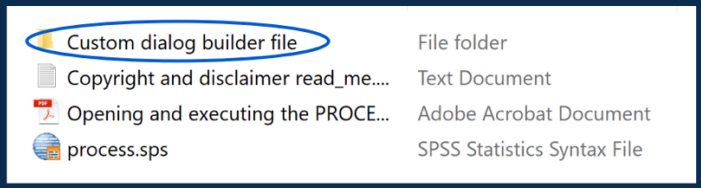
Extract and Open the process.sps file in a new syntax window in SPSS
Run all by , selecting all syntax clicking on the big green triangle

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.

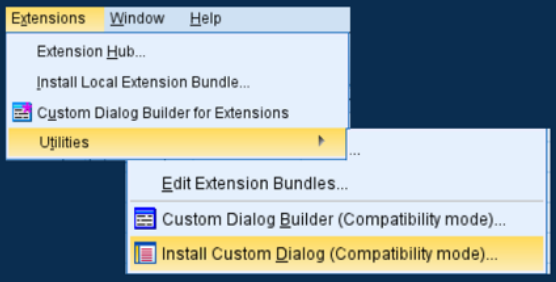


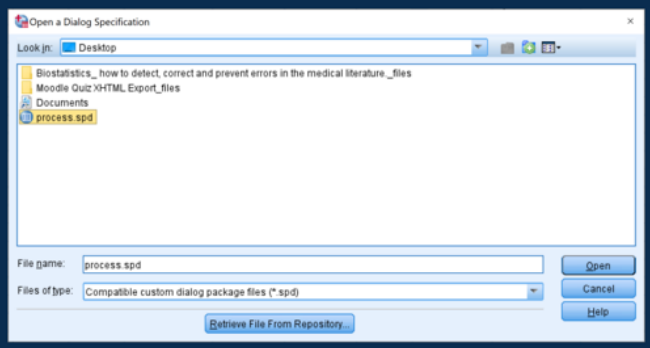


Process Macro 'how to' Option 2

Extract the process.spd file

Extensions → Utilities → Install Custom Dialog





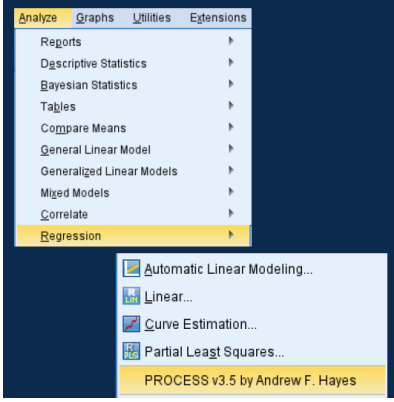
SPSS Slide: 'How to'

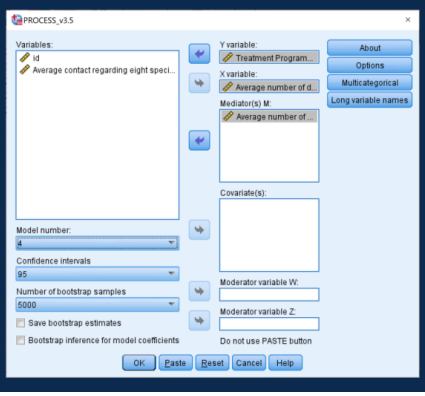
Use Lecture_8_data to test an indirect effect. In the regression menu you will see a new option PROCESS

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 eight 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.**

🧠 When Can You Say There Is Mediation?

Condition	Conclusion
a and b are significant	Mediation possible
c′ is **not significant**	**Complete** mediation
c′ is **reduced but still significant**	**Partial** mediation
a × b (indirect effect) ≠ 0 via test	Statistically supports mediation

Quiz:

Under a mediated model

Select one:

- a. The total effect of the independent variable X1 on the dependent variable Y is denoted by path c'. ×
- b. The total effect of the independent variable X1 on the dependant variable Y is denoted by the path a.
- c. A simple linear regression model can denote path b.
- d. The causal effect can be split into an indirect and direct effects.

Your answer is incorrect.

The causal effect can be split into an indirect and direct effect

The correct answer is: The causal effect can be split into an indirect and direct effects.

❗ Why your choice (a) is incorrect:

> **"The total effect... is denoted by path c′"**
> **But actually:**

- **c′** is the **direct effect**
- **c** is the **total effect**

✅ Clarifying the terms in mediation analysis

When we do mediation, we break the total effect of X on Y into:

Path	Meaning
a	Effect of X on the mediator M
b	Effect of M on Y (controlling for X)
c	Total effect of X on Y (without M in the model)
c′ (c-prime)	Direct effect of X on Y (when M is included in the model)

🔍 Key logic:

- The **total effect** is **c**
- The **indirect effect** is **a × b**
- The **direct effect** is **c′**

So, we say:

$$\text{Total effect (c)} = \text{Direct effect (c}\prime\text{)} + \text{Indirect effect (a} \times \text{b)}$$

That's why the correct answer is:

> **d. The causal effect can be split into an indirect and direct effect**

Topic 8 Knowledge Check Solutions

Q1: A mediator explains a portion (or all – if you have complete mediation) of the total association between Y and X1

Q2. Under a mediated model, the causal effect can be split into an indirect effect a*b and direct effect c’

Q3. Complete Mediation is determined when the path c’ (direct effect – coefficient for the variable x1) is not significantly different from 0

Partial Mediation is determined when the path c’ (direct effect – coefficient for the variable x1) is significantly different from 0

Q4: Which Baron and Kenny steps are essential to establish mediation

Steps 2 and 3 are essential

We can obtain path a, b and c’ from steps 2 and 3 and use these to calculate the total effect path c (which we would get from a simple linear regression (Step 1) involving the independent variable x1 and the dependent variable)

Q5. The Sobel test is used as a method for testing the indirect effect, this is a parametric test for the indirect effect.

Bonus: Mediation has occurred when: The relationship between the predictor and the outcome is completely wiped out when the mediator is included in the model.

When perfect mediation occurs, the relationship between the outcome and the predictor is lost, and all of the effect is through the mediator.