Mediation Analysis

Mark Andrews
Psychology Department, Nottingham Trent University

mark.andrews@ntu.ac.uk

Mediation

- In a mediation model, the effect of one variable x on another y is due to its effect on a third variable m, which then affects y.
- Changes in the variable x lead to changes in m that then lead to changes in y.
- As an example of a mediation effect, it is widely appreciated that tobacco smoking x raises the probability of lung cancer y, and that this effect is due to tar (tobacco residue) produced by the burning of the tobacco accumulating the lungs m.

Pure mediation

- In a *pure* or *full* mediation model, the effect of x on y is entirely due to its effect on m.

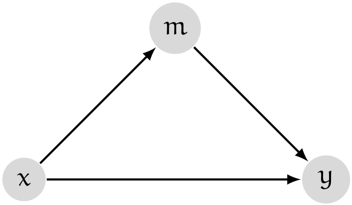
$$x \longrightarrow m \longrightarrow y$$

- Assuming that we are dealing with a normal linear model, we can write a pure mediation model as follows:

$$\begin{aligned}
\text{for } i \ldots 1 \ldots n, \quad y_i &\sim N(\mu_i^y, \sigma_y^2), \quad \mu_i^y = \beta_{y0} + \beta_{ym} m_i, \\
m_i &\sim N(\mu_i^m, \sigma_m^2), \quad \mu_i^m = \beta_{m0} + \beta_{mx} x_i,
\end{aligned}$$

Partial mediation

- In *partial mediation*, we assume that x affects m and m affects y as before, but there is also a direct effect of x on y, as in the following diagram.



- The partial mediation model can be written as follows.

$$\text{for } i \ldots 1 \ldots n, \quad y_i \sim N(\mu_i^y, \sigma_y^2), \quad \mu_i^y = \beta_{y0} + \beta_{ym} m_i + \beta_{yx} x_i,$$
$$m_i \sim N(\mu_i^m, \sigma_m^2), \quad \mu_i^m = \beta_{m0} + \beta_{mx} x_i,$$

Example 1

- In order to explore mediation models, let us begin with data generated according to a specific model.

```
N <- 100

b_m0 <- 1.25; b_mx <- 1.25;
b_y0 <- -0.5; b_ym <- 1.75; b_yx <- 0.75;
sigma_m <- 1.5; sigma_y <- 2.0

mediation_df <- tibble(
  x = rnorm(N, sd = 2),
  m = b_m0 + b_mx * x + rnorm(N, sd = sigma_m),
  y = b_y0 + b_ym * m + b_yx * x + rnorm(N, sd = sigma_y)
)
```

Using lavaan

- Let us now set up this model using `lavaan`.

```
library(lavaan)

mediation_model_spec_1 <- '
y ~ m + x
m ~ x
'
```

- For example, by writing `y ~ m`, we are assuming that for each i, $y_i = \beta_{y0} + \beta_{ym} m_i + \epsilon_i^y$.
- However, by default, unless we explicitly state in the formula that we are using an intercept term, as follows, we will not get information about it.

```
mediation_model_spec_1 <- '
y ~ 1 + m + x
m ~ 1 + x
'
```

Using lavaan

- Now we call `lavaan::sem` with reference to `mediation_model_spec_1`, and this fits the model using maximum likelihood estimation.

```
mediation_model_1 <- sem(mediation_model_spec_1,
                         data = mediation_df)

parameterEstimates(mediation_model_1)
```

```
##   lhs op rhs    est    se      z pvalue ci.lower ci.upper
## 1   y ~1      -0.493 0.243 -2.030  0.042   -0.970   -0.017
## 2   y  ~   m  1.843 0.118 15.648  0.000    1.612    2.073
## 3   y  ~   x  0.639 0.180  3.558  0.000    0.287    0.991
## 4   m ~1       1.256 0.164  7.665  0.000    0.935    1.577
## 5   m  ~   x  1.280 0.083 15.471  0.000    1.118    1.443
## 6   y ~~   y  3.633 0.514  7.071  0.000    2.626    4.640
## 7   m ~~   m  2.620 0.371  7.071  0.000    1.894    3.347
## 8   x ~~   x  3.826 0.000     NA     NA    3.826    3.826
## 9   x ~1       0.305 0.000     NA     NA    0.305    0.305
```

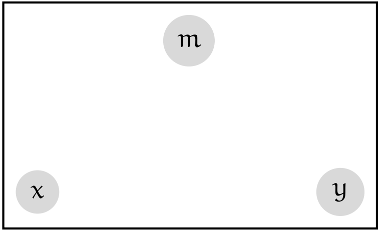
Model comparison

- In mediation analysis, a major aim is evaluating first whether there is evidence of a mediation of the effect of x on y by m, and then whether this is pure or partial mediation.
- To do so, we first specify and then fit the full mediation model‘.

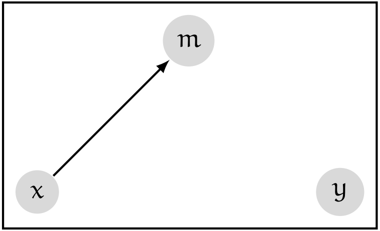
```
mediation_model_spec_0 <- '
  y ~ 1 + m
  m ~ 1 + x
'
mediation_model_0 <- sem(mediation_model_spec_0,
                         data = mediation_df)

mediation_models <- c(model_0 = mediation_model_0,
                      model_1 = mediation_model_1)

map_dbl(mediation_models, AIC)
```

```
##  model_0  model_1
## 816.8326 806.9121
```

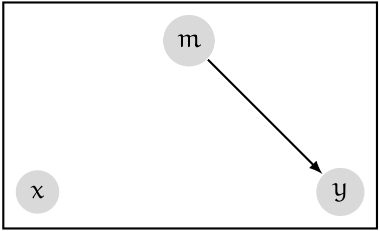
Alternative models: 1 to 4



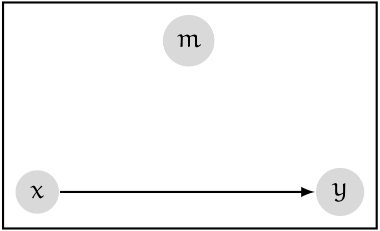
(a)



(b)

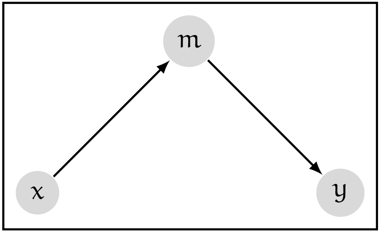


(c)

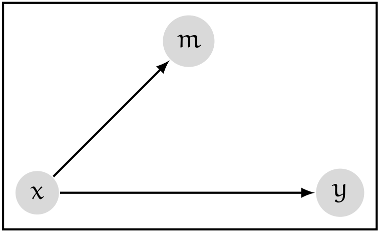


(d)

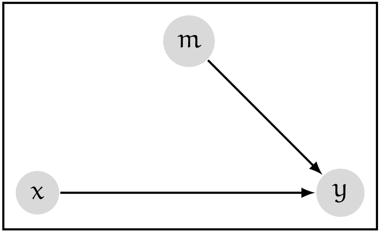
Alternative models: 5 to 8



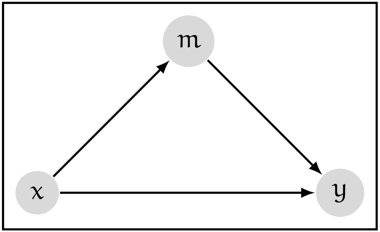
(e)



(f)



(g)



(h)

Model evalation

- Read alternative models specifications from file:

```
source('mediation_models_specs.R')
```

- Having each model specified as an element of a list, we now use `purrr::map` to fit each model and calculate its AIC score.

```
mediation_models <- map(mediation_models_specs,
                        ~sem(., data = mediation_df)
)

map_dbl(mediation_models, AIC) %>%
  sort()
```

```
##  model_h  model_e  model_g  model_f  model_c  model_d  model_
## 1228.884 1238.804 1349.074 1350.682 1358.995 1470.873 1480.70
```

Model evaluation

- Let us now generate some data from a pure mediation model, and then fit all 8 possible versions of the mediation model to the data, and evaluate the fit.

```
mediation_df_new <- tibble(
  x = rnorm(N, sd = 2),
  m = b_m0 + b_mx * x + rnorm(N, sd = sigma_m),
  y = b_y0 + b_ym * m + rnorm(N, sd = sigma_y)
)

mediation_models_new <- map(mediation_models_specs,
                            ~sem(., data = mediation_df_new)
)

map_dbl(mediation_models_new, AIC) %>%
  sort()
```

```
##  model_e  model_h  model_f  model_c  model_g  model_b  model_
## 1268.894 1270.717 1363.463 1401.127 1402.949 1466.873 1495.69
```

Direct versus indirect effects

- In a standard linear regression model of the following kind

$$y_i \sim N(\mu_i, \sigma^2), \quad \mu_i = \beta_0 + \beta_1 x_i, \quad i \in 1 \ldots n,$$

  a change in any x_i by 1 unit, i.e., $x_i + 1$, would always lead to a change of β_1 in the expected, i.e. the average, value of the outcome variable.
- This is easy to see. Let $x'_i = x_i + 1$, and

$$\mu_i = \beta_0 + \beta_1 x_i, \quad \begin{aligned} \mu'_i &= \beta_0 + \beta_1 x'_i, \\ &= \beta_0 + \beta_1 (x_i + 1), \\ &= \beta_0 + \beta_1 x_i + \beta_1, \\ &= \mu_i + \beta_1, \end{aligned}$$

  and so $\mu' - \mu = \beta_1$.
- Regardless of how many predictor variables there are in the linear regression, a change in predictor k by one unit, always leads to a change β_k in the average value of the outcome variable.

Direct versus indirect effects

- In a mediation model, the effect of a change in the predictor x on the outcome y is not as simple.
- In pure mediation model, we can write each y_i and m_i as follows

$$
\begin{aligned}
y_i &= \beta_{y0} + \beta_{ym} m_i + \epsilon_i^y, \quad \epsilon_i^y \sim N(0, \sigma_y^2),\\
m_i &= \beta_{m0} + \beta_{mx} x_i + \epsilon_i^m, \quad \epsilon_i^m \sim N(0, \sigma_m^2).
\end{aligned}
$$

  From this, we have

$$
\begin{aligned}
y_i &= \beta_{y0} + \beta_{ym} \left(\beta_{m0} + \beta_{mx} x_i + \epsilon_i^m\right) + \epsilon_i^y,\\
&= \beta_{y0} + \beta_{ym}\beta_{m0} + \beta_{ym}\beta_{mx} x_i + \beta_{ym}\epsilon_i^m + \epsilon_i^y,
\end{aligned}
$$

  and this entails

$$
y_i \sim N(\mu_i, \beta_{ym}^2 \sigma_m^2 + \sigma_y^2), \quad \mu_i = \beta_{y0} + \beta_{ym}\beta_{m0} + \beta_{ym}\beta_{mx} x_i.
$$

- Following the same reasoning as above for the case of standard linear regression, this entails that in a pure mediation model a unit change in x_i leads to a change of $\beta_{ym}\beta_{mx}$ in the expected value of y.

Direct versus indirect effects

- In the case of the partial mediation model, we saw already that each y_i and m_i in the model can be defined as follows:

$$
\begin{aligned}
y_i &= \beta_{y0} + \beta_{ym} m_i + \beta_{yx} x_i + \epsilon_i^y, \quad \epsilon_i^y \sim N(0, \sigma_y^2), \\
m_i &= \beta_{m0} + \beta_{mx} x_i + \epsilon_i^m, \quad \epsilon_i^m \sim N(0, \sigma_m^2).
\end{aligned}
$$

From this, we have

$$
\begin{aligned}
y_i &= \beta_{y0} + \beta_{ym}(\beta_{m0} + \beta_{mx} x_i + \epsilon_i^m) + \beta_{yx} x_i + \epsilon_i^y, \\
y_i &= \beta_{y0} + \beta_{ym}\beta_{m0} + \beta_{ym}\beta_{mx} x_i + \beta_{yx} x_i + \beta_{ym}\epsilon_i^m + \epsilon_i^y,
\end{aligned}
$$

which entails

$$
y_i \sim N(\mu_i, \beta_{ym}^2 \sigma_m^2 + \sigma_y^2), \quad \mu_i = \beta_{y0} + \beta_{ym}\beta_{m0} + (\beta_{ym}\beta_{mx} + \beta_{yx}) x_i,
$$

and following the reasoning above, this entails that unit change in x_i leads to a change of $(\beta_{ym}\beta_{mx} + \beta_{yx})$ in the expected values of y_i.

Direct versus indirect effects

In general in a mediation model, we have following:

$$\underbrace{\overbrace{\beta_{ym}\beta_{mx}}^{\text{indirect effect}} + \overbrace{\beta_{yx}}^{\text{direct effect}}}_{\text{total effect}}.$$

If there is no direct effect, as would be the case in pure mediation model, then the total effect is equal to the indirect effect.

Direct versus indirect effects

- In a `lavaan` mediation model, we can create single variables that measure the direct, indirect and total effects. *To do so, we must first use labels for our original parameters, i.e. the coefficients, and then use the `:=` operator to create new variables that are functions of the original parameters.

```
mediation_model_spec_1 <- '
y ~ 1 + b_ym * m + b_yx * x
m ~ 1 + b_mx * x

# Define effects
indirect := b_ym * b_mx
direct   := b_yx
total    := b_yx + (b_ym * b_mx)
'
```

Direct versus indirect effects

- We can fit this model as per usual.

```
mediation_model_1 <- sem(mediation_model_spec_1,
                         data = mediation_df)
```

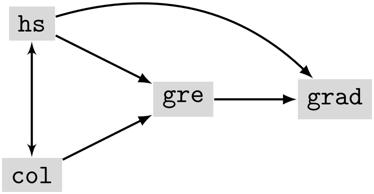
In the usual parameter estimates output, we can use `dplyr::filter` to isolate these effects:

```
parameterEstimates(mediation_model_1) %>%
  filter(label %in% c('indirect', 'direct', 'total')) %>%
  select(label, est,pvalue:ci.upper)
```

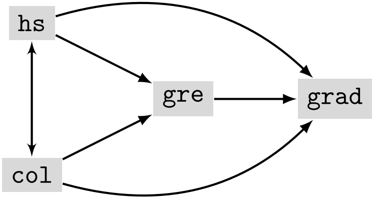
```
##      label       est       pvalue  ci.lower  ci.upper
## 1 indirect 2.3591526 0.0000000000 1.9388750 2.7794303
## 2   direct 0.6387239 0.0003735514 0.2868853 0.9905624
## 3    total 2.9978765 0.0000000000 2.6431974 3.3525556
```

As we can see, for example, the estimated effect for the total effect is 2.998, and the 95% confidence interval on this effect is $(2.643, 3.353)$.

Modelling Graduate School Performance



(i)



(j)

Modelling Graduate School Performance

```r
grad_mediation_models_specs <- within(list(),{
  model_0 <- '
      grad ~ hs + gre
      gre ~ hs + col
  '

  model_1 <- '
      grad ~ hs + b_grad_gre*gre + b_grad_col*col
      gre ~ hs + b_gre_col*col

      # labels for indirect, direct, and total
      direct := b_grad_col
      indirect := b_gre_col*b_grad_gre
      total := b_grad_col + (b_gre_col*b_grad_gre)
'
})
```