Data Analysis and Machine-Learning

Chapter 11.5.

*Mediation Effect Analysis – Application*



1. 3-Steps Conceptualization of Mediation Effect (Baron & Kenny, 1986)

Mediation effect of a model can be validated when,

for:

①

②

③

[Equation 1 (※X→M)] is statistically significant (since X is to effect Y via M), and,

[Equation 2 (※X→Y)] is statistically significant (since if X does not explain Y, mediation relationship cannot hold), and,

for [Equation 3 (※X+M→Y)]:

(1) if X is not statistically significant (c’=0) → Full Mediation

(2) if X is statistically significant, with 0<c’<c → Partial Mediation

2. Mediating Effect (a\*b) and the Sobel Test

For above conceptualization, total effect for partial mediation can be calculated as the sum of direct effect and indirect effect, as:

And for full mediation:

Logically, therefore, this implies that there is always an overlapping variance between the independent variable X and mediation variable M in explaining the variance of independent variable Y. Sobel test is designed to measure this variance, i.e., the indirect effect using standard errors (z-value = a\*b/SQRT(b2\*sa2 + a2\*sb2).

With the null hypothesis of the Sobel test set as a\*b=0, the alternative hypothesis (p<.05) is that the mediating effect is statistically significant.

3. Demonstration

For Bootstrapping (as possible alternative to the limitations of relying on the Sobel Test), at least 1,000 to 10,000 is usually considered appropriate.

For demonstration, Student Alcohol Consumption data from Kaggle is used as the dataset, with:

Independent Variable: Dalc (workday alcohol consumption: 1~5)

Mediation Variable: Studytime (weekly study time)

Dependent Variable: Failures (number of class failures)

What we want to test, therefore, would be:

If students drink during workdays, total weekly study time would decrease, which would result in the increasing number of class failures.

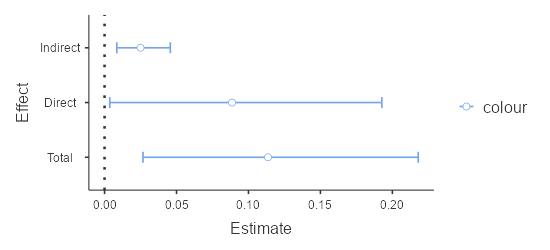
4.1. Jamovi MedMod with Bootstrapping Result

**Results**

**Mediation**

| Mediation Estimates | | | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  | | | |  |  | | |  |  |  | |  | |  | |
| **Effect** | | **Estimate** | | | | | **SE** | | | | **Z** | | | **p** | | | |
| Indirect |  | 0.0250 | | | |  | 0.00988 | | |  | 2.53 |  | | 0.011 | |  | |
| Direct |  | 0.0886 | | | |  | 0.04875 | | |  | 1.82 |  | | 0.069 | |  | |
| Total |  | 0.1136 | | | |  | 0.04917 | | |  | 2.31 |  | | 0.021 | |  | |
|  | | | | | | | | | | | | | | | | | |
| Path Estimates | | | | | | | | | | | | | | | | | | | | | | |
|  | |  |  |  |  | | |  |  | | | |  | |  | |  | |  |  |  |  |
|  | | |  | |  | | | | **Estimate** | | | | | | **SE** | | | | **Z** | | **p** | |
| Dalc | |  | → |  | studytime | | |  | -0.1847 | | | |  | | 0.0476 | |  | | -3.88 |  | < .001 |  |
| studytime | |  | → |  | failures | | |  | -0.1354 | | | |  | | 0.0409 | |  | | -3.31 |  | < .001 |  |
| Dalc | |  | → |  | failures | | |  | 0.0886 | | | |  | | 0.0488 | |  | | 1.82 |  | 0.069 |  |
|  | | | | | | | | | | | | | | | | | | | | | | |

**Estimate Plot**



4.2. Conducting Manually with three linear regressions

**Linear Regression**

| Model Fit Measures | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  | | |  |  | | |  |
| **Model** | | **R** | | | | **R²** | | | |
| 1 |  | 0.196 | | |  | 0.0384 | | |  |
|  | | | | | | | | | |
| Model Coefficients - studytime | | | | | | | | | | | | | | | |
|  | | |  |  | | |  |  | | |  |  |  |  |  |
| **Predictor** | | | | **Estimate** | | | | **SE** | | | | **t** | | **p** | |
| Intercept | | |  | 2.309 | | |  | 0.0805 | | |  | 28.68 |  | < .001 |  |
| Dalc | | |  | -0.185 | | |  | 0.0466 | | |  | -3.96 |  | < .001 |  |
|  | | | | | | | | | | | | | | | |

# Linear Regression

| Model Fit Measures | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  | | |  |  | | |  |
| **Model** | | **R** | | | | **R²** | | | |
| 1 |  | 0.136 | | |  | 0.0185 | | |  |
|  | | | | | | | | | |
| Model Coefficients - failures | | | | | | | | | | | | | | | |
|  | | |  |  | | |  |  | | |  |  |  |  |  |
| **Predictor** | | | | **Estimate** | | | | **SE** | | | | **t** | | **p** | |
| Intercept | | |  | 0.166 | | |  | 0.0721 | | |  | 2.30 |  | 0.022 |  |
| Dalc | | |  | 0.114 | | |  | 0.0417 | | |  | 2.72 |  | 0.007 |  |
|  | | | | | | | | | | | | | | | |

# Linear Regression

| Model Fit Measures | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  | | |  |  | | |  |
| **Model** | | **R** | | | | **R²** | | | |
| 1 |  | 0.202 | | |  | 0.0409 | | |  |
|  | | | | | | | | | |
| Model Coefficients - failures | | | | | | | | | | | | | | | |
|  | | |  |  | | |  |  | | |  |  |  |  |  |
| **Predictor** | | | | **Estimate** | | | | **SE** | | | | **t** | | **p** | |
| Intercept | | |  | 0.4785 | | |  | 0.1255 | | |  | 3.81 |  | < .001 |  |
| Dalc | | |  | 0.0886 | | |  | 0.0421 | | |  | 2.10 |  | 0.036 |  |
| studytime | | |  | -0.1354 | | |  | 0.0447 | | |  | -3.03 |  | 0.003 |  |
|  | | | | | | | | | | | | | | | |

As the model results imply, partial mediation is detected, with:

Total Effects (direct + indirect effect) = **-.185** (1-step regression coefficient of the independent variable) \* **-.1354** (3-step regression coefficient of the mediating variable) + **.0886** (3-step regression coefficient of the independent variable) = **.1136** (which is equal to the ‘total’ estimation calculated via Jamovi MedMod)

For Sobel Test, we need to know:

a-coefficient = -.185

b-coefficient = -.1354

Sa (standard error of the a-coefficient) = .0466

Sb (standard error of the b-coefficient) = .0447

which, throughout the calculation gives us the value of:

Test statistic: 2.41 (p<.05)

Thus, it is possible to reject the null hypothesis and state that the mediating effect (a\*b) is statistically significant.