2019.10 月

資料科學下的因果機制的探索—中介效應分析統計方法與軟體簡介

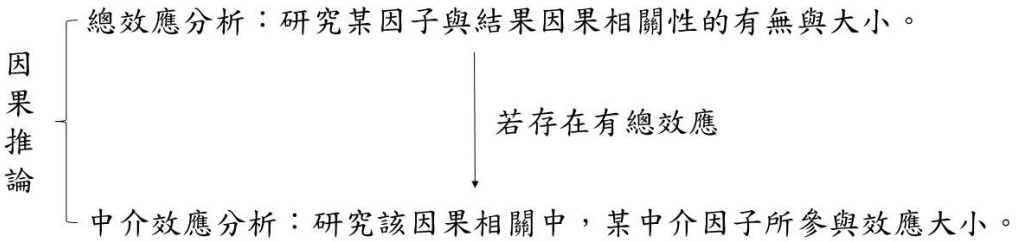
林聖軒 醫師/博士

國立交通大學統計學研究所助理教授

1. 中介效應分析簡介

因果推論為近代探索變數間因果機制的熱門方法之一，而因果推論是從「反事實架構」(Counterfactual framework)的角度，以基礎的邏輯與數理方法，估計暴露因子(Exposure)造成結果變量(Outcome)之因果關係。當暴露因子與結果變量的因果關係(在本文中概稱為「總效應」)確立以後，下一個重要的問題則是進一步探討總效應是透過哪些機制所造成、以及每種機制所解釋的比例為何。由於每種機制皆透過中介因子(Mediator)所代表，因此這種機制分析又稱為因果中介效應分析(Causal mediation analysis)。

因果中介效應分析最早由 Baron 與 Kenny 在 1986 年所提出，其包含差異法、結構方程式(structural equation modeling)之路徑分析(path analysis)[1]、以及因果中介效應分析(causal mediation analysis)[2-9]，而前兩者局限於線性模型，對於非線性結構必須使用根據因果推論學發展的因果中介效應分析(causal mediation analysis)方能得到具有因果機制意義的估計量[2-9]。因果推論與因果中介效應分析之關聯如圖一所述。

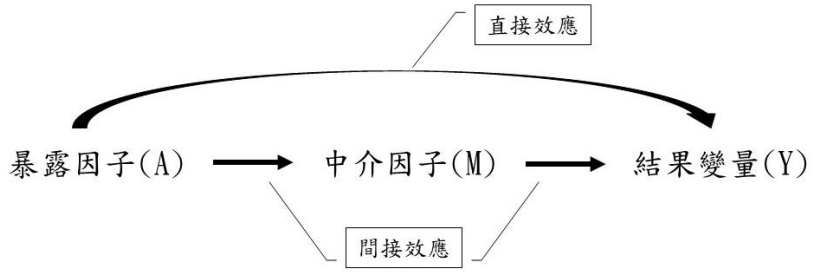


圖一 因果推論與中介效應分析之關聯。

在此我們將舉一個簡單的例子說明何謂中介因子與相關的效應解釋。假設我

們想探討機制的總效應為「抽菸是否導致死亡」，此例所提到的抽菸及死亡分別為暴露因子與結果變量。在過去的文獻中我們得知罹患肺癌可能是抽菸導致死亡的重要機制；不過除了肺癌以外，抽菸也有可能透過其他的機制導致死亡：例如食道癌、頭頸癌、或慢性阻塞性肺病等。假設我們想要知道抽菸透過肺癌導致死亡所占總效應的比例，就將肺癌視為中介因子；使用中介效應分析，將總效應分解成透過中介因子所造成的間接效應(indirect effect，或稱中介效應 mediation effect)以及不透過中介因子所造成的直接效應(direct effect，或稱替代效應 alternative effect)，如圖二的因果圖所示。間接效應與總效應的比值稱作中介比例(proportion mediated, **PM**)，代表能夠被該中介因子所解釋的比例。如果比例越接近 1，代表該中介因子所代表的路徑可以完全解釋總效應所有的機制；如果比例接近 0，則代表該中介因子所代表的路徑並不是總效應的機制。中介效應分析的目的便是要盡可能地蒐集到能夠代表所有路徑的中介因子，使中介比例接近 100%。這樣即意味著所有的路徑能夠用這些中介因子所代表的路徑所解釋。

要估計直接效應與間接效應，須建立在一些假設上，學者 VanderWeele 及 Vansteelandt 與 Imai 等人分別於 2009、2010 年，提出因果中介分析所需要的辨識假設[9, 10]，其中包含了暴露因子、中介因子及結果變量之間所有的干擾因子都要被測量且校正才能達成。這個相對於要估計總效應的假設(只需要校正「暴露因子及結果變量之間所有的干擾因子」)還需要校正更多的干擾因子。



圖二 直接效應及間接效應在中介效應分析之關係圖。

2. 單一中介因子進行因果中介分析之軟體簡介(mediation.r)

為了資料分析上的便利，已有學者將單中介因子的模型程式化在統計軟體R上，其名稱為 mediation.r[11]。對於不同的資料型態，該軟體都能支援並計算出間接效應、直接效應、以及中介比例。諸如傳統生物醫學研究常關注的存活率、機密工業上的產品使用年限以及各式各樣的連續型變量都可運用此軟體來探討其因果關係。以下將介紹上述軟體之操作。

mediation.r 為R軟體中的套件，透過以下四個步驟完成套件下載、模型設定、以及中介效應分析。以下所使用的範例為 Brader 等人於 2008 年進行的一項隨機實驗[12]，此實驗的暴露因子為受試者有無接觸有關移民的不同媒體報導(treat)，以焦慮為中介因子(emo)，結果變量為移民政策的態度和政治行為(cong_mesg)。該範例想要想要進行因果中介效應分析以得知的問題是：由於不同媒體報導影響移民政策的態度和政治行為的總效應中，有多少比例能夠由「媒體報導導致焦慮、而焦慮導致影響移民政策的態度和政治行為」這樣的機制所解釋？

步驟 1. 套件下載

```
#於統計軟體 R 中，輸入以下指令
> install.packages("mediation")
> library("mediation")
```

步驟 2. 輸入資料數據

```
#以下為軟體提供之範例數據(數據名稱是"framing")
> data("framing", package = "mediation")
```

步驟 3. 設定中介因子以及結果變量之統計模型

```
#中介因子之模型設定：本範例中中介因子使用線性模型。
> med.fit <- lm(emo ~ treat + age + educ + gender + income, data= framing)
```

#結果變量之模型設定：本範例中結果變量是二元變量、因此使用廣義線性模型中的邏輯斯回歸。

> out.fit <- glm(cong_mesg ~ emo + treat + age + educ + gender + income, data = framing, family= binomial("logit"))

此套件至當前版本適用於許多統計模型，其統計模型種類如表一。

	結果變量之統計模型						
中介因子之統計模型	**Linear**	**GLM**	**Ordered**	**Censored**	**Quantile**	**GAM**	**Survival**
Linear (lm/lmer)	✔	✔	✔*	✔	✔	✔*	✔
GLM (glm/bayesglm/glmer)	✔	✔	✔*	✔	✔	✔*	✔
Ordered (polr/bayespolr)	✔	✔	✔*	✔	✔	✔*	✔
Censored (tobit via vglm)	-	-	-	-	-	-	-
Quantile (rq)	✔*	✔*	✔*	✔*	✔*	✔*	✔*
GAM (gam)	✔*	✔*	✔*	✔*	✔*	✔*	✔*
Survival (survreg)	✔	✔	✔*	✔	✔	✔*	✔

表一 mediation.r 套件適用之統計模型種類。本表格改編自 Tingley et al 之論文圖表[11]。

步驟 4. 運用函數 mediate( )進行因果中介效應分析

#運用函數 mediate( )，並透過函數 summary( )，可得到中介效應分析之結果，包含

> med.out <- mediate(med.fit, out.fit, treat = "treat", mediator = "emo")

> summary(med.out)

使用 Tingley et al 論文[11]提供範例的結果如表二。

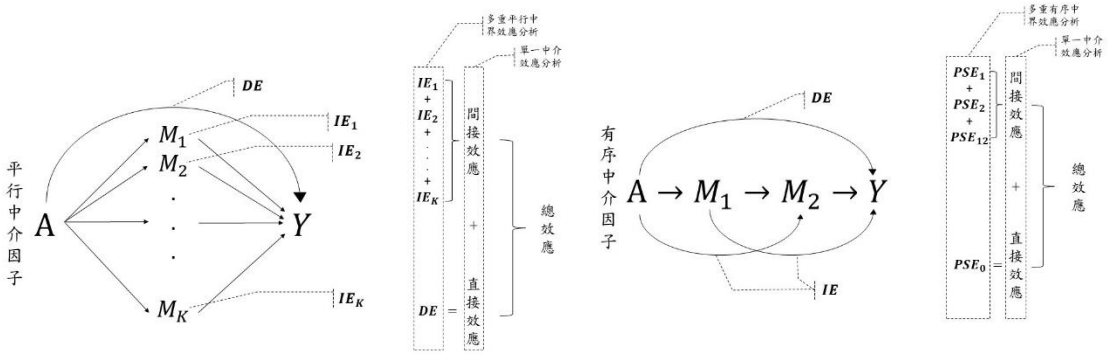
	Estimate	95% CI Lower	95% CI Upper	p-value
ACME(control)	0.0791	0.0351	0.15	<0.01
ACME(treated)	0.0804	0.0367	0.16	<0.01
ADE(control)	0.0206	-0.0976	0.12	0.70
ADE(treated)	0.0218	-0.1053	0.12	0.70
Total Effect	0.1009	-0.0497	0.23	0.14
Prop. Mediated(control)	0.6946	-6.3109	3.68	0.14
Prop. Mediated(treated)	0.7118	-5.793	3.50	0.14
ACME(average)	0.0798	0.0359	0.15	<0.01
ADE(average)	0.0212	-0.1014	0.12	0.70
Prop. Mediated(average)	0.7032	-6.0523	3.59	0.14

表二 範例使用 mediation.r 進行因果中介效應分析之結果圖表。習慣上間接效應使用 ACME(treated) 之數值、直接效應使用 ADE(control) 之數值、總效應使用 Total Effect 之數值、中介比例使用 Prop. Mediated(treated)之數值。

由表二可得，間接效應(ACME(treated))之 p-value 達到統計上顯著，亦即讓受試者接觸有關移民的不同媒體報導，可能會增強情緒焦慮反應，從而使受試者更有可能向其國會議員發送信息。而中介比例高達七成(Prop. Mediated(treated) = 0.7118)，代表該路徑可以解釋「不同媒體報導影響移民政策的態度和政治行為的總效應」所有路徑中的七成。

3. 因果多重中介效應分析(causal multi-mediation analysis)

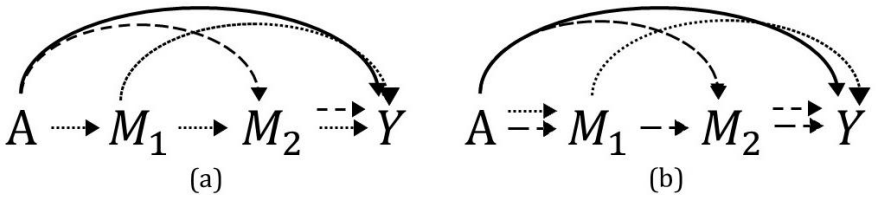
傳統的中介模型著重在單一中介因子，但是許多科學研究中的總效應很難由單一中介因子所解釋。隨著資料蒐集技術的成熟，能獲得更多的變量資料。因此提供了多重中介效應分析的發展契機。如圖三所示，K 個多重中介因子$\mathbf{M} = (M_1, M_2, \ldots, M_{K-1}, M_K)$又依照多重因子彼此之間關係的獨立性與否可以分成「平行中介因子(parallel mediator)」以及「有序中介因子(ordering mediator)」兩類，前者為相互平行的多重中介因子，後者為橫向多重且隨時間變化之中介因子。分析則有三種策略：(1) 單一中介效應分析：將$\boldsymbol{M}$為整體而造成的結果變量，並沒有探討$(M_1, M_2, \ldots, M_{K-1}, M_K)$間之因果關係，視所有平行中介因子為多變量之單一中介因子，將總效應分解成與平行中介因子有關之間接效應以及與平行中介因子無關之直接效應。(2)多重平行中介效應分析：把中介效應進一步區分為 K 個不同平行中介因子所代表的K個間接效應(IE_1~IE_K)。(3)有序多重中介效應分析：將$(M_1, M_2, \ldots, M_{K-1}, M_K)$之間因果關係納入分析，並根據 K 個不同有序中介因子的參與與否，把總效應進一步區分為2^K個所代表的 K 個特定路徑效應(Path Specific Effect, PSE)。



圖三 多重中介因子之中介效應。

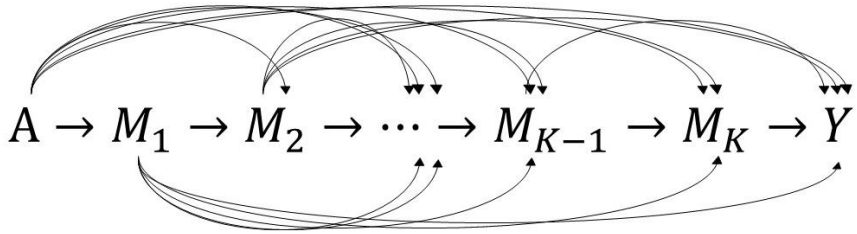
傳統上計算特定路徑效應是使用結構方程式路徑分析, SEM) ，在社會科學中較廣泛地被應用。然而 SEM 只能適用於所有的變項皆為連續變數方能得到符合因果推論下定義的特定路徑效應之不偏估計，因此較無法適用於生物醫學的資

料分析。因此 Avin et al.和 VanderWeele et al.發展出非參數式(Nonparametrically)方法來識別(Identification)部分的特定路徑效應(如圖四(a))，不過以上學者所提出的方法無法藉由非參數式的方法去識別所有的特定路徑效應。



圖四 中介因子數量為二之特定路徑效應。(a)可分為三種特定路徑效應：A → Y，由 A 不透過中介因子到 Y 之因果效應(實線路徑)；A → M_2 → Y，由 A 透過M_2到 Y 之因果效應(虛線路徑)；A → M_1 → Y 或 A → M_1 → M_2 → Y，由 A 透過M_1到 Y 之因果效應(點線路徑)。(b)可分為四種特定路徑效應：A → Y，由A不透過中介因子到 Y 之因果效應(實線路徑，PSE_0)；A → M_1 → Y，由 A 僅透過M_1到Y之因果效應(點線路徑，PSE_1)；A → M_2 → Y，由 A 僅透過M_2到Y之因果效應(虛線路徑，PSE_2)；A → M_1 → M_2 → Y，由 A 透過M_1及M_2到Y之因果效應(長虛線路徑，PSE_{12})。

為了解決此問題， Lin 和 VanderWeele 引入隨機介入(Itervention)的方法，便能獲得所有的特定路徑效應(如圖四(b)) [13]。在此假設只有兩個中介因子，即K = 2，所有的特定路徑效應分別為(1)與兩個中介因子皆無關聯的效應(PSE_0，即直接效應)，(2)只透過中介因子 1 參與的效應(PSE_1)，(3)只透過中介因子 2 參與的效應(PSE_2)，以及(4)兩個中介因子皆有參與的效應(PSE_{12})，如圖五(b)。Lin 更於 2019 年將兩個中介因子拓展至多重中介因子，並提供了使用者可操作分析之軟體在有序多重中介因子且多干擾因子模型複雜的架構之下，如圖五所示[14]。



圖五 有序多重中介因子之有向非循環圖。A為暴露因子，**M** = $(M_1, M_2, \ldots, M_{K-1}, M_K)$為有序多重中介因子， Y為結果變量。

4. 多重中介效應分析之軟體應用

Lin 的團隊將 mediation.r 軟體改編，使之能夠進行因果多重中介因子分析。更進一步發展使用者友善介面(http://shenghsuanlin.blog.nctu.edu.tw/)，此介面不需要編寫程式，僅透過拖拉點選的方式，即可分析出總效應(TE)及特定路徑效應(PSE)等數值，以下略述該使用者友善介面之操作步驟。

步驟一. 資料讀取

將所需之資料讀入。本軟體可針對缺失值可以使用“Deal with Missing value”功能進行兩種前處理：(1)將遺失值移除或(2)利用“Multiple Imputation”法插補缺失值。

步驟二. 選取暴露因子與干擾因子

步驟三. 選取多重中介因子以及對應的統計模型

以下拉選單方式，填入暴露因子、多重中介因子，其中可以選擇線性模型或羅吉斯回歸模型；並可於“Select Interaction Term”處選擇是否分析交互作用項；

步驟四. 選取結果變項

以下拉選單方式，填入結果變項以及相對應之統計模型。最後點選分析(Analysis)，即可獲得分析結果，且此結果可供使用者下載成 excel 檔。其分析結果如表三所示。

	Estimate	SE	95% CI		p-value
PSE0	0.024	0.005	0.014	0.033	<0.001
PSE1	0.039	0.002	0.035	0.042	<0.001
PSE2	0.058	0.003	0.053	0.064	<0.001
PSE12	0.054	0.003	0.049	0.059	<0.001
Total effect	0.175	0.006	0.164	0.186	<0.001
Prob. PSE0	0.136	0.024	0.088	0.184	<0.001
Prob. PSE1	0.222	0.015	0.192	0.251	<0.001
Prob. PSE2	0.333	0.013	0.308	0.358	<0.001
Prob. PSE12	0.309	0.014	0.281	0.337	<0.001

表三 多重中介因子分析之軟體應用與分析結果。其中包含所有的特定路徑效應(PSE、旁邊之數字代表參與該路徑之中介因子代號。)、總效應(total)及路徑效應比例(Prob. PSE)，亦提供了 p-value 以說明其顯著性。此處中介因子個數為二。本軟體可以適用於任意數目之中介因子。

5. 結論

因果中介效應分析為統計分析中的新興領域，其主要目的是探討因果效應透過哪些機制所造成、以及每種機制所解釋的比例為何。傳統的差異法與結構方程式局限於線性模型；因果中介效應分析突破此一限制，得以廣泛應用於公共衛生以及生物醫學的研究領域。針對單一中介因子進行因果中介分析已經有許多方法與軟體。其中 mediation.r 最廣為使用，透過四行指令即可完成套件下載、模型設定、以及資料分析。針對多重中介因子，有三種分析策略：(1) 單一中介效應分析、(2)多重平行中介效應分析、以及(3)有序多重中介效應分析。以上諸多方法能使研究者量化估計有興趣的因果效應以及其機制組成。

参考文献

1. MacKinnon, D., *Introduction to statistical mediation analysis*. 2012: Routledge.
2. Robins, J.M. and S. Greenland, *Identifiability and exchangeability for direct and indirect effects.* Epidemiology, 1992: p. 143-155.
3. Albert, J.M. and S. Nelson, *Generalized causal mediation analysis.* Biometrics, 2011. **67**(3): p. 1028-1038.
4. Avin, C., I. Shpitser, and J. Pearl, *Identifiability of path-specific effects.* 2005.
5. Imai, K., L. Keele, and D. Tingley, *A general approach to causal mediation analysis.* Psychological methods, 2010. **15**(4): p. 309.
6. Pearl, J. *Direct and indirect effects*. in *Proceedings of the seventeenth conference on uncertainty in artificial intelligence*. 2001. Morgan Kaufmann Publishers Inc.
7. van der Laan, M.J. and M.L. Petersen, *Direct effect models.* The international journal of biostatistics, 2008. **4**(1).
8. VanderWeele, T., *Explanation in causal inference: methods for mediation and interaction*. 2015: Oxford University Press.
9. VanderWeele, T.J. and S. Vansteelandt, *Conceptual issues concerning mediation, interventions and composition.* Statistics and its Interface, 2009. **2**(4): p. 457-468.
10. Imai, K., L. Keele, and T. Yamamoto, *Identification, inference and sensitivity analysis for causal mediation effects.* Statistical science, 2010: p. 51-71.
11. Tingley, D., et al., *Mediation: R package for causal mediation analysis.* 2014.
12. Brader, T., N.A. Valentino, and E. Suhay, *What triggers public opposition to immigration? Anxiety, group cues, and immigration threat.* American Journal of Political Science, 2008. **52**(4): p. 959-978.

13. Lin, S.-H. and T. VanderWeele, *Interventional Approach for Path-Specific Effects.* Journal of Causal Inference, 2017. **5**(1).

14. Lin, S.-H., *Generalized interventional approach for causal mediation analysis with causally ordered multiple mediators.* 2019.