

R 을 이용한 Conditional Process Analysis

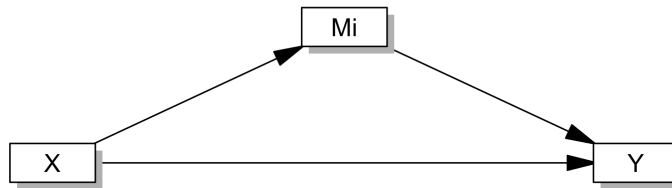
II. 회귀에서 매개모형으로

문건웅

2019/6/1

매개모형

```
library(processR)  
pmacroModel(4)
```



사용할 데이터

```
library(processR)
library(klaR)
data(countries)
head(countries)
```

	Country	Popul	PopDens	GDPpp	LifeEx	InfMor	Illit
AFG	Afghanistan	28717213	44.350908	661.6241	46.97	14.248	64.0
DZ	Algeria	32818500	13.779206	5295.7935	70.54	3.774	30.0
RA	Argentina	38740807	16.265758	10423.1179	75.48	1.616	2.9
BD	Bangladesh	138448210	961.445903	1720.4990	61.33	6.608	56.9
BR	Brazil	182032604	21.385497	7559.0854	71.13	3.174	13.6
CDN	Canada	32207113	3.225656	29002.9100	79.83	0.488	3.0

```
DT::datatable(countries[c(1,2,4:7)])
```

Show entriesSearch:

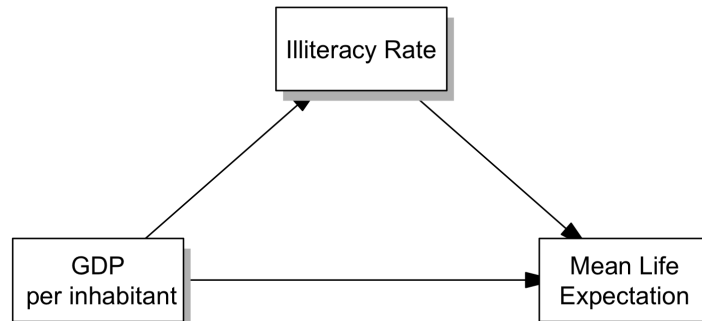
	Country	Popul	GDPpp	LifeEx	InfMor	Illit
AFG	Afghanistan	28717213	661.6241	46.97	14.248	64
DZ	Algeria	32818500	5295.7935	70.54	3.774	30
RA	Argentina	38740807	10423.1179	75.48	1.616	2.9
BD	Bangladesh	138448210	1720.499	61.33	6.608	56.9
BR	Brazil	182032604	7559.0854	71.13	3.174	13.6
CDN	Canada	32207113	29002.91	79.83	0.488	3
VRC	China	1286975468	4653.5464	72.22	2.526	14
CO	Colombia	41662073	6039.0658	71.14	2.247	7.5
RDC	Democratic Republic of the Congo	56625039	600.4411	48.93	9.656	34.5
ET	Egypt	74718797	3878.5421	70.41	3.526	42.3

Showing 1 to 10 of 42 entries

[Previous](#)[1](#)[2](#)[3](#)[4](#)[5](#)[Next](#)

통계적모형

```
labels=list(X="GDPpp",M="Illit",Y="LifeEx")
nodeslabels=list(X="GDP\n per inhabitant",M="Illiteracy Rate",
                 Y="Mean Life\nExpectation")
statisticalDiagram(4,radx=0.12,radY=0.09,labels=nodeslabels,
                  arrowslabels=c("a","b","c"),whatLabel="label")
```



전체효과모형

X 를 독립변수로 Y 를 종속변수로 하는 회귀모형을 만든다.

```
statisticalDiagram(0,radx=0.12,rad=0.09,labels=nodeslabels,  
arrowslabels=c("c"),whatLabel="label",addprime=FALSE)
```



processR을 이용한 Baron과 Kenny의 방법

```
result=mediationBK(labels=labels,data=countries)
result
```

Results of Baron and Kenny Method

Step 1 : Path C (Total Effect) : $c = 0.001$ ($p < 0.001$)

Step 2 : Path A (X on M) : $a = -0.001$ ($p < 0.001$)

Step 3 : Path B (M on Y, controlling for X) : $b = -0.200$ ($p = 0.013$)

Step 4 : Path C' (Direct Effect, X on Y, controlling for M) : $c' = 0.001$ (p

Result : Partial mediation

Results of `bda::mediation.test`

	Sobel	Aroian	Goodman
z.value	2.2080915	2.16446846	2.25446305
p.value	0.0272379	0.03042841	0.02416705

```
for(i in 1:3) {  
  cat(paste0("fit",i),"=", result$equations[[i]],"\n")  
}
```

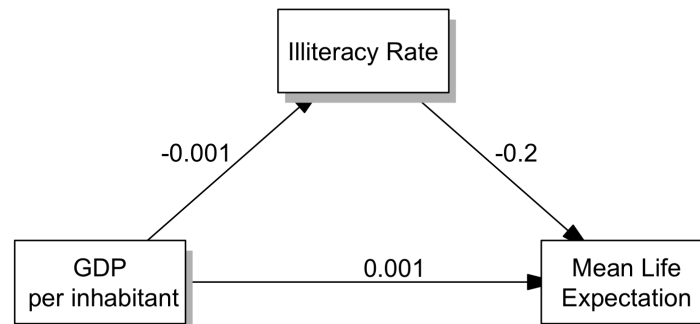
```
fit1 = lm(LifeEx~GDPpp,data=countries)  
fit2 = lm(Illit~GDPpp,data=countries)  
fit3 = lm(LifeEx~Illit+GDPpp,data=countries)
```



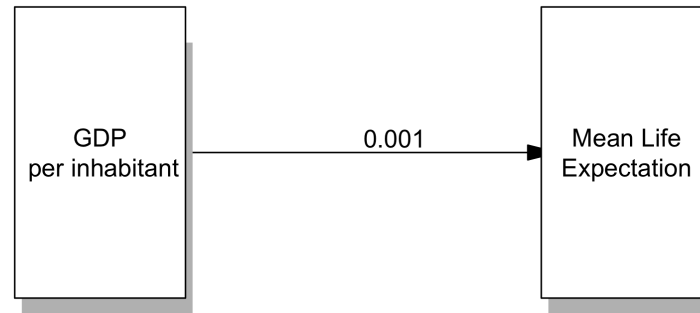
```
modelsSummaryTable(result$fit, labels=labels)
```

Consequent															
LifeEx(Y)						Illit(M)						LifeEx(Y)			
Antecedent	Coef	SE	t	p		Coef	SE	t	p		Coef	SE	t	p	
GDPpp(X)	<i>c</i>	0.001	0.000	5.686	<.001	<i>a</i>	-0.001	0.000	-4.225	<.001	<i>c'</i>	0.001	0.000	3.615	.001
Illit(M)											<i>b</i>	-0.200	0.077	-2.590	.013
Constant	<i>i_Y</i>	60.037	1.774	33.849	<.001	<i>i_M</i>	26.979	3.394	7.950	<.001	<i>i_Y</i>	65.439	2.665	24.554	<.001
Observations	42					42					42				
R2	0.447					0.309					0.528				
Adjusted R2	0.433					0.291					0.504				
Residual SE	8.282 (df = 40)					15.848 (df = 40)					7.748 (df = 39)				
F statistic	F(1,40) = 32.329, p < .001					F(1,40) = 17.854, p < .001					F(2,39) = 21.824, p < .001				

```
plot(result,radx=0.12,rad=0.09,nodeslabels=nodeslabels)
```



```
plot(result,type=1,radx=0.12,rad=0.09,nodeslabels=nodeslabels)
```



Baron과 Kenny의 방법에 대한 비판(1)

- 첫번째 문제는 간접효과 $ab = 0$ 이라는 가설에 대한 검정이 아니라 $a = 0, b = 0$ 이라는 귀무가설에 대한 검정을 따로 실시하여 $ab = 0$ 이라는 가설을 간접적으로 추론한다는 점이다. 드물기는 하지만 a 와 b 가 모두 통계적으로 0과 다르지만 간접효과가 없는 경우도 있으며 더 중요한 것은 a 와 b 가 모두 유의하지 않지만 ab 가 유의한 경우도 있을 수 있다. 간접효과의 추정치는 ab 즉 a 와 b 의 곱이며 a 와 b 각각의 통계적인 유의성은 매개효과의 필요조건이 아니다.
- 둘째, 이 방법으로 매개효과를 증명하기 위해서는 세개의 귀무가설($c = 0, a = 0, b = 0$)을 연속하여 성공적으로 기각하여야 한다. 하지만 가설에 대한 검정은 가정들이 만족하여야 하며 또한 항상 1종 또는 2종의 오류가 일어날 수 있다. 가설에 대한 가정들이 만족된다고 하더라도 하나의 주장을 하기 위해 여러 개의 가설을 검정한다면 그만큼 오류를 일으킬 가능성이 많아지므로 간접효과에 대한 하나의 총괄적 추론만으로 검정하는 것이 보다 바람직하다.

Baron과 Kenny의 방법에 대한 비판(2)

- 세번째로 단계적 인과관계 접근을 사용하는 많은 사람들이 매개변수로 이분형변수를 생각한다. 즉 X 의 Y 에 대한 효과가 M 에 의해 매개되는가 아닌가 하는 정성적인 방법을 생각한다. 간접효과는 정량화되어야 하며 간접효과가 있다 또는 없다는 흑백논리가 아닌 간접효과의 신뢰구간을 제시하는 것이 보다 바람직하다.
- 네째, 이 방법은 첫 단계로 X 가 Y 에 영향을 미치는가 하는 검정에서 시작한다. 전체효과 $c = 0$ 이라는 귀무가설을 기각하지 못하면 M 의 매개효과에 대한 다른 검정들은 부적절한 것으로 취급하여 중단된다. 이 논리는 효과가 존재하지 않으면 매개될 것도 없으며 효과가 존재하지 않는 기전을 설명할 필요도 없다는 주장이다. 이러한 주장은 잘못된 것으로 전체효과가 0과 다르지 않더라도 X 가 M 을 통하여 Y 에 영향을 미칠 수 있다. 전체효과의 크기가 간접효과의 크기를 결정하거나 제한하지 않는다. 전체 효과가 0과 다르지 않더라도 간접효과는 얼마든지 0과 다를 수 있다.

간접효과에 대한 검정

- 간접효과의 유의성에 대한 검정은 크게 두 가지가 있는데 하나는 소벨(Sobel)검정이고 또 하나는 부트스트랩(bootstrapping)을 이용한 신뢰구간이다.
- 소벨검정은 표본의 분포가 정규분포 가정을 만족할때 사용할 수 있다.
- 또한 소벨 검정은 표본수가 클 때에만 적절한 검정력을 갖기 때문에 부트스트랩 방법이 보다 추천된다.

lavaan을 사용한 매개효과 분석

```
library(lavaan)
model <- '
Illit ~ a*GDPpp
LifeEx ~ b*Illit + c*GDPpp
indirect := a*b
direct := c
total := direct + indirect
prop.mediates := indirect / total
'

set.seed(123)
semfit=sem(model=model,data=countries, se="boot", bootstrap=999)
```

lavaan을 이용한 결과 출력

```
parameterEstimates(semfit, boot.ci.type = "bca.simple",
  ci=TRUE)[c(1:5,9,10,8)]
```

	lhs	op	rhs	label	est
1	Illit	~	GDPpp	a	-0.001
2	LifeEx	~	Illit	b	-0.200
3	LifeEx	~	GDPpp	c	0.001
4	Illit	~~	Illit		239.199
5	LifeEx	~~	LifeEx		55.745
6	GDPpp	~~	GDPpp		91762387.208
7	indirect	:=	a*b	indirect	0.000
8	direct	:=	c	direct	0.001
9	total	:=	direct+indirect	total	0.001
10	prop.mediated	:=	indirect/total	prop.mediated	0.285
	ci.lower		ci.upper	pvalue	
1	-0.002		-0.001	0.000	
2	-0.364		-0.039	0.013	
3	0.000		0.001	0.000	
4	157.642		329.355	0.000	
5	31.148		91.495	0.000	
6	91762387.208		91762387.208	NA	
7	0.000		0.000	0.030	

processR 패키지의 역할

```
library(processR)
labels=list(X="GDPpp",M="Illit",Y="LifeEx")
model=tripleEquation(labels=labels)
cat(model)
```

```
Illit~a*GDPpp
LifeEx~c*GDPpp+b*Illit
indirect :=(a)*(b)
direct :=c
total := direct + indirect
prop.mediates := indirect / total
```

```
set.seed(123)
semfit=sem(model=model,data=countries, se="boot", bootstrap=999)
medSummaryTable(semfit)
```

Effect	Equation	estimate	95% Bootstrap CI	p
indirect	(a)*(b)	0.000	(0.000 to 0.000)	.030
direct	c	0.001	(0.000 to 0.001)	<.001
total	direct+indirect	0.001	(0.001 to 0.001)	<.001
prop.mediated	indirect/total	0.285	(0.065 to 0.514)	.014

boot.ci.type = bca.simple