

# 매개모형분석

문건웅

2019/12/6

# 내용

- 독립변수  $X$ 가 이분형 변수인 매개모형(예제 1)
- 독립변수  $X$ 가 연속형 변수인 매개모형(예제 2)
- 공변량을 포함하는 매개모형(예제 3)

# 사용하는 패키지

이번 장에서는 다음과 같은 패키지를 사용한다.

```
require(processR)
require(lavaan)
require(nortest)    # ad.test()의 사용을 위해
```

# 예제 1: $X$ 가 이분형 변수인 매개모형

- 사용하는 데이터 :미디어 영향력 연구(Presumed Media Influence, pmi)
- 데이터 설명
  - Tal-Or 등(2010)에 의해 이스라엘에서 수행된 연구
  - 총 123명의 참가자에게 설탕부족사태가 올 것이라는 신문 기사를 읽게 하였다.
  - 참가자들 중 반에게는 신문기사가 1면에 실릴 것이라고 얘기하였으며 (cond=1) 다른 참가자들에게는 경제면 부록에 실릴 것이라고 얘기하였다 (cond=0).
  - 기사를 읽은 후 다른 사람들이 얼마나 이 기사에 영향을 받을 것인지에 대한 믿음(즉, 일반 대중이 이 기사의 영향으로 설탕을 사러 가는 정도)을 측정하여 추정 미디어 영향력(pmi)변수에 기록하였다.
  - 또한 설탕구매의도를 측정하여 reaction변수에 기록하였다.
  - pmi 와 reaction은 값이 클수록 해당변수의 강도가 강한 것이다.

# 통계적 모형

- 독립변수(  $X$  )는 신문기사의 위치(cond)이며 이분형 변수이다.
- 매개변수(  $M$  )는 추정 미디어 영향력 변수(pmi)
- 종속변수(  $Y$  )는 설탕구매의도(reaction)
- 매개변수와 종속변수는 연속형 변수이다.

```
labels=list(X="cond",M="pmi",Y="reaction")  
drawModel(labels=labels,label.pos=2)
```

## 연구 가설

Tal-Or 등(2010)은 신문기사가 중간에 실렸다고 들은 경우보다 1면에 실렸다고 들은 경우가 다른 사람들이 설탕재고부족 가능성에 보다 영향을 받을 것이라는 믿음을 유발시키고, 그 결과 설탕을 보다 많이 구매할 것이라고 추론하였다.

# 기술통계량 요약

이 연구에서 독립변수인 신문기사의 위치(**cond**)에 따른 매개변수 **pmi**와 종속변수 **reaction**의 기술통계량은 다음과 같이 요약할 수 있다.

```
labels=list(X="cond",Y="reaction",M="pmi")
xlabels=c("Interior Page", "Front Page")
meanSummaryTable(data=pmi,labels=labels,xlabels=xlabels)
```

		Y	M	Y
		<i>reaction</i>	<i>pmi</i>	<i>adjusted</i>
Interior Page	Mean	3.250	5.377	3.364
cond(X) = 0	SD	1.608	1.338	
Front Page	Mean	3.746	5.853	3.618
cond(X) = 1	SD	1.452	1.267	
		Mean	3.484	5.602
		SD	1.550	1.321

## 회귀모형 요약

단순매개모형은 다음과 같은 두 개의 회귀식으로 나타낼 수 있다.

$$M = i_M + aX + \varepsilon_M$$

$$Y = i_Y + c'X + bM + \varepsilon_Y$$

## R을 이용한 회귀모형 분석

```
fit1=lm(pmi~cond,data=pmi)  
fit2=lm(reaction~ cond+pmi,data=pmi)
```



# 회귀분석의 결과 요약(1)

```
summary(fit1)
```

Call:

```
lm(formula = pmi ~ cond, data = pmi)
```

Residuals:

Min	1Q	Median	3Q	Max
-4.8534	-0.8534	0.1466	1.1231	1.6231

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	5.3769	0.1618	33.222	<2e-16 ***
cond	0.4765	0.2357	2.022	0.0454 *

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.305 on 121 degrees of freedom

Multiple R-squared: 0.03268, Adjusted R-squared: 0.02468

F-statistic: 4.088 on 1 and 121 DF, p-value: 0.0454

```
summary(fit2)
```

Call:

```
lm(formula = reaction ~ cond + pmi, data = pmi)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-3.07636	-1.06128	-0.06346	0.94573	2.94299

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	0.52687	0.54968	0.958	0.340
cond	0.25435	0.25582	0.994	0.322
pmi	0.50645	0.09705	5.219	7.66e-07 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.393 on 120 degrees of freedom

Multiple R-squared: 0.2059, Adjusted R-squared: 0.1927

F-statistic: 15.56 on 2 and 120 DF, p-value: 9.83e-07

# 모형의 회귀계수 요약

```
modelsSummaryTable(labels=labels,data=pmi)
```

Antecedent	Consequent									
	pmi(M)					reaction(Y)				
		Coef	SE	t	p		Coef	SE	t	p
cond(X)	<i>a</i>	0.477	0.236	2.022	.045	<i>c'</i>	0.254	0.256	0.994	.322
pmi(M)						<i>b</i>	0.506	0.097	5.219	<.001
Constant	<i>i<sub>M</sub></i>	5.377	0.162	33.222	<.001	<i>i<sub>Y</sub></i>	0.527	0.550	0.958	.340
Observations	123					123				
R2	0.033					0.206				
Adjusted R2	0.025					0.193				
Residual SE	1.305 ( df = 121)					1.393 ( df = 120)				
F statistic	F(1,121) = 4.088, p = .045					F(2,120) = 15.557, p < .001				

## 회귀모형 요약

단순매개모형은 다음과 같은 두 개의 회귀식으로 나타낼 수 있다.

$$M = i_M + aX + \varepsilon_M$$

$$Y = i_Y + c'X + bM + \varepsilon_Y$$

회귀식을 회귀분석의 결과를 이용하여 다시 써 보면 다음과 같다.

$$\hat{M} = 5.377 + 0.477X$$

$$\hat{Y} = 0.527 + 0.254X + 0.506M$$

위 식에서  $\hat{M}$ 은  $M$ -"**hat**"으로 읽으며  $M$ 의 추정치 또는 기대치를 뜻한다.

## 회귀계수 $a$

회귀계수  $a$ 는  $X$ 가 1단위만큼 차이 나는 두 예에서  $M$ 의 변화는 0.477단위로 추정된다는 것을 말해준다. 따라서 신문기사가 안쪽에 실린 경우에 비해 신문기사가 1면에 실린 경우 추정 미디어 영향력은 평균 0.477단위만큼 높다.

이 데이터에서  $X$ 는 이분형 변수로 0과 1로 입력되어 있으므로  $a$ 는  $X = 0$ 일 때의  $M$ 의 평균과  $X = 1$ 일 때의  $M$ 의 평균의 차이와 같다.

$$a = [\bar{M}|(X = 1)] - [\bar{M}|(X = 0)]$$

위의 식에서  $\bar{M}$ 은  $M$ 의 평균이며 "|"는 주어진 경우에(given)로 해석한다.

즉,  $a$ 는  $X = 1$ 인 경우의  $M$ 의 평균  $\bar{M}|(X = 1)$ 과  $X = 0$ 인 경우의  $M$ 의 평균  $\bar{M}|(X = 0)$ 의 차이로  $5.853 - 5.377 \approx 0.477$ 이다.

## 회귀계수 $b$

회귀계수  $b = 0.506$ 는 같은 실험군에 속한 두 사람에서 추정 미디어 영향력이 1단위만큼 차이 나는 경우 보이는 설탕구매의도의 차이의 추정치이다. 즉,

$$b = [\hat{Y}|(M = m, X = x)] - [\hat{Y}|(M = m - 1, X = x)]$$

이다. 회귀계수  $b$ 의 부호가 양성이므로 추정 미디어 영향력이 높은 경우 설탕구매의도도 높다고 할 수 있다.

## 간접효과 $ab$

간접효과  $ab$ 는 신문기사의 위치 조작이 추정 미디어 영향력에 미치는 효과( $a$ )와 신문기사의 위치가 일정할 때 추정 미디어 영향력이 설탕구매의도에 미치는 효과( $b$ )의 곱으로 계산된다.

즉, 신문기사의 위치가 추정 미디어 영향력을 매개로 설탕구매의도에 미치는 효과로  $ab = 0.477(0.506) = 0.241$ 로 계산된다.

즉, 신문기사가 중간에 실린 경우에 비해 1면에 실렸다고 들은 경우 추정 미디어 영향력을 매개로 한 간접효과로 설탕구매의도가 0.241단위만큼 높게 나타났다.

## 직접효과 $c'$

신문기사의 위치가 설탕구매의도에 미치는 직접효과  $c' = 0.254$ 이다. 즉,  $M$ 이 같고  $X$ 가 1단위만큼 차이 나는 두 예에서 보이는  $Y$ 의 값의 차이는 0.254단위로 추정된다.

이 연구에서는 신문기사가 1면에 실린 경우 추정 미디어 영향력과는 독립적으로 신문기사가 중간에 실렸다고 들은 경우에 비해 설탕구매의도가 0.254단위만큼 높다고 할 수 있다. 즉,

$$[\bar{Y}|(X = 1, M = m)] - [\bar{Y}|(X = 0, M = m)] = 0.254$$

이다.



## $Y$ 의 수정평균 $\bar{Y}_*$

회귀식에  $M$ 의 값으로  $M$ 의 평균( $\bar{M}$ )을 넣으면  $Y$ 의 수정평균  $\bar{Y}_*$ 를 얻을 수 있다

$$\begin{aligned} Y &= i_Y + bM + c'X + \varepsilon_Y \\ \bar{Y}_* &= i_Y + b\bar{M} + c'X \end{aligned}$$

추정 미디어 영향력이 평균( $\bar{M} = 5.602$ )인 사람이 신문기사가 1면에 실릴 예정이라고 들을 경우( $X = 1$ )  $Y$ 의 수정평균  $\bar{Y}_*$ 는 다음과 같다.

$$\bar{Y}_* = 0.527 + 0.506(5.602) + 0.254(1) \approx 3.618$$

위의 식에서 실제 값 3.618과 소수점 아래 셋째 자리에서 반올림한 값들을 계산한 3.616의 차이는 반올림에러에 의한 것이다. 추정 미디어 영향력이 평균( $\bar{M} = 5.602$ )인 사람이 신문기사가 중간에 실릴 예정이라고 들을 경우( $X = 0$ )  $Y$ 의 수정평균  $\bar{Y}_*$ 는 다음과 같다.

$$\bar{Y}_* = 0.527 + 0.506(5.602) + 0.254(0) \approx 3.364$$

위의 식에서도 실제 값 3.364과 소수점아래 세째 자리에서 반올림한 값들을 계산한 3.362의 차이는 반올림에러에 의한 것이다. 두 수정평균의 차이는 직접효과인 0.254이며 이 값은  $M$ 의 값이 달라져도 변하지 않는다.

# 총효과

총효과는 직접효과와 간접효과의 합으로 계산된다. 이 연구에서 총효과는  $c' + ab = 0.254 + 0.241 = 0.495$ 이다. 이는 신문기사가 1면에 실릴 것이라고 들은 사람들은 중간면에 실릴 것이라고 들은 사람들보다 설탕구매의도가 0.495단위만큼 높다고 해석할 수 있다. 이 연구와 같은 단순매개모형에서 총효과는 반응변수를  $Y$ , 설명변수를  $X$ 로 하는 단순회귀분석으로 추정할 수 있다.

```
fit3=lm(reaction~cond,data=pmi)
fit3
```

Call:

```
lm(formula = reaction ~ cond, data = pmi)
```

Coefficients:

(Intercept)	cond
3.2500	0.4957

위 회귀분석의 결과 회귀계수  $c$ 는 총효과이며 이 값은 두 군 간의  $Y$ 값의 평균차이와 같다(즉,  $\bar{Y}_{front} - \bar{Y}_{interior} = 0.496$ ).

## 총효과, 간접효과, 직접효과

$X$ 가 이분형변수로 두 군을 0과 1로 코딩한 경우 총효과, 간접효과, 직접효과는 다음과 같다.

$$\begin{aligned}(\bar{Y}_{X=1} - \bar{Y}_{X=0}) &= (\bar{Y}^*_{X=1} - \bar{Y}^*_{X=0}) + (\bar{M}_{X=1} - \bar{M}_{X=0})b \\ Total\ effect(c) &= Direct\ effect(c') + Indirect\ effect(ab)\end{aligned}$$

위의 식을 위의 분석 결과로 대치하면 다음과 같다.

$$\begin{aligned}(3.746 - 3.250) &= (3.618 - 3.364) + (5.853 - 5.377)0.506 \\ Total\ effect(c) &= Direct\ effect(c') + Indirect\ effect(ab)\end{aligned}$$

# processR을 이용한 간접효과 분석

이 모형을 분석하기 위한 lavaan syntax를 만든다.

```
labels=list(X="cond",Y="reaction",M="pmi")  
model=tripleEquation(labels=labels)  
cat(model)
```

```
pmi~a*cond  
reaction~c*cond+b*pmi  
indirect :=(a)*(b)  
direct :=c  
total := direct + indirect  
prop.mediates := indirect / total
```

위에서 만든 lavaan syntax를 이용하여 분석한다. 부트스트래핑을 통해 재현가능한 결과를 얻기 위해 set.seed()함수에 시드값을 준다. 표준오차는 부트스트랩을 이용하여 구하고(se="boot") 부트스트랩 횟수는 1,000회 시행한다(bootstrap=1000, 생략 가능).

```
set.seed(123)
semfit=sem(model=model,data=pmi,se="boot",bootstrap=1000)
```

```
summary(semfit,ci=TRUE)
```

lavaan 0.6-5.1453 ended normally after 21 iterations

Optimization method	NLMINB
Number of free parameters	5
Number of observations	123
Estimator	ML
Model Fit Test Statistic	0.000
Degrees of freedom	0

Parameter Estimates:

Standard Errors	Bootstrap
Number of requested bootstrap draws	1000
Number of successful bootstrap draws	985

# 매개효과요약

위의 분석 결과 중 매개효과 분석을 표로 정리해 보면 다음 표와 같다.

```
medSummaryTable(semfit)
```

Effect	Equation	estimate	95% Bootstrap CI
indirect	(a)*(b)	0.241	(0.003 to 0.526)
direct	c	0.254	(-0.251 to 0.766)
total	direct+indirect	0.496	(-0.060 to 1.041)
prop.mediated	indirect/total	0.487	(-0.993 to 2.906)

boot.ci.type = perc

boot.ci.type인수에는 "norm","basic","perc","bca.simple" 등을 지정할 수 있다. 이 중 summary() 함수로 출력되는 것은 "perc"이다. 이들 값에 대한 자세한 설명은 boot 패키지의 boot.ci함수의 도움말을 참조한다.

# 통계적인 추론

- 단순매개모형에서 총효과는 직접효과와 간접효과로 나뉘어지며 회귀분석(OLS)을 이용하여 추정할 경우 어떤 데이터에서도  $c = c' + ab$ 는 항상 참이다.
- 하지만  $c, c', ab$ 로 표현되는 효과들은 참값인  $Tc, Tc', Ta_Tb$ 에 대하여 표본에서 얻어진 추정치이다. 이 값들은 주어진 표본에서의 관계를 의미할 뿐 일반화할 수는 없다.
- 연구자들은 이들 분석 결과들을 일반화하기 위해 가설 검정을 통해 "우연"성을 배제하거나 효과들에 대한 구간추정치의 설정을 통해 표집분산을 고려하여야 한다.

# 총효과에 대한 추론

총효과에 대한 추론은  $M$ 을 배제하고  $Y$ 를  $X$ 로 회귀분석함으로써 추정할 수 있다.  
즉, 이 모형에서는 다음과 같이 총효과를 추정할 수 있다.

```
fit3=lm(reaction~cond,data=pmi)
fit3
```

Call:

```
lm(formula = reaction ~ cond, data = pmi)
```

Coefficients:

(Intercept)	cond
3.2500	0.4957



위의 분석 결과에서 총효과  $c$ 의 추정치는  $0.496(p = .077)$ 임을 알 수 있다. 총효과  
의 95% 신뢰구간은 다음과 같이 구할 수 있다.

```
confint(fit3)
```

	2.5 %	97.5 %
(Intercept)	2.87268088	3.627319
cond	-0.05378462	1.045164

즉,  $Tc$ 는 -0.054에서 1.045 사이에 존재할 것이라고 95%의 확신을 갖고 이야기할 수  
있다.

# 직접효과에 대한 추론

직접효과는  $M$ 이 일정한 두 예에서  $X$ 의 값이 1단위만큼 차이 나는 경우 보이는  $Y$ 값의 차이의 추정치이다. 직접효과에 대한 추론은 회귀분석에서 회귀계수에 대한 표준적인 추론과 같다. 즉,  $TC'$ 이 0과 같다는 귀무가설을 검정하거나  $TC'$ 에 대한 신뢰구간의 설정을 통해 이루어 진다.

```
fit2=lm(reaction~pmi+cond,data=pmi)
summary(fit2)
```

Call:

```
lm(formula = reaction ~ pmi + cond, data = pmi)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-3.07636	-1.06128	-0.06346	0.94573	2.94299

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	0.52687	0.54968	0.958	0.340
pmi	0.50645	0.09705	5.219	7.66e-07 ***
cond	0.25435	0.25582	0.994	0.322

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

또한  $TC'$ 에 대한 신뢰구간은 다음과 같이 구할 수 있다.

```
confint(fit2)
```

	2.5 %	97.5 %
(Intercept)	-0.5614647	1.6151957
pmi	0.3142996	0.6985974
cond	-0.2521567	0.7608651

위의 분석 결과  $TC'$ 에 대한 95% 신뢰구간은  $-0.252 - 0.761$ 로 0을 포함하고 있으므로 직접효과가 0이라는 주장을 기각할 수 없다.

물론 귀무가설에 대한 검정 결과와 95% 신뢰구간을 이용한 추론은 같은 정보를 다르게 포장한 것에 불과하므로 항상 같은 결과를 보인다.

# 간접효과에 대한 추론

간접효과는  $X$ 의 값이 1단위 차이 나는 두 예에서  $X$ 가  $M$ 에 영향을 미치고 다시  $M$ 이  $Y$ 에 영향을 미침으로써 보이는  $Y$ 값의 차이의 추정치이다. 연구자들은 주어진 자료를 이용하여 이러한  $Y$ 값의 차이의 추정치가 0과 다르다는 주장을 할 수 있는지 알고 싶어 한다. 만일 이 차이가 0과 다르다면  $M$ 은  $Y$ 에 대한  $M$ 의 효과를 매개한다고 할 수 있다.

직접효과에 대한 추론과 마찬가지로 간접효과에 대한 추론도  $ta_Tb = 0$ 라는 귀무가설에 대한 검정 또는 신뢰구간에 근거한 추론을 할 수 있다.

- 정규이론방법
- 부트스트랩에 의한 신뢰구간

# 정규이론방법

정규이론방법은 "계수들의 곱" 방법, 소벨검정, 델타방법 등으로도 불린다. 이 방법은 직접효과 또는 총효과의 추론과 같은 방법을 사용하여 간접효과를 추론하는 방법이다.

먼저 회귀계수의 곱인  $ab$ 의 표준오차를 추정하고  $ab$ 의 표본분포를 정규분포로 가정한 후 특정한  $Ta_Tb$ 에 대한 귀무가설에 대하여  $ab$ 의  $p$ 값을 계산하거나 신뢰구간을 도출하여 검정한다.

정규이론방법을 사용하기 위하여는  $ab$ 의 표준오차에 대한 추정치가 필요하다. 이 추정치를 구하는 공식은 몇 개가 알려져 있으나 그중 표준오차의 이차추정량을 구하는 공식은 다음과 같다.

$$se_{ab} = \sqrt{a^2 se_b^2 + b^2 se_a^2 + se_a^2 se_b^2}$$

위 식에서  $se_a^2$ 와  $se_b^2$ 는 각각  $a$ 와  $b$ 의 제곱표준오차이다. 미디어 영향력 연구에서 간접효과의 표준오차의 이차추정량을 계산하면 다음과 같다.

$$se_{ab} = \sqrt{0.477^2 0.097^2 + 0.506^2 0.236^2 + 0.236^2 0.097^2} = 0.130$$

$\tau a_T b = 0$ 이라는 귀무가설에 대한 검정은  $ab$ 와 그 표준오차의 비인  $Z$ 통계량을 계산하여 표준정규분포로부터  $\pm Z$ 보다 더 극단적인 값을 갖는 비율을 계산한다.

$$Z = \frac{ab}{se_{ab}} = \frac{0.241}{0.130} = 1.854$$

양측검정을 이용하여  $Z = 1.854$ 보다 극단적인 값을 갖는  $p$ 값은 .064가 된다. 따라서  $\alpha = 0.05$ 의 의사결정수준에서 귀무가설  $\tau a_T b = 0$ 은 기각되지 않는다.

귀무가설 검정보다 신뢰구간을 이용한 방법을 선호하는 경우  $ab$ 의 표본분포의 정규성을 가정하여 다음과 같이 구간추정치를 계산한다.

$$ab - Z_{ci\%}se_{ab} \leq_T a_Tb \leq ab + Z_{ci\%}se_{ab}$$

위 식에서  $ci$ 는 신뢰수준(예를 들어, 95)이고  $Z_{ci\%}$ 는 표준정규분포에서 이 값 이상의 확률이  $(100 - ci)/2\%$ 의 값을 가지는  $Z$ 값을 의미한다. 신뢰구간이 95%이면  $Z = 1.96$ 이 된다. 이 연구에서  $ab = 0.241$ 이므로 95% 구간추정치를 계산하면 다음과 같다.

$$\begin{aligned} 0.241 - 1.96(0.130) &\leq_T a_Tb \leq 0.241 + 1.96(0.130) \\ -0.014 &\leq_T a_Tb \leq 0.496 \end{aligned}$$

따라서 연구자는  $a_Tb$ 가 -0.014 부터 0.496 사이의 구간 어딘가에 속한다고 95% 확신할 수 있다. 귀무가설 검정과 마찬가지로  $a_Tb$ 의 가능한 값으로 0을 배제할 수 없기 때문에 신문기사의 위치가 추정 미디어 영향력을 매개로 설탕구매의도에 미치는 영향이 0과 다르다고 주장할 수 없다. 즉, 추정 미디어 영향력을 매개로 한 간접효과가 나타났다고 주장할 수 없다.

## 정규이론방법의 단점

하지만 소벨검정은 두 가지 단점 때문에 권장되지 않는다. 첫째, 이 방법은  $ab$ 의 표본 분포의 정규성을 가정하고 있다. 여러 연구에서  $ab$ 의 표본분포는 정규분포를 따르지 않는 것으로 나타났다.  $ab$ 의 표본분포가 정규성을 가정하는 소벨검정을 사용할 수 있을 만큼 정규분포에 가까운지 알 수 있는 방법이 없기 때문에 정규성을 가정하지 않는 방법을 사용하는 것이 바람직하다. 둘째, 시뮬레이션을 사용하여 다른 추론 방법들과 비교해 본 결과 소벨검정의 검정력이 가장 낮게 나타났고 신뢰구간도 부정확하게 나타났다. 이러한 두 가지 이유로 정규이론방법은 권장되지 않는다.



**참고.** 위에서 계산한  $Z$ 통계량은 회귀계수들을 소수점 셋째자리에서 반올림한 후 제곱을 하고 곱하는 등의 계산을 하였으므로 반올림에 따른 오차가 많이 발생한다. `processR` 패키지의 `mediationBK()`함수를 이용하면 반올림하지 않은 회귀계수들을 이용한  $ab$ ,  $se_{ab}$  및  $Z$ 통계량 및  $p$ 값 및 95% 구간추정치를 구할 수 있다.

```
labels=list(X="cond",M="pmi",Y="reaction")
result=mediationBK(labels=labels,data=pmi)
result$normalTheory
```

ab	seab	z	p	ci.lower	ci.upper
0.24133546	0.13003848	1.85587728	0.06347105	-0.01353528	0.49620621

$Z$ 통계량을 이용하여  $p$ 값을 구하려면 다음과 같이 할 수 있다.

```
Z=result$normalTheory[3]
pnorm(Z,lower.tail=FALSE)*2
```

```
z
0.06347105
```

# 부트스트랩에 의한 신뢰구간

위에서 언급한 정규이론방법에 의한 검정은 간접효과의 표본분포의 정규성을 가정한 방법이다. 과거 컴퓨터가 발전되지 않고 손으로 계산을 하던 시절에는 계산을 가능하게 하기 위해 통계학에서 많은 가정들을 사용해 왔다. 하지만 문제가 되는 가정이 필요 없는 다른 대체 방법을 통해 추론 검정이 가능하다면 문제가 되는 방법을 사용할 이유가 없다.

부트스트랩 방법은 재표본추출(resampling) 방법으로 알려져 있으며 연구자들이 직면하게 되는 많은 통계적인 추론 문제에 광범위하게 적용할 수 있는 방법으로 반복 추출 시 통계량의 행태를 알지 못하거나 계산이 너무 복잡하거나 상황에 따라 달라지는 경우 특히 유용하다.

부트스트랩을 이용하는 경우  $n$ 개의 원표본은 모집단의 축소된 대표표본으로 간주된다. 이 원표본 내의 관찰치들을 반복적으로 복원재추출하여  $n$ 개의 새로운 표본에서 관심 있는 통계량을 다시 계산한다. 이러한 과정을 계속 반복 수행하여(이상적으로는 수천번 이상) 통계량의 표본분포를 얻고, 이러한 경험적 표본분포에서 설정한 대표치를 추론작업에 사용한다.

# 미디어 영향 연구에서 매개효과

```
medSummaryTable(semfit)
```

Effect	Equation	estimate	95% Bootstrap CI
indirect	(a)*(b)	0.241	(0.003 to 0.526)
direct	c	0.254	(-0.251 to 0.766)
total	direct+indirect	0.496	(-0.060 to 1.041)
prop.mediates	indirect/total	0.487	(-0.993 to 2.906)

boot.ci.type = perc

processR을 이용한 매개효과의 검정에서는 먼저 분석에 사용할 수 있는 lavaan syntax를 만든 후 이를 이용하여 모형을 만들어 검정을 하였고, 이때 신뢰구간을 부트스트랩 방법으로 구하였으며 부트스트랩은 1,000번 반복하였다. 이때 만들어진 모형에서 부트스트랩에 의한 통계량을 추출하여 bootData에 저장한다. processR 패키지의 getBootData() 함수는 모형에서 부트스트랩 데이터를 추출해 준다.

```
bootData=getBootData(semfit)
```

우리가 보고자 하는 간접효과는 bootData의 a(bootData\$a)와 b(bootData\$b)를 곱한 것이다.

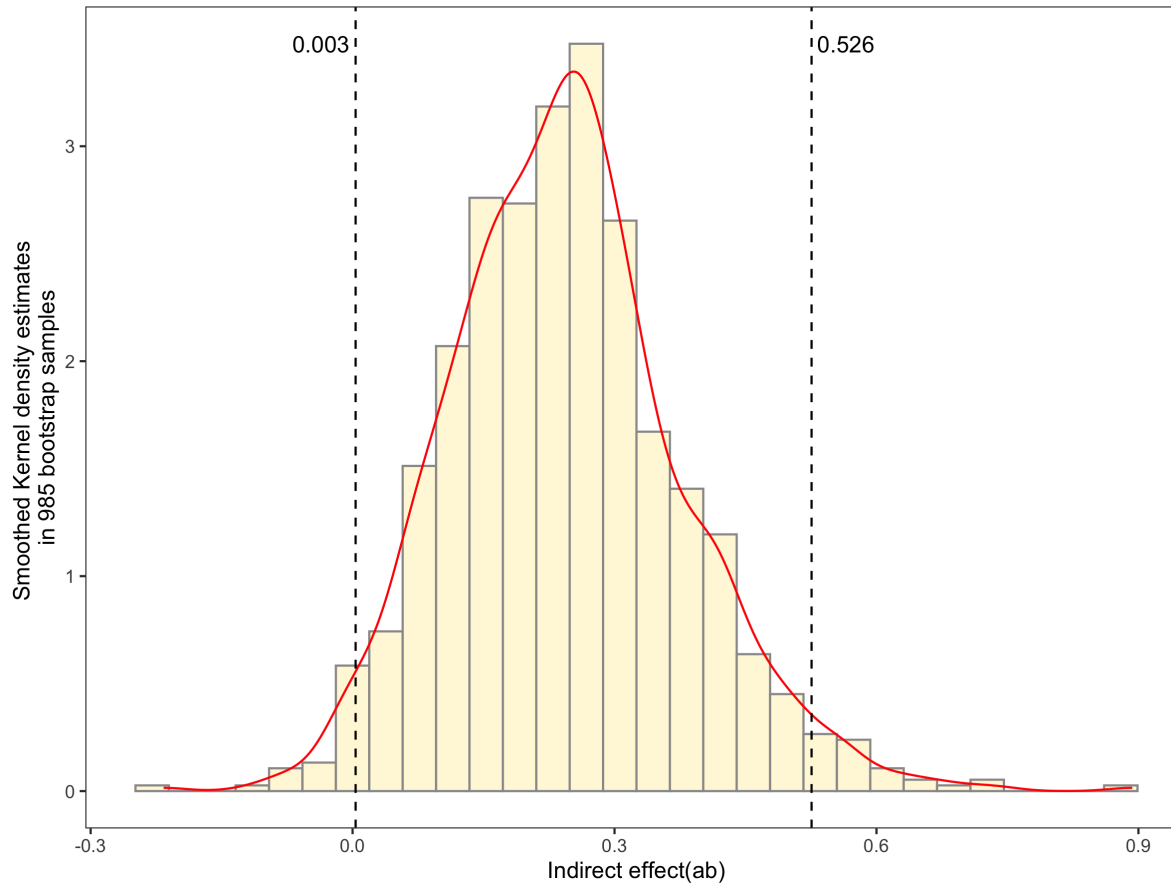
```
ab=bootData$a*bootData$b
```

*ab*의 부트스트랩에 의한 95% 신뢰구간의 추정치는 다음과 같이 구할 수 있다.

```
quantile(ab,probs=c(0.025,1-0.025),type=6)
```

2.5%	97.5%
0.003471443	0.525655909

```
densityPlot(ab)
```

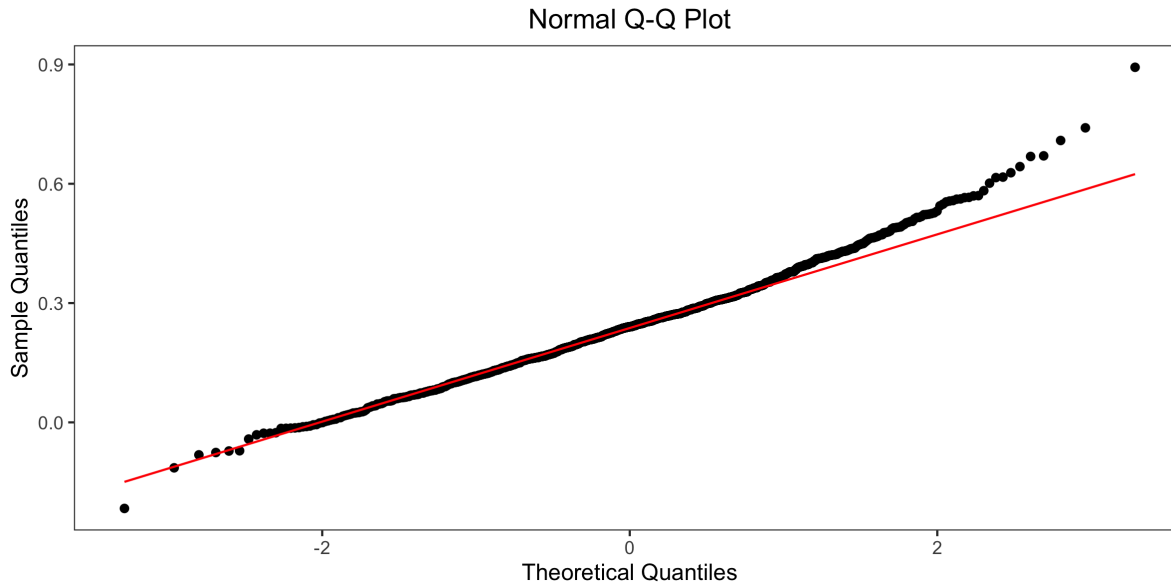


위의 `densityPlot`에서 살펴보면 *ab*의 부트스트랩에 의한 95% 신뢰구간의 추정치는 0.003 – 0.526 임을 알 수 있다. 신뢰구간이 0을 포함하지 않으므로 이 결과는 간접효과가 통계적으로 유의한 양의 값을 갖는다고 결론 내릴 수 있다.

# 간접효과는 정규분포하는가?

위의 `densityPlot`에서 *ab*의 분포는 좌우가 비대칭이며 정규분포를 이루지 않는다. 이를 시각적으로 확인하기 위해 Q-Q plot을 그려 볼 수 있다.

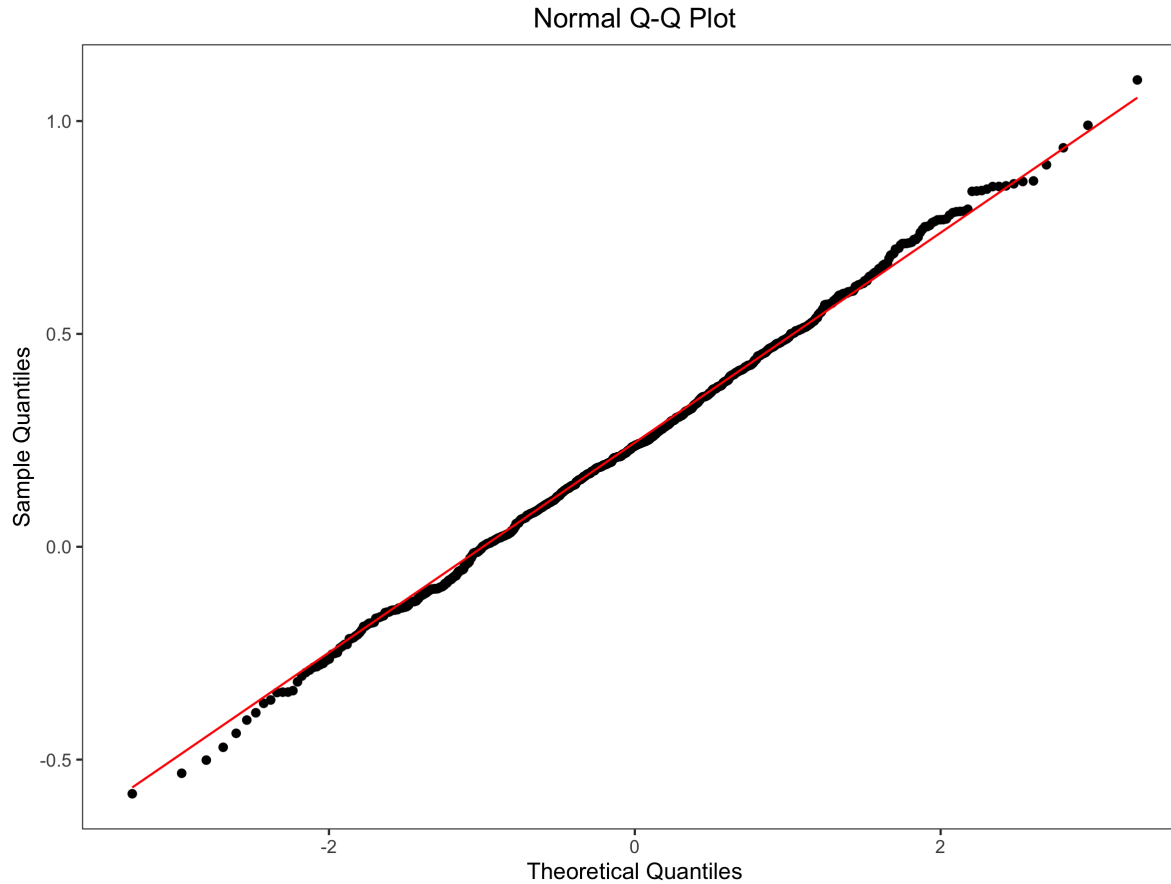
```
qqPlot(ab)
```



위의 Q-Q plot에서 참조선에 비해 값들이 한쪽으로 치우치므로 정규성이 위배된다고 할 수 있다.

참고로 직접효과인  $c'$ 의 Q-Q plot과 비교해 보면 차이를 뚜렷이 알 수 있다.

```
qqPlot(bootData$c)
```



하지만 Q-Q plot에서 한쪽으로 치우쳤다는 것에 대한 판단은 주관적이라는 단점이 있다. 보다 객관적인 방법으로 Shapiro-Wilk 정규성 검정과 Anderson-Darling 정규성 검정이 있다. Shapiro-Wilk 정규성 검정은 표본수가 5,000개 이하인 경우 적합한 정규성 검정 방법이다. Anderson-Darling 정규성 검정은 표본수에 제한을 받지 않는다. Shapiro-Wilk 정규성 검정과 Anderson-Darling 검정은 둘 다 데이터가 정규분포를 따른다는 가설을 귀무가설로 하는 검정 방법으로,  $p$ 값이 0.05 이하인 경우 데이터가 정규분포하지 않는다는 대립가설을 채택하게 된다.

```
shapiro.test(ab)
```

Shapiro-Wilk normality test

```
data:  ab
W = 0.98849, p-value = 5.553e-07
```

```
nortest::ad.test(ab)
```

Anderson-Darling normality test

```
data:  ab
A = 2.1272, p-value = 2.09e-05
```

두 검정 결과 모두  $p < .05$ 로 부트스트랩에 따른 간접효과  $ab$ 의 분포는 정규분포를 따르지 않는다고 결론 내릴 수 있다.



참고로 직접효과의 경우를 검정해 보면 다음과 같다.

```
x=bootData$c  
shapiro.test(x)
```

Shapiro-Wilk normality test

```
data:  x  
W = 0.99886, p-value = 0.8011
```

```
nortest::ad.test(x)
```

Anderson-Darling normality test

```
data:  x  
A = 0.37848, p-value = 0.4061
```

직접효과의 경우 정규성 검정 결과  $p > .05$ 로 직접효과의 분포가 정규분포라는 귀무가설을 기각할 수 없다.

참고로 processR에서는 부트스트랩 방법으로 norm, basic, perc, bca.simple의 네 가지를 지원한다. medSummaryTable()함수를 호출할 때 boot.ci.type인수를 "all"로 지정하는 경우 네 가지 부트스트랩 방법에 따른 95% 신뢰구간을 모두 볼 수 있다. 방법에 따라 약간의 차이는 있으나 비슷한 신뢰구간을 보여 준다.

```
medSummaryTable(semfit,boot.ci.type="all")
```

	Indirect Effect			Direct Effect		
		(a)*(b)		c		
type	estimate	95% Bootstrap CI	p	estimate	95% Bootstrap CI	p
norm	0.241	(-0.018 to 0.499)	.067	0.254	(-0.229 to 0.764)	.315
basic	0.241	(-0.043 to 0.479)	.067	0.254	(-0.257 to 0.760)	.315
perc	0.241	(0.003 to 0.526)	.067	0.254	(-0.251 to 0.766)	.315
bca.simple	0.241	(0.007 to 0.534)	.067	0.254	(-0.186 to 0.792)	.315