

계량분석

Moderating Variables and Interaction Terms

김현우, PhD¹

¹충북대학교 사회학과 조교수

November 15, 2021

진행 순서

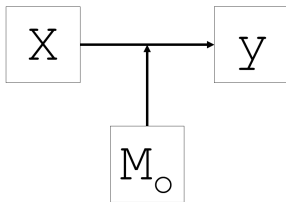
- 1 조절변수와 상호작용항
- 2 조절변수의 연습

조절변수와 상호작용항

조절변수와 상호작용항

조절변수(moderating variables)를 통해 조건부 효과(conditional effects)를 살펴볼 수 있다.

- 어떤 독립변수가 종속변수에 대해 갖는 효과가 “조절변수에 따라 조건부(conditional)로 상이하게 작동하는” 경우가 있을 수 있다.
- 독립변수(X)가 종속변수(y)에 미치는 영향력이 단선적이지 않고, 조절변수(M_o)의 개입(intervention)에 따라 달라지는 영향력을 살펴볼 필요가 있는 셈이다.
- 독립변수(X)와 조절변수(M_o) 각각의 주효과(main effects)를 살펴보고, 그 위에 두 변수가 얹혀 만들어내는 상호작용 효과(interaction effect)도 추가적으로 살펴보게 된다.



조절변수와 상호작용항

노조원 신분(M_o)은 임금(y)과 재직경력(X) 사이의 관계를 어떻게 “조절”하는지 살펴보자.

- 노조원(union)의 경우 비노조원보다 더 많은 임금을 보장받을 것이다(이른바 union premium).
- 재직경력(tenure)이 증가하면 임금은 더 커질 것이다.
- 노조원(union)의 경우 재직경력(tenure)이 없거나 짧더라도 상대적으로 더 많은 임금을 보장받을 것이다. 그러므로 y 축(임금) 절편이 비노조원보다 높으리라고 예상할 수 있다.
- 다른 한편, 이미 초봉을 높게 출발했으므로 tenure가 쌓이더라도 그 상승분은 비노조원보다 적을 것이다. 그렇다면 노조원의 tenure 회귀계수는 비노조원의 tenure 회귀계수보다 작으리라고 예상할 수 있다.
- 이 사례에서 중요한 것은 (1) $\text{tenure} \rightarrow \ln_wage$ 와 (2) $\text{union} \rightarrow \ln_wage$ 의 관계 (=주효과)만을 보는데 그치지 않고, (3) $\text{tenure} \rightarrow \ln_wage$ 관과가 union에 따라 어떻게 달라지는가(=상호작용 효과)도 함께 보는 것이다.

조절변수와 상호작용항

- 결국, 독립변수인 재직경력(TENURE)와 조절변수인 노조원 신분(UNION) 외에도 두 변수를 곱한 새로운 TENURE · UNION 변수가 포함되어 있다. 이 변수를 특별히 **상호작용항(interaction term)**이라고 한다. 나머지 두 변수를 **주요항(main terms)**이라고 한다.
- 위 회귀모형의 상호작용항 회귀계수(\hat{b}_3)에 대해 유의성 검정($H_0 : \beta_3 = 0$)을 하여, 이것이 만일 통계적으로 유의하다면 **상호작용 효과(interaction effect)**가 존재한다고 말할 수 있다. 즉, 모집단에서도 재직경력의 임금 효과가 노조원 신분에 따라 달라지는 것이다.
- **상호작용 효과를 살펴보기 위해 상호작용항(TENURE · UNION)을 넣을 때는 두 독립변수(UNION과 TENURE)를 “반드시” 함께 모델 안에 넣어야 한다(Why?)**

조절변수와 상호작용항

- 상호작용항을 포함한 회귀모형이 아래와 같았다.

$$\text{LN_WAGE}_i = b_0 + b_1 \text{TENURE}_i + b_2 \text{UNION}_i + b_3 (\text{TENURE}_i \cdot \text{UNION}_i) + e_i$$

- 상호작용항을 포함한 회귀모형의 해석은 더미변수의 해석과 크게 다르지 않다.

$$\text{LN_WAGE}_i = \hat{b}_0 + \hat{b}_1 \text{TENURE}_i \quad (\text{if UNION}=0)$$

$$\begin{aligned} \text{LN_WAGE}_i &= \hat{b}_0 + \hat{b}_1 \text{TENURE}_i + \hat{b}_2 + \hat{b}_3 \text{TENURE}_i \quad (\text{if UNION}=1) \\ &= (\hat{b}_0 + \hat{b}_2) + (\hat{b}_1 + \hat{b}_3) \text{TENURE}_i \end{aligned}$$

- 노조원이 아닌 경우 단순히 상수(\hat{b}_0)와 회귀계수(\hat{b}_1)를 가지고 해석한다.
- 노조원인 경우 TENURE의 변화에 따른 LN_WAGE가 변화 정도에는 매번 $\hat{b}_3 \text{TENURE}$ 만큼의 부스터가 더해진다. 여기서 주목할 부분은 UNION에 따라 TENURE의 회귀계수와 상수 부분 모두 달라진다는 점이다.

조절변수와 상호작용항

Stata에서 상호작용항의 생성 방법을 연습하자([Stata 코드] 참고).

- 상호작용항은 결국 독립변수와 조절변수를 곱해서 쉽게 만들수 있다. 그 뒤, 새롭게 곱한 변수를 기존의 두 변수와 함께 모형에 넣으면 된다.
- 좀 더 세련된 방법은 #나 ##를 붙이는 **이항 연산자(binary operator)**를 사용하는 것이다. 어느 쪽이던 결과는 똑같다.
- 상호작용항이 통계적으로 유의한지 확인해보자. 단순히 회귀분석 결과표만 가지고는 상호작용항의 실질적인 의미를 해석하기 곤란할 때가 많다. 실제로 표만 보고 상호작용항을 해석해보라.
- Stata의 **margins**와 **marginsplot** 명령어를 활용해 상호작용효과를 그림으로 표현하는 것이 바람직하다. 실제로 그래프를 그려 상호작용항을 해석해보라.

조절변수와 상호작용항

두 번째 유형으로 독립변수와 조절변수 둘 다 숫자형 변수인 경우를 살펴보자([Stata 코드] 참고).

- 독립변수는 재직경력(TTL_EXP), 조절변수는 나이(AGE)이고 상호작용항을 포함한 회귀모형을 다음과 같이 구성하자.

$$\text{LN_WAGE}_i = b_0 + b_1 \text{TTL_EXP}_i + b_2 \text{AGE}_i + b_3 (\text{TTL_EXP}_i \cdot \text{AGE}_i) + e_i$$

- 해석에서 혼동하기 쉬우므로 약간 주의해야 한다. 핵심은 조절변수를 마치 범주형처럼 판단하는 것이다. 먼저 대표적인 나이로 30세를 기준으로 상호작용 효과를 계산한다.

$$\begin{aligned} \text{LN_WAGE}_i &= \hat{b}_0 + \hat{b}_1 \text{TTL_EXP}_i + \hat{b}_2 \cdot 30 + \hat{b}_3 (\text{TTL_EXP}_i \cdot 30) \\ &= (\hat{b}_0 + \hat{b}_2 \cdot 30 + \hat{b}_3 \cdot 30) + (\hat{b}_1 + \hat{b}_3) \cdot \text{TTL_EXP}_i \end{aligned}$$

조절변수와 상호작용항

마지막 유형으로 독립변수와 조절변수 모두가 더미 변수인 경우를 살펴보자([Stata 코드] 참고).

- 독립변수는 남부(SOUTH), 조절변수는 노조원 신분(UNION)이고 상호작용항을 포함한 회귀모형을 다음과 같이 구성하자.

$$\text{LN_WAGE}_i = b_0 + b_1 * \text{SOUTH}_i + b_2 \text{UNION}_i + b_3(\text{SOUTH}_i \cdot \text{UNION}_i) + e_i$$

- {0, 1}로 더미 코딩하는 이상 이 유형도 매우 쉽게 해석할 수 있다.

$$\text{LN_WAGE}_i = \hat{b}_0 \quad (\text{if } \text{UNION}=0 \text{ \& } \text{SOUTH}=0)$$

$$\text{LN_WAGE}_i = \hat{b}_0 + \hat{b}_1 \quad (\text{if } \text{UNION}=0 \text{ \& } \text{SOUTH}=1)$$

$$\text{LN_WAGE}_i = \hat{b}_0 + \hat{b}_2 \quad (\text{if } \text{UNION}=1 \text{ \& } \text{SOUTH}=0)$$

$$\text{LN_WAGE}_i = \hat{b}_0 + \hat{b}_1 + \hat{b}_2 + \hat{b}_3 \quad (\text{if } \text{UNION}=1 \text{ \& } \text{SOUTH}=1)$$

조절변수의 연습

조절변수의 연습

송리라·이민아(2013)를 보고 조절변수가 활용되는 방식을 살펴보자.

- 종속변수, 독립변수, 그리고 조절변수는 각각 무엇인가(97-98페이지)?
- 본문의 <표 2>에서 우울 및 성역할 태도는 각각 남녀 어느 쪽에서 높은가(165페이지)? 연령집단별 통계적으로 유의한 차이가 있는가?
- 본문의 <표 4>에 따르면 상호작용항(성역할태도 \times 성별)이 통계적으로 유의한 것으로 나타나 있다. 이를 그림과 함께 해석해 보라(103-104페이지).

송리라 · 이민아, 2012. “성역할 태도와 우울,” *한국인구학* 35(3): 87-116.

조절변수의 연습

제시된 <동아행복지수> 신문 기사를 읽고 이 아이디어를 재현(replicate)해보자([Stata 코드] 참고).

- eCampus에서 2003-2018_KGSS_public_v3.dta 데이터를 다운받아 위 신문기사에서 제시한 첫번째 가설을 테스트해보자.
- 이 신문기사에서 **연구가설(research hypothesis)**은 무엇인가? (1) 여성은 나이를 먹을수록 행복해진다; (2) 결혼하지 않은 여성은 결혼한 여성에 비해 나이를 먹으면서 더 빨리 행복해진다.
- 종속변수(y)은 무엇을 사용할 수 있는가? 독립변수(X)은 무엇을 사용할 수 있는가? 조절변수(M_o)는 무엇을 사용할 수 있는가? 이유는 무엇인가?
- 상호작용항(M_o)을 포함한 회귀모형을 작성하시오.
- 실제로 분석을 수행하고 상호작용 효과가 통계적으로 유의한지 밝히고 해석하시오.

