계량분석

Moderating Variables and Interaction Terms

김현우, PhD¹

1 충북대학교 사회학과 조교수

November 15, 2021

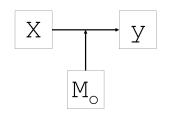


진행 순서

- 조절변수와 상호작용항
- ② 조절변수의 연습

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 얽혀 만들어내는 상호작용 효과(interaction effect)도 추가적으로 살펴보게 되다.





- 단순한 선형관계로 이루어진 가설과 이론으로 만족하지 않고, 한층 더 복합적인 분기 (junction)와 상호작용을 고려한 세련된 형태로 발전시킬 수 있다.
- 사실 우리는 이미 삼원교차표(three-way contingency table)를 연습하면서 이미 이론적인 고민을 통해 독립변수와 종속변수의 사이에 존재하는 조절변수를 찾아내 보았다. 우리 예제에서는 독립변수로 "when jobs are scarce, employers should give priority to people of this country over immigrants"를, 종속변수로 "(이민자가 늘어나면) Increases the crime rate"를 설정하였다. 두 변수 사이의 연관성을 교차표로 분석할 때, "Losing my job or not finding a job"가 조절변수로 작용하는가를 살펴보았다.
- 삼원교차표를 통해 상호작용 효과를 부분적으로 살펴볼 수 있었지만, 회귀모형속에서 훨씬 세련된 형태로 이를 검증할 수 있다.



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 y축(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) tenure → ln_wage 와 (2) union → ln_wage 의 관계 (=주효과)만을 보는데 그치지 않고, (3) tenure → ln_wage 관과가 union에 따라 어떻게 달라지는가(=상호작용 효과)도 함께 보는 것이다.



 이 예제에서 조절변수의 영향력을 살펴보기 위해 먼저 다음과 같이 회귀모형을 구성하자.

$$LN_WAGE_i = b_0 + b_1TENURE_i + b_2UNION_i + e_i$$

• 이때, 재직경력(TENURE)과 노조원 신분(UNION) 변수가 포함되어 있지만, 이것들을 통해서는 각각의 개별적인 효과만을 볼 수 있을 뿐이다. 그러나 만일 노조원 신분에 따라 재직경력의 영향력이 변화한다면, 재직경력의 효과(=회귀계수)는 다음과 같이 두 부분으로 분해될 수 있다.

$$b_1 = \gamma_0 + \gamma_1 UNION_i$$

• 두 식을 결합하면 아래와 같다.

$$\begin{split} LN_WAGE_i &= b_0 + \gamma_0 TENURE_i + b_2 UNION_i + \gamma_1 (TENURE_i \cdot UNION_i) + e_i \\ &= b_0 + b_1 TENURE_i + b_2 UNION_i + b_3 (TENURE_i \cdot UNION_i) + e_i \end{split}$$



- 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 희귀모형의 상호작용항 회귀계수(b̂₃)에 대해 유의성 검정(H₀: β₃ = 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?)



• 상호작용항을 포함한 회귀모형이 아래와 같았다.

$$LN_WAGE_i = b_0 + b_1 TENURE_i + b_2 UNION_i + b_3 (TENURE_i \cdot UNION_i) + e_i$$

• 상호작용항을 포함한 회귀모형의 해석은 더미변수의 해석과 크게 다르지 않다.

$$\begin{split} LN_WAGE_i &= \widehat{b}_0 + \widehat{b}_1 TENURE_i & \text{ (if UNION==0)} \\ LN_WAGE_i &= \widehat{b}_0 + \widehat{b}_1 TENURE_i + \widehat{b}_2 + \widehat{b}_3 TENURE_i & \text{ (if UNION==1)} \\ &= (\widehat{b}_0 + \widehat{b}_2) + (\widehat{b}_1 + \widehat{b}_3) TENURE_i & \end{split}$$

- 노조원이 아닌 경우 단순히 상수(b̂₀)와 회귀계수(b̂₁)를 가지고 가지고 해석한다.
- 노조원인 경우 TENURE의 변화에 따른 LN_WAGE가 변화 정도에는 매번 \hat{b}_3 TENURE 만큼의 부스터가 더해진다. 여기서 주목할 부분은 UNION에 따라 TENURE의 회귀계수와 상수 부분 모두 달라진다는 점이다.



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 margins와 marginsplot 명령어를 활용해 상호작용효과를 그림으로 표현하는 것이 바람직하다. 실제로 그래프를 그려 상호작용항을 해석해보라.



위 예제에서 독립변수(TENURE)는 숫자형 변수였고, 조절변수(UNION)는 더미변수였다.

- 그러므로 독립변수와 조절변수의 척도가 어떻게 구성되어 있는가에 따라 네 가지 유형이 나올 수 있을 것이다:
 - (1) 숫자형 변수 × 더미변수
 - (2) 숫자형 변수 × 숫자형 변수
 - (3) 더미변수 × 더미변수
 - (4) 더미변수 × 숫자형 변수.
- 그런데 회귀모형에서 조절변수이니 통제변수니 관심변수니 하는 건 모두 주관적인 문제일 뿐이고, 실제 모델 안에는 들어가는 방식에는 차이가 없다. 따라서 사실상 (4) 는 (1)과 중복이므로 첫 세 가지 유형만 고려하면 된다.
- 위 예제는 (1)의 유형이였다.



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.

$$LN_WAGE_i = b_0 + b_1TTL_EXP_i + b_2AGE_i + b_3(TTL_EXP_i \cdot AGE_i) + e_i$$

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

$$\begin{split} LN_WAGE_i &= \widehat{b}_0 + \widehat{b}_1TTL_EXP_i + \widehat{b}_2 \cdot 30 + \widehat{b}_3(TTL_EXP_i \cdot 30) \\ &= (\widehat{b}_0 + \widehat{b}_2 \cdot 30 + \widehat{b}_3 \cdot 30) + (\widehat{b}_1 + \widehat{b}_3) \cdot TTL_EXP_i \end{split}$$



• 유사하게 연령이 40인 경우를 대표로 삼아 상호작용 효과를 구한다.

$$\begin{split} LN_WAGE_i &= \widehat{b}_0 + \widehat{b}_1TTL_EXP_i + \widehat{b}_2 \cdot 40 + \widehat{b}_3(TTL_EXP_i \cdot 40) \\ &= (\widehat{b}_0 + \widehat{b}_2 \cdot 40 + \widehat{b}_3 \cdot 40) + (\widehat{b}_1 + \widehat{b}_3) \cdot TTL_EXP_i \end{split}$$

• 한 번 더 연령이 50인 경도 대표로 삼아 상호작용 효과를 구한다.

$$\begin{split} LN_WAGE_i &= \widehat{b}_0 + \widehat{b}_1TTL_EXP_i + \widehat{b}_2 \cdot 50 + \widehat{b}_3(TTL_EXP_i \cdot 50) \\ &= (\widehat{b}_0 + \widehat{b}_2 \cdot 50 + \widehat{b}_3 \cdot 50) + (\widehat{b}_1 + \widehat{b}_3) \cdot TTL_EXP_i \end{split}$$

- 조절변수를 숫자형 변수로 그대로 두면 무척 해석이 까다로우므로 대표적인 범주 몇 개만을 골라 거기에 초점을 두고 해석하는 것이 요령이다!
- 앞서와 마찬가지로 Stata의 margins와 marginsplot 명령어를 활용해 상호작용 효과를 그래프로 표현하고 해석하는 것이 바람직하다.
- 사실 숫자형 변수 × 숫자형 변수의 상호작용 효과를 해석하기는 제법 까다로우므로 조절변수를 아예 처음부터 범주화하는 경우도 제법 있다.



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.

$$LN_WAGE_i = b_0 + b_1 * SOUTH_i + b_2 UNION_i + b_3 (SOUTH_i \cdot UNION_i) + e_i$$

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

$$LN_WAGE_i = \hat{b}_0$$
 (if UNION==0 & SOUTH==0)
 $LN_WAGE_i = \hat{b}_0 + \hat{b}_1$ (if UNION==0 & SOUTH==1)

$$LN_WAGE_i = \hat{b}_0 + \hat{b}_2$$
 (if UNION==1 & SOUTH==0)

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



- 부스터 효과에 초점을 맞추어 해석한다면 남부이고(south==1) 노조원(union==1)인 경우 .035 만큼 로그임금의 "추가적인 감소"가 관찰된다(이 값은 상호작용항의 회귀계수이다). 이 상호작용 효과는 통계적으로 유의하다. 이 "추가적인 감소"는 둘 중 하나의 조건만 없었어도 나타나지 않았을 현상이었다.
- 독립변수와 조절변수가 둘 다 더미변수인 경우에는 일반적인 그래프와 살짝 다르게 그리는 편이 좋다(Why?). Stata의 marginsplot 명령어의 옵션도 조금 다르다.



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(165 페이지)? 연령집단별 통계적으로 유의한 차이가 있는가?
- 본문의 〈표 4〉에 따르면 상호작용항(성역할태도 × 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〉 신문기사를 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.

