

계량분석

Moderating Variables and Interaction Terms

김현우, PhD¹

¹충북대학교 사회학과 조교수

November 7, 2024



진행 순서

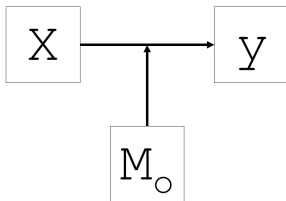
- 1 조절변수와 상호작용항
- 2 조절변수의 연습
- 3 연습문제

조절변수와 상호작용항

조절변수와 상호작용항

조절변수를 통해 조건부 효과를 살펴볼 수 있다.

- 어떤 독립변수가 종속변수에 대하여 **조절변수(moderating variables)**에 따라 조건부(conditional)로 상이한 효과를 가질 수도 있다.
- 독립변수 X 의 종속변수 Y 에 미치는 영향력이 단선적이지 않고, 조절변수 M_o 의 **개입(intervention)**에 따라 달라지는 영향력을 살펴보게 된다.



조절변수와 상호작용항

- 독립변수 X 와 조절변수 M 의 각각의 주효과(main effects)를 살펴보고, 그 위에 두 변수가 얹혀 만들어내는 상호작용 효과(interaction effect) 내지 조건부 효과(conditional effects)도 함께 살펴보게 된다.
- 단순한 선형관계로 이루어진 가설과 이론으로 만족하지 않고, 한층 더 복잡한 분기(junction)와 상호작용을 고려한 가설과 이론을 만들어낼 수 있다.
- 실제로 상호작용 효과를 살펴보는 수많은 석박사 논문이 존재한다!



조절변수와 상호작용항

노조원 신분은 임금과 재직경력 사이의 관계를 조절하는지 살펴보자.

- 재직경력 *tenure*가 증가하면 임금 *ln_wage*는 더 커질 것이다.
- 노조원의 경우 비노조원보다 더 많은 임금을 보장받을 것이다. 이른바 **노조 프리미엄 (union premium)**이 그것이다.
- 노조원의 경우 재직경력이 없거나 짧더라도 상대적으로 더 많은 임금을 보장받을 것이다. 그러므로 노조원의 *y*축(임금) 절편은 비노조원의 그것보다 높으리라고 예상할 수 있다.
- 다른 한편, 이미 초봉을 높게 출발했으므로 재직경력이 쌓이더라도 그 상승분은 비노조원보다 적을 것이다. 그렇다면 노조원의 *tenure* 회귀계수는 비노조원의 그것보다 작으리라고 예상할 수 있다.
- 중요한 것은 (주효과 만을 보는데 그치지 않고), *tenure* → *ln_wage* 관계가 union에 따라 어떻게 달라지는가(=상호작용 효과)도 함께 본다는 것이다.



조절변수와 상호작용항

- 먼저 다음과 같이 회귀모형을 설정해보자.

$$LN_WAGE = b_0 + b_1TENURE + b_2UNION + e$$

- 재직경력 TENURE와 노조원 신분 UNION이 포함되어 있지만, 오로지 개별적인 효과만을 볼 수 있을 뿐이다.
- 그런데 만일 UNION에 따라 TENURE의 영향력이 변화한다면, TENURE의 효과 (=회귀계수)는 다음과 같이 추가적으로 쓸 수 있다.

$$b_1 = \gamma_0 + \gamma_1UNION$$

- 두 식을 결합하면 아래와 같다.

$$LN_WAGE = b_0 + \gamma_0TENURE + b_2UNION + \gamma_1(TENURE \cdot UNION) + e$$



조절변수와 상호작용항

- TENURE와 UNION 외에도 (두 변수를 곱한) TENURE·UNION가 포함되어 있다.
- 이 변수를 특별히 **상호작용항(interaction term)**이라고 한다. 나머지 두 변수를 **주요항(main terms)**이라고 한다.
- 위 회귀모형의 상호작용항 회귀계수 $\hat{\gamma}_1$ 에 대해 유의성 검정($H_0 : \gamma_1 = 0$)을 하여, 이것이 만일 통계적으로 유의하다면 **상호작용 효과(interaction effect)**가 존재한다고 말할 수 있다(모집단에서도 재직경력의 임금 효과가 노조원 신분에 따라 다르다).
- 상호작용 효과를 살펴보기 위해 상호작용항을 넣을 때는 두 변수를 반드시 함께 회귀모형 안에 넣어야 한다(Why?).



조절변수와 상호작용항

- 상호작용항을 포함한 회귀모형을 아래와 같이 다시 쓸 수 있다.

$$LN_WAGE = b_0 + b_1TENURE + b_2UNION + b_3(TENURE \cdot UNION) + e$$

- 상호작용항을 포함한 회귀모형의 해석은 가변수의 해석과 크게 다르지 않다.

$$LN_WAGE = \hat{b}_0 + \hat{b}_1TENURE \quad (\text{if } UNION=0)$$

$$LN_WAGE = \hat{b}_0 + \hat{b}_1TENURE + \hat{b}_2 + \hat{b}_3TENURE \quad (\text{if } UNION=1)$$

$$= (\hat{b}_0 + \hat{b}_2) + (\hat{b}_1 + \hat{b}_3)TENURE$$

- 노조원이 아닌 경우, 단순히 상수 \hat{b}_0 와 회귀계수 \hat{b}_1 를 가지고 해석한다.
- 노조원인 경우, TENURE의 회귀계수와 상수 부분 모두 부스터가 추가된다!



조절변수와 상호작용항

Stata에서 상호작용항의 생성 방법을 연습하자.

- 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.



조절변수와 상호작용항

독립변수와 조절변수의 척도에 따라 세 가지 유형이 나온다.

- (1) 양적변수 \times 가변수, (2) 양적변수 \times 양적변수, (3) 가변수 \times 가변수.
- 회귀모형에서 조절변수이니 통제변수니 관심변수니 하는 건 모두 주관적인 문제일 뿐이고, 실제 모델 안에는 들어가는 방식에는 차이가 없므로 (4) 가변수 \times 양적변수를 추가적으로 고려할 필요는 없다.
- 위 예제에서 TENURE는 양적변수였고, UNION는 가변수이므로 (1)의 유형이었다.



조절변수와 상호작용항

독립변수와 조절변수 둘 다 양적변수인 경우를 살펴보자.

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

$$LN_WAGE = b_0 + b_1TTL_EXP + b_2AGE + b_3(TTL_EXP \cdot AGE) + e$$

- 조절변수를 양적변수로 그대로 두면 무척 해석이 까다로우므로, 대표적인 범주 몇 개만을 골라 거기에 초점을 두고 해석할 수 있다.
- 사실 양적변수 \times 양적변수의 상호작용 효과를 해석하기는 제법 까다로우므로 조절변수를 아예 처음부터 범주화할 수도 있다.



조절변수와 상호작용항

- 먼저 대표적인 나이로 20세를 기준으로 상호작용 효과를 계산해보자.

$$\begin{aligned}LN_WAGE &= \hat{b}_0 + \hat{b}_1 TTL_EXP + \hat{b}_2 \cdot 20 + \hat{b}_3 (TTL_EXP \cdot 20) \\ &= (\hat{b}_0 + \hat{b}_2 \cdot 20 + \hat{b}_3 \cdot 20) + (\hat{b}_1 + \hat{b}_3) \cdot TTL_EXP\end{aligned}$$

- 유사하게 연령이 30인 경우를 대표로 삼아 상호작용 효과를 구한다.

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

- 마지막으로 연령이 40인 경우도 대표로 삼아 상호작용 효과를 구한다.

$$\begin{aligned}LN_WAGE &= \hat{b}_0 + \hat{b}_1 TTL_EXP + \hat{b}_2 \cdot 40 + \hat{b}_3 (TTL_EXP \cdot 40) \\ &= (\hat{b}_0 + \hat{b}_2 \cdot 40 + \hat{b}_3 \cdot 40) + (\hat{b}_1 + \hat{b}_3) \cdot TTL_EXP\end{aligned}$$



조절변수와 상호작용항

- Stata에서 margins와 marginsplot을 활용해 그래프로 표현하고 해석하는 것이 바람직하다.
- 이때 at(.) 옵션을 익숙하게 활용해야 한다. 이 안에 독립변수와 조절변수를 함께 집어넣고 범위와 대표값들을 부여하였다.
- 조절변수의 기술통계를 살펴보고 나서 그것의 대표값을 정하는 편이 낫다(Why?). 가령 AGE라면 20, 30, 40으로 대표값을 정해도 괜찮은 것 같다.



조절변수와 상호작용항

마지막 유형으로 독립변수와 조절변수 모두가 가변수인 경우를 살펴보자.

- 독립변수는 남부 SOUTH, 조절변수는 노조원 신분 UNION이고 상호작용항을 포함한 회귀모형을 다음과 같이 구성하자.

$$LN_WAGE = b_0 + b_1SOUTH + b_2UNION + b_3(SOUTH \cdot UNION) + e$$

- {0, 1}로 가부호화(dummy coding)하는 이상 이 유형도 매우 쉽게 해석할 수 있다.

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

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

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

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



조절변수와 상호작용항

- 부스터 효과에 초점을 맞추어 해석한다면, 남부이고(SOUTH==1) 노조원(UNION==1)인 경우 .035 만큼 임금 LN_WAGE의 추가적인 감소가 관찰된다(이 값은 상호작용항의 회귀계수이다).
- 이 상호작용 효과는 통계적으로 유의하다. 의미를 해석해보자.
- 독립변수와 조절변수가 둘 다 가변수인 경우에는 일반적인 그래프와 살짝 다르게 그리는 편이 좋다(Why?).
- 당연히 Stata의 marginsplot도 조금 다르게 사용해야 한다. 특히 recast(·) 옵션으로 막대차트를 그려야 하며, xdim(·) 옵션으로 x 축도 다르게 지정해야 한다.



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 \times 성별)이 통계적으로 유의한 것으로 나타나 있다. 이를 <그림 1>과 함께 해석해 보라(103-104페이지).

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



조절변수의 연습

<표 3> 우울에 미치는 성별과 성역할 태도의 영향

		모형1		모형2		모형3	
		계수	표준오차	계수	표준오차	계수	표준오차
성역할 태도		0.01**	0.010	0.01**	0.01	0.01*	0.01
성별(1=여성)		0.12**	0.050	0.15***	0.05	0.10**	0.05
연령		-0.01**	0.004	-0.01**	0.01	-0.01***	0.01
결혼유무(1=결혼)		-0.15**	0.060	-0.06	0.09	0.02	0.09
교육수준	고졸(준거집단)						
	2-4년제 대졸	0.11*	0.060	-0.01	0.07	-0.06	0.06
	석사이상	-0.20**	0.100	-0.20*	0.11	-0.15	0.10
연평균가구소득		-0.02***	0.010	-0.01**	0.01	-0.01	0.01
직업형태	정규직(준거집단)						
	비정규직	0.02	0.080	0.03	0.08	0.01	0.08
	자영업	0.03	0.090	0.012	0.10	0.003	0.09
	무직	-0.09	0.070	-0.08	0.07	-0.12*	0.07
	학생	-0.22**	0.090	-0.23**	0.09	-0.23	0.09
자녀유무	없음(준거집단)						
	미취학아동			-0.13	0.10	-0.2**	0.09
	초등학생이상			-0.14	0.10	-0.17*	0.10
집안일 정도	저녁식사준비			0.03	0.02	0.03	0.02
	빨래			-0.05*	0.02	-0.04	0.02
	청소			0.002	0.03	0.05	0.03
주관적 건강평가						-0.28***	0.03
상수항		2.45***	0.190	2.38***	0.24	1.74***	0.24
표본수		969		912		912	
F		5.03***		3.52***		9.61***	
결정계수		0.0547		0.0593		0.1545	

주: * p<0.10, **p<0.05, ***p<0.01



조절변수의 연습

<표 4> 성별과 성역할 태도 간의 상호작용 효과분석

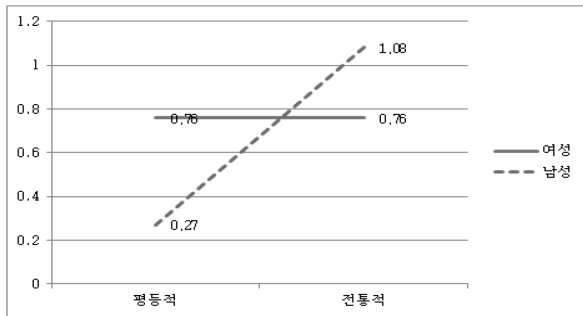
		모형	
		계수	표준오차
성역할 태도*성별		-0.030**	0.01
성역할 태도		0.030***	0.01
성별(1 = 여성)		0.760***	0.26
연령		-0.010***	0.01
결혼여부(1 = 기혼)		0.010	0.09
교육수준	고졸(준거집단)		
	2-4년제 대졸	-0.070	0.06
	석사이상	-0.170*	0.10
연평균가구소득		-0.010	0.01
직업형태	정규직(준거집단)		
	비정규직	0.020	0.08
	자영업	0.002	0.09
	무직	-0.100	0.07
	학생	-0.230	0.09***
자녀유무	없음(준거집단)		
	미취학아동	-0.190**	0.09
	초등학생이상	-0.180*	0.10
집안일 정도	저녁식사준비	0.030	0.02
	빨래	-0.030	0.02
	청소	0.010	0.03
주관적 건강평가		-0.280***	0.03
상수항		1.330***	0.29
표본수		912	
F		9.50***	
결정계수		0.1607	

주: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$



조절변수의 연습

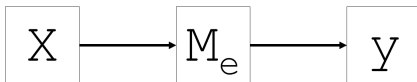
<그림 1> 성별과 성역할 태도에 따른 우울



조절변수의 연습

조절효과를 공부하기 위해 읽어야 할 많은 논문들이 있다.

- 구조방정식모형(structural equation modeling; SEM)을 나중에 공부하면 좀 더 수준있게 조절효과를 다룬다.
- 조절효과(moderation effect)와 매개효과(mediation effect)를 혼동해선 안된다.



- 좀 더 고급으로 나아가면 매개된 조절효과(mediated moderation effect)나 조절된 매개효과(moderated mediation effect)도 모형화할 수 있다(정선호·서동기 2016).

정선호·서동기. 2016. “회귀분석을 이용한 매개된 조절효과와 조절된 매개효과 검증 방법.” 『한국심리학회지: 일반』 35(1): 257-282.



조절변수의 연습

- Brambor et al. (2006)과 Berry (2012)는 정치학 문헌에서 상호작용 효과 해석의 원리와 오류, 시각화 방법 등을 꼼꼼히 설명한다. Stata 코드 부록도 참고할 것.
- 심리학 계열에서는 **Johnson-Neyman 방법**에 따라 조절효과를 살펴보기도 한다. 수많은 책과 논문이 있으므로 적절히 살펴보자(e.g., 고길곤 2021; 박인숙·이재림 2023).

고길곤. 2021. 『매개효과와 조절효과』. 문우사.

박인숙·이재림. 2023. “초등학교 저학년 자녀의 학교생활 파악과 취업모의 일-자녀교육 갈등의 관계: 양육역량의 조절효과.” 『가정과 삶의 질 연구』 41(3): 17-30.

Berry, William D., Matt Golder, Daniel Milton. 2012. “Improving Tests of Theories Positing Interaction.” *Journal of Politics* 74(3): 653-671.

Brambor, Thomas, William Roberts Clark, Matt Golder. 2006. “Understanding Interaction Models: Improving Empirical Analyses.” *Political Analysis* 14(1): 63-82.



연습문제

연습문제

연습 1. 여성 표본에서 연령의 행복효과에 결혼 여부가 미치는 조절 효과를 살펴보고자 한다. hap을 활용하여 다음의 질문에 모두 답하시오.

- hapinss1을 종속변수로, age와 married의 상호작용 효과를 살펴보세요(married는 marital에서 재부호화해야 한다).
- margins와 marginsplot을 사용하여 시각화하시오. 이를 언어로 해석하시오.
- 유의성 검정을 수행하고 어떤 조절효과에 관한 연구가설을 검정하고 있는지 명시적으로 해석하시오.

