

# 사회통계연습

교차표 분석의 활용

김현우, PhD<sup>1</sup>

<sup>1</sup>충북대학교 사회학과 부교수



# 진행 순서

- 1 교차표에서 범주의 통폐합
- 2 평균값을 나타내는 교차표
- 3 마르코프 체인



## 교차표에서 범주의 통폐합



# 교차표에서 범주의 통폐합

질적변수가 아니라 양적변수의 경우에도 교차표를 그릴 수 있을까?

- 물론이다. 정보의 손실(information loss)을 감수한다면 양적변수를 질적변수로 얼마든지 **재부호화(recoding)**할 수 있다. 그러나 그 역은 불가능하다.
- 엑셀에서는 “그룹” 기능으로 쉽게 범주화할 수 있었고, SPSS에서도 “다른 변수로 코딩변경”을 활용하여 범주화할 수 있다.
- 다만 (범주를 설정할 때) 너무 많은 셀에서 0에 가까운 값이 나오지 않도록 주의할 기울여야 한다!



# 교차표에서 범주의 통폐합

표본 크기가 작는데 셀에 아주 작은 숫자만 들어간다면 문제가 된다.

- 먼저 특정 셀에 아주 작은 숫자(특히 5보다 작은 수)가 들어간다는 것의 의미를 고민해야 한다.
- 이 문제는 아주 작은 표본( $n < 40$ )에 기반하여  $2 \times 2$  교차표를 만들 때 제기되는 문제이다. 사실 오늘날 사회과학에서는 좀처럼 이런 상황이 없다.
- ( $2 \times 2$ 보다) 큰 교차표에서는 최소한 80% 이상의 셀이 5보다 큰 값을 가져야 한다.
- 이런 경우 다음의 세 가지 대응방법이 있다.
  - (1) 범주를 다시 만든다.  $2 \times 2$  교차표에서는 이 방법이 유효하지 않다.
  - (2) ( $\chi^2$  독립성 검정에서) Yates의 연속성 보정(continuity correction)을 한다.
  - (3) ( $\chi^2$  독립성 검정 대신) Fisher의 정확 검정(exact test)을 사용한다.



# 교차표에서 범주의 통폐합

연습 1. ASAN.SAV에서 향후 10년 내 한반도에서의 무력충돌 발생가능성과 주한미군의 한국주둔 필요성 인지 사이의 연관성에 대해 살펴보고 유의성 검정을 수행하시오. 그 다음, 무력충돌 발생가능성을 3개의 범주로 재부호화한 뒤 같은 방식으로 유의성 검정을 다시 수행하시오. 두 분석의 차이에 대해 논하시오.



# 교차표에서 범주의 통폐합

- 응답 범주를 통폐합하여 교차표의 행렬을 줄이거나 늘리면  $\chi^2$  검정통계량과 유의도 검정 결과가 달라진다. 통계 분석 자체는 거짓말을 한 것이 아니다.
- 다만 연구자가 자신에게 유리한 결과를 이끌어내기 위해 이런 재부호화를 시도할 수도 있다. 이것은 연구윤리 위반인가?
- 확실한 것은 연구자와 독자 사이에 분석 과정에 관한 정보 비대칭(information asymmetry)이 제법 크다는 사실이다(Why?).



## 평균값을 나타내는 교차표





# 평균값을 나타내는 교차표

Durhkeim의 〈자살론〉 교차표는 살짝 다른 종류의 것이다.

- Durkheim (1951[1897] 208)의 표를 잠깐 살펴보자.

<i>Suicides per million inhabitants</i>		
	<i>Urban population</i>	<i>Rural population</i>
1866-69	202	104
1870-72	161	110

- Jamovi에서는 [분석]-[탐구]-[기술통계]에서 평균값이라도 구해볼 수 있다. 그러나 교차표는 직접 만들어야 한다.
- SPSS에서는 [분석]-[표]-[사용자 정의 표]에서 만든다. 변수의 측도를 ‘척도(scale)’로 바꿔두어야만 평균을 보고할 수 있음에 주의해야 한다.
- 이런 종류의 교차표가 앞서 배운 교차표와 어떻게 다른지 꼼꼼하게 고민할 것! 문제가 주어졌을 때 어떤 교차표를 만들어야 하는지 먼저 파악해야 한다.
- 이러한 교차표에 대해서는 보통  $\chi^2$  검정을 수행하지 않는다.



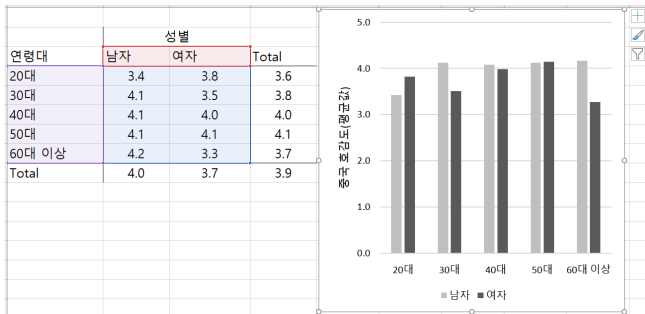
# 평균값을 나타내는 교차표

연습 2. ASAN.SAV를 사용하여 연령대와 성별에 따라 상이한 평균 중국 호감도를 표로 나타내시오.



# 평균값을 나타내는 교차표

- Jamovi에서 평균값을 구한 뒤 직접 만들거나 아니면 엑셀에서 피벗 테이블(pivot table)을 사용하자. SPSS에서는 곧바로 만들 수 있다.



# 마르코프 체인



# 마르코프 체인

사회동태의 구조를 교차표로 나타낼 수도 있다.

- 마르코프 체인(Markov Chain)은 상태변동 또는 이행(transition)의 확률을 교차표로 나타낸 것이다.
- 사회학에서는 세대간 직업이동(intergenerational occupational mobility)과 관련한 분석기법으로 잘 알려져 있다.
- 특히 직업이동 유출표(outflow table)에서 행(row)은 투입(input)이고 열(column)은 산출(output)로 생각할 수 있다.



〈표 1〉 1, 2, 3차 불평등조사의 세대간 직업이동표의 유출률 비교: 1990-2000년

1. 1990년

	정신근로자	육체근로자	농민	계
정신근로자	66.1%	30.9%	3.0%	100.0%
육체근로자	39.2%	57.3%	3.5%	100.0%
농민	37.1%	36.0%	26.8%	100.0%

2. 1995년

	정신근로자	육체근로자	농민	계
정신근로자	61.5%	36.0%	2.5%	100.0%
육체근로자	37.5%	60.1%	2.4%	100.0%
농민	36.7%	43.8%	20.5%	100.0%

3. 2000년

	정신근로자	육체근로자	농민	계
정신근로자	68.4%	28.7%	3.0%	100.0%
육체근로자	43.2%	52.9%	3.9%	100.0%
농민	34.3%	48.3%	17.5%	100.0%

주: 행과 열은 각각 아버지와 아들의 직업을 가리킴.

차종천. 2002. “최근 한국사회의 사회이동 추세: 1990-2000.” 『한국사회학』 36(2): 1-22.



# 마르코프 체인

- 행렬의 곱셈을 알면 앞으로 직업이 어떻게 변화하게 될지 예상할 수 있다.

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \times \begin{bmatrix} 10 & 11 \\ 20 & 21 \\ 30 & 31 \end{bmatrix}$$
$$= \begin{bmatrix} 1 \times 10 + 2 \times 20 + 3 \times 30 & 1 \times 11 + 2 \times 21 + 3 \times 31 \\ 4 \times 10 + 5 \times 20 + 6 \times 30 & 4 \times 11 + 5 \times 21 + 6 \times 31 \end{bmatrix}$$
$$= \begin{bmatrix} 140 & 146 \\ 320 & 335 \end{bmatrix}$$

- 만약 2000년에 정신근로자가 100명, 육체근로자가 120명, 농민이 60명이였다면, 5년 뒤에는 각각 140.8명, 121명, 18명이 된다!

$$\begin{bmatrix} 100 & 120 & 60 \end{bmatrix} \times \begin{bmatrix} 0.684 & 0.287 & 0.03 \\ 0.432 & 0.529 & 0.039 \\ 0.343 & 0.483 & 0.175 \end{bmatrix} = \begin{bmatrix} 140.82 & 121.16 & 18.18 \end{bmatrix}$$



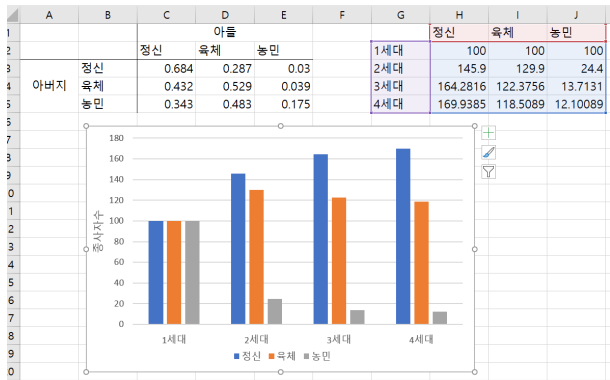
연습 3. 차종전(2002)에서 추정된 2000년 유출표를 기준으로 4세대가 경과했을 때 우리나라의 직업구조를 예측하시오(최초 구성원은 각각 100명, 100명, 100명이라고 가정하자).





# 마르코프 체인

- 행렬의 곱셈은 Python이나 MATLAB 같은 전문적인 프로그래밍 언어 또는 계산 프로그램을 사용해야 한다.
- 그래도 엑셀에서 MMULT() 함수를 사용하면 간단한 행렬 곱셈을 수행할 수 있다.



연습 4. MCHAIN.SAV 자료를 토대로 불러들여 직업이동의 유출표를 작성하시오.



# 마르코프 체인

- 행렬의 대각선 부분에 주의하여 살펴보자. 이것은 무엇을 의미하는가?
- 상당히 큰 자료를 사용하지 않으면 각 셀 안의 사례수가 너무 적아 신뢰하기 어렵다.
- 사회가 격변을 겪게 되면 직업이동 유출표의 안정성은 의심받을 수 밖에 없다(Why?).

		본인 직업(2016)										전체
		1	2	3	4	5	6	7	8	9	10	
아버지 직업 (2016)	1. 관리자	0.16	0.00	<b>0.42</b>	0.32	0.00	0.05	0.00	0.05	0.00	0.00	1.00
	2. 전문가 및 관련 종사자	0.05	<b>0.39</b>	0.28	0.09	0.09	0.04	0.02	0.04	0.02	0.00	1.00
	3. 사무 종사자	0.02	0.25	<b>0.34</b>	0.14	0.14	0.00	0.03	0.03	0.06	0.00	1.00
	4. 서비스 종사자	0.04	0.07	0.24	<b>0.31</b>	0.20	0.00	0.04	0.00	0.09	0.00	1.00
	5. 판매 종사자	0.03	0.11	0.24	0.16	<b>0.32</b>	0.03	0.00	0.03	0.11	0.00	1.00
	6. 농림어업 숙련 종사자	0.02	0.08	0.16	0.13	0.13	<b>0.19</b>	0.06	0.07	0.17	0.00	1.00
	7. 기능원 및 관련 기능 종사자	0.04	0.14	<b>0.24</b>	0.11	0.16	0.00	0.14	0.05	0.13	0.00	1.00
	8. 장치,기계조작 및 조립종사자	0.03	0.10	<b>0.22</b>	0.12	0.22	0.00	0.03	0.14	0.10	0.02	1.00
	9. 단순노무 종사자	0.00	0.10	0.17	0.10	0.17	0.03	0.10	0.17	<b>0.17</b>	0.00	1.00
	10. 군인	0.00	0.13	0.13	<b>0.38</b>	0.13	0.13	0.00	0.00	0.13	0.00	1.00
전체		0.03	0.13	0.22	0.14	0.15	0.09	0.05	0.06	0.12	0.00	1.00



# 마르코프 체인

- 직업이동 유출표 외에도 여러분이 관심있는 주제에서 어떤 식으로 표를 만들 수 있는지 상상해보자.
- (1) 광역 지자체간 **인구 이동(population mobility)**, (2) **브랜드 이동(brand switching)** 구조 분석, (3) **부동층(swing voters)**의 지지후보의 변화 등을 측정하는데 마르코프 체인이 활용되기도 한다.

