

# 계량분석

## 다변량 회귀분석 (II)

김현우, PhD<sup>1</sup>

<sup>1</sup>충북대학교 사회학과 조교수

November 8, 2021

# 진행 순서

- 1 다변량 회귀분석의 실제 활용
- 2 다변량 회귀분석 연습

## 다변량 회귀분석의 실제 활용

## 다변량 회귀분석의 실제 활용

## 일-생활 균형 인식은 임금노동자의 행복에 어떤 영향을 미치는가?

- 윤수인 · 이홍직 (2020)의 모든 통계표를 순서대로 살펴보자. 어떤 식으로 진행되고 있는가?
- 주요 종속변수 및 독립변수는 어떤 항목들로 측정되고 있는가? 기술통계표는 어떻게 제시되고 있는가? 상관관계분석의 결과는 어떻게 요약되었는가?
- t 검정과 일원분산분석 결과는 어떻게 제시되었는가?
- 마지막으로 <표 8>의 다변량 회귀분석 결과표는 어떤 특징을 가지고 있나? 일-생활 균형 인식의 회귀계수를 해석하시오.

윤수인 · 이홍직. 2020. “임금노동자의 행복에 영향을 미치는 요인에 관한 연구: 인구 사회학적 요인, 좋은 일자리 요인, 일-생활 균형 요인을 중심으로.” *한국사회복지학* 72(2): 373-400.

## 다변량 회귀분석의 실제 활용

집합적 효능은 우울증상과 주관적 건강평가를 설명하는 다변량 회귀모형을 살펴보자.

- 우울증상에 관한 집합적 효능의 회귀계수를 해석하시오.
- 주관적 건강평가에 관한 집합적 효능의 회귀계수를 해석하시오.
- 우울증상에 관한 개인소득의 회귀계수가 집합적 효능을 추가하면서 어떻게 변화하는지 해석하시오.
- 주관적 건강평가에 관한 연령의 회귀계수가 집합적 효능을 추가하면서 어떻게 변화하는지 해석하시오.
- 참고로 <표 2>와 <표 3>에서는 서로 독립변수와 종속변수를 뒤바꾸어 모형을 추정하고 있다. 이것은 다소 일반적이지 않은 관행으로 보이니 주의할 것.

임훈민 · 김중백. 2014. “집합적 효능과 건강의 상호관계에 관한 연구.” 보건사회연구 34(4): 441-446.

## 다변량 회귀분석 연습

## 다변량 회귀분석 연습

## 회귀분석 연습(I)

- 다음의 데이터 파일을 여시오: ATTEND.DTA (N=680, cross-sectional individual data on classes attended).
- 종속변수는 termgpa이고, 독립변수는 termgpa, atndrte, hwrte, priGPA, final인 회귀모형을 구상하시오.
- 모형에 사용되는 모든 변수를 간단히 점검하고 의미를 파악하시오. 결측치는 listwise deletion으로 제거하시오.
- 변수를 추가하는 순서에 관하여 나름의 스토리텔링을 구상하면서 위계적 회귀모형을 구축하시오. 결과표를 제시하되 반드시 사례수(number of observations)와 (adj.)  $R^2$ 를 함께 제시하시오.
- 최종 모형에서 볼 때 무엇이 가장 강력한 변수인가? 그 회귀계수를 해석하시오. 무엇에 근거하여 그렇게 판단하였는가? 왜 그것이 가장 강력한 변수인지 사회학적인 설명을 제시하시오.

## 회귀분석 연습(II)

- 다음의 데이터 파일을 여시오: WAGE2.DTA (N=935, cross-sectional data on wages).
- 종속변수는 lwage이고, 독립변수는 meduc, feduc, brthord, IQ, KWW, educ, exper, tenure, age인 회귀모형을 구상하시오.
- 모형에 사용되는 모든 변수를 간단히 점검하고 의미를 파악하시오. 결측치는 listwise deletion으로 제거하시오.
- 변수를 추가하는 순서에 관하여 나름의 스토리텔링을 구상하면서 위계적 회귀모형을 구축하시오. 결과표를 제시하되 반드시 사례수와 (adj.)  $R^2$ 를 함께 제시하시오.
- 모형을 urban==1과 urban==0로 나누어 각각 추정하고 결과표에서 대조하시오.
- 위의 두 모형에서 무엇이 각각 가장 강력한 변수인가? 그 회귀계수를 해석하시오. 왜 그것이 가장 강력한 변수인지 사회학적인 설명을 간략히 제시하시오.



## 회귀분석 연습(III)

- 다음의 데이터 파일을 여시오: LOWBRTH.DTA (N=1989, cross-sectional individual data).
- 데이터 구조를 먼저 꼼꼼히 살펴보고 무엇이 예전과 다른가 확인하시오.
- 종속변수는 lwage이고, 독립변수는 meduc, feduc, brthord, IQ, KWW, educ, exper, tenure, age인 회귀모형을 구상하시오.
- 모형에 사용되는 모든 변수를 간단히 점검하고 의미를 파악하시오. 결측치는 listwise deletion으로 제거하시오.
- 변수를 추가하는 순서에 관하여 나름의 스토리텔링을 구상하면서 위계적 회귀모형을 구축하시오. 결과표를 제시하되 반드시 사례수와 (adj.)  $R^2$ 를 함께 제시하시오.
- 모형을 year==1987와 year==1990로 나누어 각각 추정하고 결과표에서 대조하시오.
- 위의 두 모형에서 무엇이 각각 가장 강력한 변수인가? 그 회귀계수를 해석하시오. 왜 그것이 가장 강력한 변수인지 사회학적인 설명을 간략히 제시하시오.

## 다변량 회귀분석 연습

다변량 회귀분석은 석사 초년생 레벨에서 사회통계학의 허리에 해당한다.

- 앞으로 무거운 짐을 더 높게 쌓아올릴 수 있기 위해 허리를 튼튼히 단련해야 한다.
- 응용통계학의 일종인 사회통계학 측면에서 보면 기초적인 이해를 마친 다음 가장 먼저 필요한 것은 반복 숙련이다. 많은 데이터를 가지고 계속해서 다변량 회귀분석을 연습하고 감을 얻어야 한다.
- 조금씩 더 숙달되면 그 안에 숨겨져 있는 수학적 원리를 공부하여 더 체계적으로 이해하고 향후 수준높은 공부를 준비한다.
- 위 3개의 연습을 모두 마치고 다음 주까지 숙제로 제시하시오(코드 첨부 필수!).
- (숙제는 아니지만) 반드시 자신만의 데이터셋을 가지고 무엇이 적절한 종속변수이고 독립변수인지 판단하여 연습하시오. 절대적으로 시간을 많이 들여야 충분한 감각을 체득할 수 있다. 일단 종속변수와 독립변수 모두가 숫자형 척도인 경우를 중심으로 연습하시오.