

범주형 자료분석
CATEGORICAL DATA ANALYSIS
(3-3-0)

2025년 1학기

Instructor	김현우, PhD (hxk271@cbnu.ac.kr)
Office	N15동 414호
Classroom	N15동 407호(사회조사실습실)
Class hours	01:00 PM – 4:50 PM, 금요일
Course website	https://github.com/hxk271/CatData

강의 개요

사람들의 투표 여부, 최종학력 수준, 주된 교통 수단 등에 관한 여러 설문조사 자료, 실험 자료, 행정 자료에서 흔히 관심변수는 범주형으로 주어집니다. 이는 어떤 질병의 치료를 위한 약의 종류나 치료 결과의 판정 등 다양한 보건의학 연구에서도 마찬가지입니다. 범주형 변수의 분석은 연구 결과의 정확성과 타당성을 확보하는 데 필수적인 지식의 일부라고 할 수 있습니다.

이 수업은 이와 같은 범주형 자료 분석에 대한 이론과 실습을 통합적으로 다루는 대학원 수준의 강의로, 사회학을 비롯한 사회과학 분야의 연구자들이 데이터를 효과적으로 분석할 수 있는 전문 지식을 제공하는 데 초점을 맞추고 있습니다. 이 수업에서는 (범주형 자료의 기초 개념부터 고급 분석 방법론까지 포괄하는) 이론 강의와 실제 데이터 분석 실습으로 구성되며, Stata를 활용하여 실질적인 데이터 분석 능력을 배양합니다.

주요 학습 목표는 다음과 같다.

1. 로짓/프로빗 모형 입문
2. 서열척도 및 명목척도의 분석
3. 범주형 자료분석의 확장

이 수업은 실증적인 연구를 수행하려는 사회과학자, 정책 연구자, 데이터 사이언티스트에게 복잡한 연구 질문에 적합한 분석 방법을 선택하고 실행하는 데 필요한 통찰과 기술을 제공하고자 합니다.

선수 과목

학부에서 하나 이상의 사회통계 관련 과목을 반드시 이수해야 합니다. 대학원에서 계량분석 (8969001) 또는 그에 준하는 수업을 반드시 이수하여야 합니다.

강의 교재

이 수업에서 우리는 다음과 같은 교재를 사용합니다.

- Long, J. Scott and Jeremy Freese. 2014. *Regression Models for Categorical Dependent Variables Using Stata (3rd Edition)*. College Station, TX: Stata Press. (필수)
- 홍세희. 2005. 『이항 및 다항 로지스틱 회귀분석』. 교육과학사. (필수)
- 고길곤. 2018. 『범주형 자료 분석』. 문우사. (참고)
- Agresti, Alan. 2020. 『범주형 자료분석 개론(제3판)』. 자유아카데미. (참고)
- 장혜정·김현우. 2023. 『경영경제 통계학 노트』. 한빛아카데미. (참고)

학습 보조자료

- YouTube 또는 Google에서 궁금한 주제를 Stata와 연관검색어로 하여 찾아보면 엄청나게 많은 자료가 있습니다. 궁금한 점이 있으면 망설이지 말고 즉시 영어로 검색해 보기를 추천합니다. 초보자가 가질만한 거의 모든 질문은 이미 누군가가 던졌고 게다가 대답도 이미 있을 가능성이 높습니다!
- Statalist (<https://www.statalist.org/>)는 Stata를 사용하는 (저 뿐만 아니라) 여러분 모두에게 든든한 우군입니다. 다른 곳에서 아무리 찾아봐도 적절한 답변을 찾을수 없었다면 여기서 직접 영어로 질문할 수 있습니다.

강의 구성

- 강의 및 실습: 모든 학생은 반드시 수업에 참여해야 합니다. 수업을 통해 기초적인 개념과 방법을 배우고 Stata를 사용하여 실습합니다. 수업 내용은 진행될수록 누적되기 때문에 결석 혹은 지각은 향후 이해에 큰 방해가 됩니다. 강의 중에 실습이 이루어집니다. 전산실에서 Stata를 사용하여 실제 분석을 진행합니다.
- 퀴즈: 거의 매주 아주 조금씩 과제(take-home quiz)가 주어집니다. 과제는 여러분을 평가하고 심사하려는데 목적이 있다가 보다, 수업에서 배운 내용을 복습하는데 도움을 주려는 목적이 더욱 큽니다. 잘 모르겠다면 다른 학생들과 토론을 해도 되지만 최대한 혼자 시도하기를 추천합니다. 다음 주 수업 시작과 함께 즉각 제출해야 하며 (수업이 끝나고 제출하는 등) 데드라인을 넘기면 감점됩니다.
- 읽을거리: 교수자가 읽을거리를 부여한 주에는 반드시 이를 읽고 수업에 참여해야 합니다. 또한 해당 주에 다루는 분석 기법이 사용된 논문 1편을 각자 선택하여 읽고, 분석 기법에 관해 여러분이 얻은 통찰을 공유합니다. 반드시 자신의 연구 주제와 부합하는 분야 중 SSCI 또는 등재지 수준의 학술 논문을 선택해야 합니다. 필수 읽을거리 이외의 다른 자료도 종종 업로드되며 이것을 읽는 것은 여러분의 선택 사항입니다.
- 중간시험 및 기말시험: 중간시험과 기말시험에서 스스로 고른 자료를 직접 분석한 리포트를 제출합니다.

학점 안내사항

최종 학점은 다음 기준에 따라 산출됩니다.

- 출석 및 수업 참여 (20%)
- 과제 (20%)
- 시험 2개 (60%)

수강생 유의사항

- 모든 수업은 별도의 안내가 없는 이상 원칙적으로 **대면**으로 진행됩니다.
- 수업 전일에 해당 주차 강의안과 자료가 다음 GitHub 레파지토리에 업로드됩니다:
<https://github.com/hxk271/CatData>
- 공결은 증빙서류를 모두 갖추어 담당교수에게 **직접** 제출된 경우만 인정하며, 개별적인 공결 행정처리는 일체 **무효**임에 주의하십시오.
- 2회 지각은 1회 결석으로 처리합니다. (**첫 수업 및 공결**을 포함하여) 3회 이상 결석한 경우 무조건 F이며 예외는 없습니다. 불가피한 사정으로 공결하였을 경우 그 외 일체 지각이나 결석을 하지 않아야 하겠습니까.
- 평균적인 학생의 수준에서 이 수업은 상당히 부담이 될 수 있습니다. 학기 중 다른 여러 과목을 함께 이수해야 한다면 이 수업은 선택하지 않는 편이 바람직합니다.
- 시험 관련 부정행위자로 판명되었을 때는 학칙 또는 내규에 의거 해당 교과목의 성적을 취소합니다.

장애학생 수업안내

장애학생은 본 수업과 관련하여 본인 희망시 다음과 같은 지원이 가능합니다. 담당교수 및 장애학생지원센터와 상담 바랍니다.

- 공통: 도우미 지원(수업, 이동), 대체평가, 별도 발표/시험장소 제공, 선수강 지원, 노트북 사용
- 시각장애: 점자/확대/녹음 교재 및 시험지 제공, 발표/시험시간 연장, 강의자료 텍스트제공
- 청각장애: 지정좌석제, 동영상 자막지원
- 지체장애: 강의실 변경, 지정좌석제, 발표/시험시간 연장

토픽 개요

로짓/프로빗 모형 입문	1주차	Introduction
	2주차	Ordinary Least Squares Revisted
	3주차	Basics in Logit and Probit Models
	4주차	Applications in Logit and Probit Models
	5주차	Cautionary Notes about Nonlinear Models
목표: 로짓과 프로빗 모형의 기초를 이해하고 관련 문헌을 학습한다		
서열척도 및 명목척도의 분석	6주차	Models for Ordinal Outcomes I
	7주차	Models for Ordinal Outcomes II
	8주차	중간시험
	9주차	Models for Nominal Outcomes I
	10주차	Models for Nominal Outcomes II
목표: 종속변수가 서열 혹은 명목척도일 때 자료를 분석할 수 있다		
범주형 자료분석의 확장	11주차	Count Models
	12주차	Models for Truncated or Censored Outcomes
	13주차	Two-Stage Models
	14주차	Repeated Measures
	15주차	기말시험
목표: 좀 더 다양한 형태의 범주형 자료분석을 학습한다		

세부 일정

1주차 Course Introduction

GOAL 이 수업의 핵심 목적과 내용을 간단히 살펴본다.

READINGS 없음

2주차 Ordinary Least Squares Revisted

GOAL 보통최소제곱(OLS)의 개념과 계산 원리를 이해하고 실제로 사용할 수 있다.

READINGS 장해정·김현우 제12장 및 제13장; 김성언 (2018); Ashraf and Galor (2013)

3주차 Basics in Logit and Probit Models

GOAL 로지스틱 회귀모형과 프로빗 회귀모형의 개념과 계산 원리를 이해한다.

READINGS Long and Freese 제5장 및 제6장; 홍세희 제1장-제7장 및 제10장; Uberti (2022)

4주차 Applications in Logit and Probit Models

GOAL 로지스틱 회귀모형과 프로빗 회귀모형을 실제로 분석에 사용할 수 있다.

READINGS Long and Freese 제5장 및 제6장; 박상현·진영재 (2017); 홍힘찬·나태준 (2020); Abrutyn and Mueller (2014); Warner and McCarthy (2014)

5주차 Cautionary Notes about Nonlinear Models

GOAL 비선형모형(nonlinear model)에서 계수 비교(coefficient comparisons)와 상호작용항(interaction terms) 사용에 수반되는 문제를 파악한다.

READINGS Ai and Norton (2003); Karaca-Mandic, Norton and Dowd (2011); Kuha (2020); Mitchell and Chen (2005); Mood (2010)

세부 일정 (계속)

6주차 Models for Ordinal Outcomes I

GOAL 서열변수에 대한 로지스틱 회귀분석에 관한 개념과 계산 원리를 이해한다.

READINGS Long and Freese 제7장; 양종민·김도균 (2022); Abromaviciute, Seebruck and Edwards (2019)

7주차 Models for Ordinal Outcomes II

GOAL 서열변수에 대한 로지스틱 회귀분석을 실제로 수행하고, 보다 일반화된 모형을 이해한다.

READINGS Long and Freese 제7장; Fullerton (2009); Mouw (2001); Williams (2010, 2016)

8주차 중간시험

9주차 Models for Nominal Outcomes I

GOAL 명목변수에 대한 로지스틱 회귀분석에 관한 개념과 계산 원리를 이해한다.

READINGS Long and Freese 제8장; 홍세희 제8장-제9장; 김재현 (2023); Paolino (2021)

10주차 Models for Nominal Outcomes II

GOAL 명목변수에 대한 로지스틱 회귀분석을 실제로 수행할 수 있다. 조건부 로지스틱 회귀분석을 수행할 수 있다.

READINGS Long and Freese 제8장; 장덕진 (2005); Dow and Endersby (2004); 조성대 (2003)

세부 일정 (계속)

11주차 Count Models

GOAL 푸아송(Poisson) 회귀모형과 음이항(negative binomial) 회귀모형을 이해하고 활용할 수 있다.

READINGS Long and Freese 제9장; 최열·임하경 (2005); Cornwell, Laumann and Schumm (2008); Agnone (2007); Wo (2024)

12주차 Models for Truncated or Censored Outcomes

GOAL 절단 혹은 검열된 종속변수가 있는 특수한 경우에 대응할 수 있다.

READINGS 이동영 (2019); Stata 메뉴얼(intreg)

13주차 Two-Stage Models

GOAL 2단계 회귀모형을 이해하고 적절한 상황에 이를 활용할 수 있다.

READINGS

14주차 Repeated Measures

GOAL 반복된 측정이 이루어진 데이터에 대응한 범주형 자료분석을 이해한다.

READINGS

15주차 기말시험