

계량분석
LINEAR REGRESSION ANALYSIS AND ITS APPLICATIONS
전공필수(3-3-0)

2021년 2학기

Instructor	김현우 (hxk271@cbnu.ac.kr)
Office Address	N15동 414호
Office Hours	메일로 연락후 면담
Classroom	N15동 427호
Class Hours	6:00 PM-9:00 PM, 월요일

강의 개요

이 수업에서는 이론과 데이터를 계량적으로 연결하는 사회연구의 절차와 방법을 배웁니다. 교차분석(cross-tabulation analysis), 선형회귀분석(linear regression analysis), 선형회귀 가정의 진단(regression assumption diagnostics), 결측치 문제와 대응(missing data), 주성분 분석(principal component analysis) 및 척도구성(scale construction) 등 여러 도구를 폭넓게 공부합니다.

이 수업의 주안점은 사회학적으로 의미있는 발견(substantive findings)을 이끌어내기 위해 필요한 절차를 배우고 연습하는 것에 있습니다. 여러분은 자신의 연구에 활용된 통계적 기법에 관해 타인에게 말과 글로써 어떻게 설명할 수 있는가를 함께 배우게 됩니다. 설령 사회통계학을 심도있게 전공하지 않더라도 다른 동료 사회학자의 계량적 논문을 읽고 비판하기 위해서도 꼭 필요한 수업입니다.

이 수업은 사회학과 대학원생(석사 및 박사)을 기준으로 설계되었으며 학부 과정에서 적어도 하나 이상의 학부 사회통계학 수업을 이수하였을 것으로 전제합니다. 통계학은 본질적으로 수학의 영역입니다. 만일 사회통계학을 보다 수준높게 공부하기로 마음 먹었다면 수학 공부는 결코 피할 수 없습니다. 그러나 모든 이들이 사회통계학 자체를 전문영역으로 삼을 필요는 없으므로 수학을 억지로 공부할 필요도 없습니다. 최소한의 수학 지식(고등학교 문과 수준)과 기초적인 컴퓨터 활용 능력만큼은 전제합니다.

이 수업은 프로젝트기반 학습(Project-Based Learning)을 지향합니다. 여러분은 스스로 사회학적 연구 주제를 개발해야만 하며, 이 수업에서 배운 도구와 절차를 활용하여 연구 주제를 발전시켜 나가야 합니다. 우리는 이 수업에서 사회과학 분야에서 국내외로 폭넓게 사용되는 Stata라는 통계분석 패키지를 사용합니다.

선수 과목

학부 과정에서 사회통계학 또는 그에 준하는 기초통계학 수업의 이수를 추천합니다. 전혀 기초통계학에 대한 이해가 없다면 면담이 필요합니다.

강의 교재

유인물을 배부하므로 필수 교재는 없습니다. 아래는 모두 참고용이므로 반드시 구매할 필요는 없습니다. Allison (1999)은 매우 우수한 입문자용 교과서입니다. 김태근(2006)의 교과서도 매우 쉽게 쓰여졌으므로 자습에 큰 도움이 됩니다. 김두섭·강남준(2008)의 교과서는 입문자에게는 다소 버거울수 있지만 회귀분석의 이해를 깊이하는데 큰 도움이 됩니다. 민인식·최필선(2020)의 교과서는 이론적 설명보다는 Stata를 활용하는 방법을 배우는데 유용합니다. Jaggia의 것은 한 권으로 통계학 기초에서 데이터과학의 주요토픽까지 모두 커버하며 Excel과 R을 사용하는 좋은 책입니다. 만일 한 권만 구입한다면 Allison (1999)을 추천합니다. 만일 사회통계학을 심도있게 공부하고 싶다면 모두 구매를 추천합니다.

- Allison, Paul D. 1999. *Multiple Regression: A Primer*. Thousand Oaks, CA: Pine Forge Press. (참고)
- 김태근. 2006. *U-CAN 회귀분석*. 인간과복지. (참고)
- 김두섭·강남준. 2008. *회귀분석: 기초와 응용*. 나남. (참고)
- 민인식·최필선. 2020. *STATA 기초통계와 회귀분석*. 지필. (참고)
- Jaggia, Sanjiv, Alison Kelly, Kevin Lertwachara, Leida Chen. 2021. 『비즈니스 애널리틱스』. 한빛아카데미. 그린. (참고)

학습 보조자료

- 강의안과 퀴즈는 모두 수업 하루 전에 깃허브에 업로드 됩니다. 갱신이 되는 경우가 있으니 종종 확인 부탁드립니다: <https://github.com/hxk271/LinearRegression/>
- 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 있다기 보다, 수업에서 배운 내용을 복습하는데 도움을 주려는 목적이 더욱 큽니다. 잘 모르겠다면 다른 학생들과 토론을 해도 되지만 최대한 혼자 시도해보기를 추천합니다. 다음 주 수업 시작과 함께 즉각 제출해야 하며 (수업이 끝나고 제출하는 등) 데드라인을 넘기면 감점됩니다.
- 중간시험: 선형회귀분석이 활용된 기존문헌을 비판적으로 리뷰하고 이를 발표합니다. 여러분의 논문에도 활용될 수 있는 형태로 요약한 내용을 함께 제출합니다.
- 기말시험: 학생 여러분의 프로젝트로 수행한 선형회귀분석 결과를 요약하고 발표합니다. 명확하게 제시된 가설, 데이터 소개, 변수 리코딩, 모형 설정, 가정 확인; 결과 해석 및 시각화를 종합적으로 심사하고 평가합니다.

학점 안내사항

모든 과제와 중간 및 기말발표 점수는 종합정보시스템에 업로드됩니다. 최종 학점은 다음 기준에 따라 산출됩니다.

- | | |
|--------------------|--------------|
| • 출석 및 수업 참여 (20%) | • 중간발표 (20%) |
| • 퀴즈 (20%) | • 기말발표 (40%) |

수강생 유의사항

- 모든 수업은 별도의 안내가 없는 이상 원칙적으로 **대면**으로 진행됩니다.
- 전체 수업의 4분의 1이상을 결석한 경우 시험에 응시할 수 없습니다.
- 시험관련 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 장애지원센터와 언제든지 상담 바랍니다.

- 학습지원 : 강의 파일 제공, 대필 도우미 및 속기 지원 허락, 강의 녹음 허락, 과제 제출 기간 연장(시각, 손사용 불편 학생), 보조기구 사용 가능 등
- 평가지원: 영어교과 듣기 시험 대체(청각장애학생), 장애종류 및 정도에 따라 시험 시간 1.5배 1.7배 연장, 별도 시험장소 및 시험지 제공, 필요한 경우 학습기자재 사용을 허용

고마우신 분들

종종 잘 알려져 있지 않지만 교육은 사실 집단적 과업의 결과물입니다. “아이 하나를 키우려면 마을 하나가 필요하다(It takes a village to raise a child)”라는 격언처럼 대학원 과목을 하나를 만드는데도 집단적인 노력이 필요합니다. 이 수업의 많은 부분은 제가 스스로 만든 것이 아닙니다. 여기에는 David Johnson (Penn State), Jennie Brand (UCLA), Molly Martin (Penn State), Marieke Van Willigen (Georgia Southern), Don Bradley (Samford), 민인식 (경희대), Richard Williams (Notre Dame) 등 많은 분들의 강의자료를 활용하였습니다. 도움을 주신 모든 분들께 감사드립니다. 여러분의 피드백은 다음 이 수업을 더욱 발전시키는데 큰 도움이 됩니다.

토픽 개요

토픽1: 교차분석에서 데이터의 이해까지	1주차	Course Introduction/Review
	2주차	Theory-Based Data Analysis
	3주차	Cross-Tabulations and Tables
	4주차	Means/Proportions Comparison and ANOVA
	5주차	Examining Data

목표: 교차분석과 그래프 및 데이터 구조를 이해한다.

토픽2: 단순회귀분석	6주차	Correlations and OLS Regression
	7주차	Multivariate OLS Regression
	8주차	Midterm
	9주차	Dummy Variables and Conditional Relationships
	10주차	Non-linear Tests and Variable Transformations

목표: 단순회귀분석에 필요한 절차와 방법에 익숙해진다.

토픽3: 회귀 가정과 그 위배	11주차	Regression Diagnostics
	12주차	Principal Component Analysis and Scale Construction
	13주차	Missing Data
	14주차	Sample Design, Data Clustering, and Weighting
	15주차	Final

목표: 회귀분석의 가정과 위배시 적절한 대응책을 판단할 수 있다.

세부 일정

1주차 Course Introduction/Review

TOPICS 사회학과 계량분석; 통계분석 패키지의 선택; 기술통계, 확률론, 이산확률분포와 연속확률분포, 표집, 통계추정, 가설검정, 평균비교, 분산비교, 카이제곱분석, 상관 분석, 회귀분석; Stata의 기초 명령어들

GOAL 이 수업에 대한 이해를 얻고 기초적인 사회통계학의 개념들을 간단히 정리한다. Stata를 처음으로 구동하고 기초적인 명령어들을 학습한다.

2주차 Theory-Based Data Analysis

TOPICS Theory; Finding; Hypothesis; Correlation and Causation; Falsifiability; Replicability; Theoretical Construct; Measurement Error; Validity and Reliability; Population and Sample; Independent and Dependent Variables; Cross-Sectional Analysis; Unit of Analysis; Confounding Variables

GOAL 이론적 모형을 경험적 모형으로 옮기는데 필요한 개념적 기초를 학습한다

3주차 Cross-Tabulations and Tables

TOPICS Crosstab; Chi-square Analysis

GOAL 교차표(cross-tabulations)와 그 논리 및 해석 방법을 이해한다; 카이제곱분석을 통해 두 변수간 연관성을 통계적으로 파악한다

4주차 Means/Proportions Comparison and ANOVA

TOPICS t -test; F test; Data Visualization

GOAL 카이제곱분석을 통해 두 변수간 연관성을 통계적으로 파악한다.

5주차 Examining Data

TOPICS Raw Data; Secondary Data; Data Sources; Aggregation; Questionnaire and Codebook; Histogram

GOAL 스프레드시트와 통계적 컴퓨팅의 기초를 살펴본다; 데이터의 기본 구조를 이해하고 분석에 필요한 형태를 명확하게 사고한다

세부 일정 (계속)

6주차 Correlations and OLS Regression

TOPICS Correlation and Causation; Regress to the Mean; Linear Model Specification; Predicted \hat{Y} ; Best-Fitting Straight Line; Least Squared Error; Standard Error; Standardized Coefficients; Goodness-of-fit

GOAL 상관계수의 계산과 해석 방법을 이해한다; OLS의 기본적인 논리를 이해한다; OLS의 한계를 이해한다

7주차 Multivariate OLS Regression

TOPICS Control Variables; Multicollinearity; Variance Inflation Factor; Path Analysis

GOAL 두 개 이상의 독립변수에 대한 회귀분석과 “Ceteris Paribus”를 이해한다; 경로분석의 아이디어를 이해한다

8주차 Midterm

9주차 Dummy Variables and Conditional Relationships

TOPICS Categorical Variables; Dummy Variables; Interaction Terms; Graph Drawing

GOAL 범주형 및 더미 변수가 요구되는 상황을 이해한다; 상호작용항이 포함된 회귀분석을 수행한다

10주차 Non-linear Tests and Variable Transformations

TOPICS Nonlinear Model Specification; Polynomial Model; Log Transformation; Wage Equation

GOAL 통계적 관계가 비선형적인가를 판단한다; 다항식이 포함된 모형을 실습하고 해석한다

세부 일정 (계속)

11주차	<u>Regression Diagnostics</u>
TOPICS	Homoscedasticity; Heteroscedasticity; Outliers; Leverage; Influential Cases; Cook's Distance
GOAL	회귀 진단을 위한 대책을 이해하고 주어진 상황에 적합한지를 판단하여 적절한 대응책을 구사할 수 있다
12주차	<u>Principal Component Analysis and Scale Construction</u>
TOPICS	Dimension Reduction; Principal Components; Eigenvalues/Eigenvector; Cronbach's Alpha
GOAL	주성분 분석의 기본적인 아이디어를 이해한다; multiple-item scales을 구성한다
13주차	<u>Missing Data</u>
TOPICS	Unobservables; Missing Completely at Random (MCAR); Missing at Random (MAR); Missing Not at Random (MNAR); Deletion; Imputation; Multiple Imputations by Chained Equations
GOAL	결측치가 내포한 문제를 이해한다; 결측치가 있는 경우 어떠한 대책이 있는가를 간략히 이해한다.
14주차	<u>Sample Design, Data Clustering, and Weighting</u>
TOPICS	Multi-stage Area Probability Sample; Stratified Random Sampling; Design Effects; Complex Data; Sampling Weights
GOAL	상이한 샘플링 디자인을 살펴본다; 샘플 가중치가 무엇이고 어떻게 사용하는지 간략히 살펴본다; 클러스터링을 동반한 샘플링 디자인의 효과를 이해한다
15주차	<u>Final</u>