

계량분석
QUANTITATIVE ANALYSIS FOR SOCIAL SCIENTISTS
전공필수(3-3-0)

2023년 2학기

Instructor	김현우 (hxk271@cbnu.ac.kr)
Office Address	N15동 414호
Office Hours	메일로 연락후 면담
Classroom	N15동 410호(사회조사분석실)
Class Hours	01:00 PM-3:50 PM, 월요일
Course website	https://github.com/hxk271/QASS

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 발견을 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 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 선수강이 요구됩니다.

강의 교재

관심은 통계학 기초 교과서가 이미 서점에 많습니다. 어떤 것을 선택해도 좋지만 그 중 특히 아래 책을 추천합니다. 기초 교과서는 반드시 수업 진도에 맞추어 함께 공부해야 합니다.

- 장혜정·김현우. 2023. 『경영경제 통계학 노트』. 한빛아카데미. **(필수)**
- 와쿠이 요시유키 · 와쿠이 사다미. 2021. 『누워서 읽는 통계학: 누구나 쉽게 시작하고 블랙홀처럼 빠져드는 통계학 이야기』. 한빛아카데미. **(참고)**

회귀분석의 이론적 내용은 아래는 참고서를 통해 공부합니다. 김태근(2006)도 매우 쉽게 쓰여졌으므로 자습에 큰 도움이 됩니다. 김두섭·강남준(2008)은 입문자에게는 다소 버거울수 있지만 회귀분석의 이해를 얻는데 큰 도움이 됩니다. Allison (1999)은 매우 우수한 입문자용 교과서입니다. Gujarati (2009)는 기초부터 중급 이상의 수준까지 계량경제학의 기법들을 굉장히 꼼꼼하게 다루고 있습니다.

- 김태근. 2006. 『U-CAN 회귀분석』. 인간과복지. **(참고)**
- 김두섭 ·강남준. 2008. 『회귀분석: 기초와 응용 (개정판)』. 나남. **(참고)**
- Allison, Paul D. 1999. *Multiple Regression: A Primer*. Thousand Oaks, CA: Pine Forge Press. **(참고)**
- Gujarati, Damodar N. 2009. 『Gujarati의 계량경제학 (제5판)』. 지필미디어. **(참고)**

Stata 자습은 아래는 참고서를 통해 공부합니다. 민인식·최필선(2020)은 이론적 설명보다는 Stata를 활용하는 방법을 배우는데 유용합니다.

- 민인식 · 최필선. 2020. 『STATA 기초통계와 회귀분석 (제2판)』. 지필. **(필수)**

학습 보조자료

- 강의안과 퀴즈는 모두 수업 하루 전에 깃허브에 업로드 됩니다. 갱신이 되는 경우가 있으니 종종 확인 부탁드립니다: <https://github.com/hxk271/QASS/>
- 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 큼니다. 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.
- 수업 전일에 해당 주차 강의안이 다음 GitHub 레파지토리에 업로드됩니다:
<https://github.com/hxk271/QASS>
- 공결은 증빙서류를 모두 갖추어 담당교수에게 **직접** 제출된 경우만 인정하며, 개별적인 공결 행정처리는 일체 무효임에 주의하십시오.
- 2회 지각은 1회 결석으로 처리합니다. (첫 수업 및 공결을 포함하여) 3회 이상 결석한 경우 무조건 F이며 예외는 없습니다. 불가피한 사정으로 공결하였을 경우 그 외 일체 지각이나 결석을 하지 않아야 하겠습니다.
- 과제를 4회 이상 제출하지 않으면 무조건 F이며 예외는 없습니다. 과제는 선택사항이 아닙니다.
- 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 시간 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
	2주차	Descriptive Statistics and Hypothesis Tests
	3주차	Examining Data
	4주차	Cross-Tabulations and Tables
	5주차	Means/Proportions Comparison and ANOVA
목표: 교차분석과 그래프 및 데이터 구조를 이해한다.		
토픽2: 단순회귀분석	6주차	Correlations and Simple Regression
	7주차	Multiple Regression
	8주차	Midterm
	9주차	Dummy Variables and Conditional Relationships
	10주차	Regression Diagnostics
목표: 단순회귀분석에 필요한 절차와 방법에 익숙해진다.		
토픽3: 회귀 가정과 그 위배	11주차	Non-linear Tests and Variable Transformations
	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

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

READINGS 장혜정·김현우 1장, 2장

2주차 Descriptive Statistics and Hypothesis Tests

TOPICS Mean; Median; Mode; Standard Deviation; Variance; Range; Confidence Interval; Null and Alternative Hypothesis; Significance Test

READINGS 장혜정·김현우 3장, 5장, 6장, 7장, 8장; 민인식·최필선 1장, 2장

3주차 Examining Data

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

4주차 Cross-Tabulations and Tables

TOPICS Crosstab; Chi-square Analysis

READINGS 장혜정·김현우 4장, 11장; 민인식·최필선 6장, 7장

5주차 Means/Proportions Comparison and ANOVA

TOPICS Degree of Freedom; t -test; ANOVA; Data Visualization

READINGS 장혜정·김현우 9장, 10장; 민인식·최필선 2장, 3장

세부 일정 (계속)

6주차 Correlations and Simple Regression

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

READINGS 장혜정·김현우 12장; 민인식·최필선 4장, 8장

7주차 Multiple Regression

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

READINGS 장혜정·김현우 13장; 민인식·최필선 8장

8주차 Midterm

9주차 Dummy Variables and Conditional Relationships

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

READINGS 장혜정·김현우 13장; 민인식·최필선 9장

10주차 Regression Diagnostics

TOPICS Homoscedasticity; Heteroscedasticity; Outliers; Leverage; Influential Cases; Cook's Distance

READINGS 민인식·최필선 10장, 11장

세부 일정 (계속)

11주차	<u>Non-linear Tests and Variable Transformations</u>
TOPICS	Nonlinear Model Specification; Polynomial Model; Log Transformation; Wage Equation
READINGS	민인식·최필선 15장
12주차	<u>Principal Component Analysis and Scale Construction</u>
TOPICS	Dimension Reduction; Principal Components; Eigenvalues/Eigenvector; Cronbach's Alpha
READINGS	민인식·최필선 5장
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
14주차	<u>Sample Design, Data Clustering, and Weighting</u>
TOPICS	Multi-stage Area Probability Sample; Stratified Random Sampling; Design Effects; Complex Data; Sampling Weights
15주차	<u>Final</u>