

사회통계
INTRODUCTION TO SOCIAL STATISTICS
전공필수(3-3-0)

2022년 1학기

Instructor	김현우 (hxk271@cbnu.ac.kr)
Office Address	N15동 414호
Office Hours	메일로 연락후 면담
Classroom	N15동 427호
Class Hours	01:00 PM-03:50 PM, 금요일

강의 개요

이 수업에서는 인구, 조직, 정치, 경제, 여론, 미디어, 종교, 교육, 금융, 환경 등을 아우르는 다양한 분야의 현실사회 데이터(공공개방 데이터 포함)를 계량적으로 분석하는 기초적인 절차와 방법에 관해 배웁니다. 여러분이 중고등학교 시절에 배운 수학 과목을 기초로 하여, 확률분포와 가설검정의 논리를 직접 실험해보고, 자료 리코딩/변수구축, 평균비교와 분산비교, 상관분석과 회귀분석 등 기초적인 통계학의 도구를 반복 숙달하게 됩니다.

이 수업은 여러분의 관련 자격증 취득과도 직접적으로 연관되어 있습니다. 이 수업을 포함하여 다음 학기 **사회통계연습**까지 학습하면 여러분은 컴퓨터활용능력 2급의 스프레드시트 및 실기 과목과 사회조사분석사 2급의 사회통계 과목에 해당하는 내용을 모두 배우게 됩니다.

이 수업은 사회학과 학부과정 2학년생을 기준으로 설계되었으며 그동안 어떠한 사회통계 과목도 이수하지 않았을 것을 전제로 합니다. 타 학과에서 이미 수학과 통계학 분야 과목을 충분히 이수한 학생을 위한 수업이 아니므로 해당 학생들은 수강 신청에 신중을 기하여야 합니다(최종 점수에 일정한 패널티가 부여됩니다).

여러분은 매일매일 빠지지 않고 수업에 참석해야 하고 수업 후에도 복습을 거듭해야 합니다. 이 수업에서 사용하는 통계분석 패키지는 마이크로소프트 엑셀(Microsoft Excel)입니다. 여러분은 충북대학교가 구입한 마이크로소프트 오피스 라이선스를 무료로 사용할 수 있습니다. 엑셀은 SPSS나 Stata, SAS, R, Python과 같은 전문적인 분석도구는 아니지만 현업에서 폭넓게 사용되는 툴입니다.

선수 과목

없음.

강의 교재

일차적으로 GitHub에서 다운로드 받을 수 있는 강의안을 충실히 공부하면서 주교재(장혜정·김현우 2022)를 통해 복습해야 합니다. 강의안과 주교재는 앞으로 1년 반 이상 전공필수 수업(사회통계연습 및 사회조사방법론)에서 계속 사용합니다. 수업시간에는 큰 그림과 논리적 흐름을 강조하므로 세밀한 내용을 모두 다룰 수 없습니다. 그러므로 수업이 끝난 뒤 강의교재를 활용해 개인적으로도 반드시 복습을 해야 합니다. (수업에서 다루어진 주제에 한하여) 책에 제시된 문제들도 충실히 풀어보아야 합니다.

- 장혜정·김현우. 2022. 『통계학노트』. 한빛아카데미. (필수)

『통계학노트』는 3월 중순경 인터넷으로 구매가 가능할 것이며, 구매방법은 다시 공지합니다. 원래 2월말 발간 예정이었으나, 선거로 인해 인쇄업체들이 물량을 소화 못해서 출판이 지연되었습니다. 3주차까지는 강의안을 충실히 복습하면 충분할 것으로 기대합니다. 첫 2주 동안 교재 인쇄본을 구입할지 스스로 결정하고, 구매 안내가 나오면 가능한 신속한 구매를 당부합니다. 이번 교재는 저자들의 수업에서만 사용하기 위해 출판하였으므로, 학생들의 일반적인 교재 구매율 60% (약 240부)를 기준으로 인쇄하게 되었습니다. 판매가 완료된 후에는 구매가 불가능함 또한 미리 알려드립니다.

학습 보조자료

- YouTube 또는 Google에서 궁금한 주제를 Excel과 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!

강의 구성

- 강의 및 실습: 모든 학생은 반드시 수업에 참여해야 합니다. 수업을 통해 기초적인 개념과 방법을 복습한 뒤, 엑셀을 사용하여 실습합니다. 수업 내용은 진행될수록 누적되기 때문에 결석 혹은 지각은 향후 이해에 큰 방해가 됩니다. 비대면 수업인만큼 카메라를 켜고 얼굴을 비추어 주십시오. 강의 중간중간에 실습이 함께 진행됩니다. 실습 시간에는 각자의 컴퓨터 환경에서 엑셀을 사용하여 실제 분석을 진행합니다.
- 퀴즈: 거의 매주 퀴즈(take-home quiz)가 아주 조금씩 주어집니다. 각 퀴즈는 반드시 혼자서 도전해야 합니다. 퀴즈 하나하나의 점수는 매우 작지만 쌓이면 큰 점수(전체의 30%)가 됩니다. 다음 주 수업 전일 자정까지 제출해야 하며, 이 데드라인을 넘기면 무효 처리됩니다.

- 시험: 중간고사와 기말고사가 예정되어 있습니다. 컴퓨터를 사용하여 시험을 치릅니다. 원한다면 책이나 인터넷을 참고할 수도 있지만 과연 그럴 시간이 있을까요?

학점 안내사항

모든 과제와 시험 점수는 종합정보시스템에 그때그때 업로드됩니다. 이의사항이 있는 경우 점수가 업로드된 즉시 제시되어야 합니다. 최종 학점은 다음 기준에 따라 산출됩니다.

- 출석 및 수업 참여 (10%)
- 중간시험 (30%)
- 퀴즈 (30%)
- 기말시험 (30%)

수강생 유의사항

- 모든 수업은 별도의 안내가 없는 이상 원칙적으로 **대면**으로 진행됩니다.
- 수업 전일에 해당 주차 강의안과 퀴즈가 다음 GitHub 레파지토리에 업로드됩니다:
<https://github.com/hxk271/SocStats>
- 전체 수업의 4분의 1이상을 결석한 경우 시험에 응시할 수 없습니다.
- 시험관련 부정행위자로 판명되었을 때는 학칙 또는 내규에 의거 해당 교과목의 성적을 취소합니다.
- 조기 취업자는 강사와 상담 후 출석과 과제, 중간/기말시험 등을 적절한 과제로 모두 대체할 수 있습니다. 이 경우 점수 상한은 B+로 합니다.

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 과목을 하나를 만드는데도 집단적인 노력이 필요합니다. 이 수업의 많은 부분은 제가 스스로 만든 것이 아닙니다. 여기에는 Sarah Font (Penn State), 장혜정(경희대), 박희제(경희대) 등 많은 분들의 강의자료를 활용하였습니다. 도움을 주신 모든 분들께 감사드립니다. 여러분의 피드백은 다음 이 수업을 더욱 발전시키는데 큰 도움이 됩니다.

토픽 개요

기술통계, 교차분석, 그리고 통계적 추정의 논리

- 1주차 강의 소개/통계학과 자료의 이해
- 2주차 자료의 시각화
- 3주차 자료의 수치요약
- 4주차 자료의 연관성 측도
- 5주차 확률이론
- 6주차 이론적 확률분포
- 7주차 표집분포와 추정
- 8주차 중간시험

목표: 기술통계분석, 교차표/그래프, 확률분포/표집, 가설검정의 논리를 설명할 수 있다.

분석기법의 활용

- 9주차 단일집단 가설검정
- 10주차 두집단 가설검정
- 11주차 분산분석
- 12주차 카이제곱분석
- 13주차 상관분석과 단순회귀분석
- 14주차 다중회귀분석
- 15주차 기말시험

목표: t test/ANOVA, 상관분석/회귀분석을 수행하고 해석할 수 있다.

세부 일정

1주차 강의 소개/통계학과 자료의 이해

TOPICS 사회통계의 필요성; 수업의 구성과 당부사항; 통계분석 패키지 비교; Empirical Science; Methods/Methodology; Statistics; Data; Variables; Observations; Quantifiability; Measurement; Descriptive/Inferential Statistics; Unit of Observations/Analysis; Nominal, Ordinal, Interval, and Ratio Scales; Continuous and Categorical Scales; Likert Scale; Approximately Interval

GOAL 이 수업의 구성과 진행 방식을 파악한다; 왜 사회통계학을 배워야 하는지 이해한다; 통계학의 정의와 기본 개념을 이해하고 용어를 학습한다; 측정의 원리와 척도 유형을 구분할 수 있다; 자료유형의 구분이 왜 중요한지 이해한다

2주차 자료의 시각화

TOPICS Cartesian Coordinates; Histogram; Boxplot; Pie Chart; Box-Whisker Plot; Time-Series Line Graph

GOAL 자료의 시각적 표현 또는 시각화의 필요성과 장점, 기본 원칙을 이해한다; 분석 목적과 자료유형에 따라 적용할 수 있는 시각화 방법을 이해한다; 단일 질적변수를 적절한 방법을 이용하여 시각적으로 표현할 수 있다; 단일 양적변수를 적절한 방법을 이용하여 시각적으로 표현할 수 있다. 다양한 그래프를 엑셀로 구현할 수 있다; 그래프의 조작을 둘러싼 트릭을 살펴본다

3주차 자료의 수치요약

TOPICS Central Tendency; Mean, Median, and Mode; Frequency Distribution Table; Dispersion; Range, Variance, and Standard Deviation; Labels

GOAL 자료 유형별로 상이한 기술통계량의 종류를 설명할 수 있다; 중심경향, 자료의 위치, 변동성을 요약하는 방법들을 이해하고 산출할 수 있다; 치우침/집중도 등 수치로 요약하는 다른 방법이 있음을 이해한다; 경험법칙(the Empirical Rule)과 체비셰프 부등식(Chebyshev's inequality)을 설명할 수 있다; 엑셀로 실제 자료로부터 기술통계치를 구할 수 있다

세부 일정 (계속)

4주차 자료의 연관성 측도

TOPICS Covariance; Correlation; Scatterplot; Time-Series Analysis; Cross-Sectional Analysis; Hypothesis; Independent and Dependent Variable; 변수간 연관성; Marginals; Mediating, Moderating, and Spurious relationships

GOAL 자료유형에 따라 사용할 수 있는 연관성 측도가 다름을 이해하고, 적절한 연관성 측도를 선택할 수 있다; 두 양적변수 간의 관계를 적절히 수치적, 시각적으로 표현할 수 있다; 두 질적변수 간의 관계를 적절히 수치적, 시각적으로 표현할 수 있다; 양적변수와 질적변수의 관계를 적절히 수치적, 시각적으로 표현할 수 있다; 조건부 관계와 허위적 관계를 이해한다

5주차 확률이론

TOPICS Probability; Marginal Probability; Conditional Probability; Bayes Rule; Random Variable; Probability Distribution; Discrete Random Variable; Continuous Random Variable; Probability Density Function

GOAL 확률의 개념과 용어를 이해한다; 상호배타성과 독립성, 덧셈법칙과 곱셈법칙을 이해한다; 결합확률, 한계확률과 조건부확률을 이해하고 계산할 수 있다; 베이즈 정리를 이해하고 사후확률을 계산할 수 있다; 확률변수와 확률분포의 개념을 이해하고, 이산형과 연속형 확률변수를 구분하여 설명할 수 있다

6주차 이론적 확률분포

TOPICS Bernoulli Distribution; Binomial Distribution; Uniform Distribution; Normal Distribution; Unit Variance; Standardization; Z Score; Standard Normal Distribution

GOAL 이산형과 연속형 각각의 대표적인 확률분포를 구분하여 적합한 상황을 설명할 수 있다; 각 확률분포 가운데 대표적인 이론적 분포를 이해한다; 각 확률분포의 확률질량함수 또는 확률밀도함수를 설명할 수 있다; 각 확률분포의 (구간) 확률을 계산하는 방법을 안다; 정규분포의 확률값에 대응되는 관찰값을 구할 수 있다

세부 일정 (계속)

7주차 표집분포와 추정

TOPICS Probability and Non-probability Sampling; Sampling Distribution; Standard Error of the Mean; Central Limit Theorem; Law of Large Numbers; The 68, 95, and 99 Rule; Statistical Estimation; Point and Interval Estimation; Confidence Interval (CI); Degree of Freedom

GOAL 표집분포의 개념과 중심극한정리의 핵심 원리를 이해한다; 평균과 비율에 대한 표집분포의 기대값과 분산을 구할 수 있다; 추정의 개념을 이해하고 점추정과 구간 추정을 할 수 있다; 구간 추정의 의미를 해석할 수 있고, 구간 길이에 영향을 미치는 요인을 이해한다; 모수와 상황에 따라 적용하는 분포가 달라짐을 이해하고 각각의 신뢰구간을 산출할 수 있다

8주차 중간시험

9주차 단일집단 가설검정

TOPICS Sampling Error; Sampling Distribution of the Mean; Standard Error of the Mean; Confidence Interval and Confidence Level; Confidence Intervals for Means/Proportions; Standard Error of the Means/Proportions; p -value; Margin of Error

GOAL 표본의 통계량을 이용하여 모집단의 모수에 관한 가설을 검정하는 5단계 절차를 이해한다; 단일모집단의 모평균, 모비율, 모분산의 가설을 검정하기 위한 검정통계량의 적절한 분포를 선택할 수 있다; 임계치와 기각역을 설정하고 귀무가설 기각 여부를 판단하여 의미를 해석하고 결론을 도출할 수 있다; 가설검정과 신뢰구간의 관계를 이해한다.

세부 일정 (계속)

10주차 두집단 가설검정

TOPICS Sampling Distribution of the Differences between Means; Null and Research Hypotheses for Comparing Means; Type I error or α ; Z test; t distribution; t test

GOAL 단일집단 가설검정을 확장하여 두 모집단의 모수를 비교하는 가설검정의 개념을 이해한다; 독립적인 두 집단의 모평균 비교를 위한 적절한 분포를 선택하여 가설을 검정할 수 있다; 두 집단이 쌍(pair)으로 이루어진 경우, 두 집단의 모평균을 비교하는 가설을 검정할 수 있다; 독립적인 두 집단의 모비율을 비교하는 가설을 검정할 수 있다; 독립적인 두 집단의 모분산을 비교하는 가설을 검정할 수 있다; 엑셀로 t -test를 수행하고 그 결과를 해석할 수 있다.

11주차 분산분석

TOPICS F value; F Distribution; One-Way ANOVA; Treatment and Control Groups

GOAL 세 그룹 이상의 평균을 비교하는 분석 방법인 분산분석을 개념적으로 이해한다; 분산분석의 기초인 분산의 역할을 이해하고 분해할 수 있다; 그룹 간 및 그룹 내 제곱합, 평균제곱을 산출하여 분산분석표를 구성할 수 있다; 일원분산분석을 이용하여 세 그룹 이상의 모평균이 같은지 가설을 검정할 수 있다; 엑셀로 일원분산분석을 수행하고 그 결과를 해석할 수 있다

12주차 카이제곱분석

TOPICS Chi-square Test; Observed Frequencies; Expected Frequencies

GOAL 범주형 자료를 요약하는 분할표를 개념적으로 이해한다; 분할표를 분석하면서 독립성, 동질성 및 적합성을 검정하는 맥락과 목적을 구분하여 설명할 수 있다; 변수 간의 관계를 분석하는 독립성(independence) 검정을 수행할 수 있다; 둘 이상의 독립적인 표본들의 분포가 같은지 분석하는 동질성(homogeneity) 검정을 수행할 수 있다; 관측된 자료가 주어진 확률분포를 따르는지 적합성(goodness of fit) 검정을 수행할 수 있다; 엑셀로 카이제곱 검정을 수행하고 그 결과를 해석할 수 있다

세부 일정 (계속)

13주차 상관분석과 단순회귀분석

TOPICS Pearson's Correlation Coefficient; Correlation and Causation; Visualization of the Correlation; Positive/negative Relations; Non-relationship; Trend Line; Best-Fitting Line; Error and the Least Square; The Sum of the Squared Error; Slopes and the Intercept of the Regression Equation; Prediction

GOAL 두 양적 확률변수간 연관성을 분석하는 상관계수를 산출하고, 상관계수의 유의성을 검정할 수 있다; 두 양적 확률변수 간의 관계를 직선 수식으로 표현하는 회귀분석의 개념을 이해한다; 최소제곱법으로 단순회귀식을 산출할 수 있다; 회귀식 모형의 적합도와 회귀계수의 유의성을 이해한다; 엑셀로 상관분석과 단순회귀분석을 수행하고 그 결과를 해석할 수 있다

14주차 다중회귀분석

TOPICS Control Variable; Total, Explained, and Unexplained Variations; The Coefficient of Determination

GOAL 단순회귀분석을 다중회귀분석으로 확장할 수 있다; 다중회귀분석 모형의 회귀계수에 대한 유의성을 검정할 수 있다; 더미변수를 다중회귀모형에 포함하여 분석할 수 있다; 적합도 검정으로 F 표와 결정계수를 계산하고 해석할 수 있다; 엑셀로 다중회귀분석을 수행하고 그 결과를 해석할 수 있다

15주차 기말시험