

사회통계연습  
INTERMEDIATE SOCIAL STATISTICS  
전공필수(3-3-0)

2021년 2학기

Instructor	김현우 (hxx271@psu.edu)
Office Address	N15동 414호
Office Hours	메일로 연락후 면담
Classroom	N15동 427호
Class Hours	10:00 AM-1:00 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 절차와 방법에 관해 배웁니다. 여러분이 1학기에 배운 사회통계학 내용을 기초로 하여, 확률분포와 가설검정의 논리를 직접 실험해보고, 자료 리코딩/변수구축, 평균비교와 분산비교, 상관분석과 회귀분석 등 기초적인 통계학의 도구를 반복 숙달하게 됩니다. 이 수업을 통해 여러분은 그동안 배운 사회학적 상상력의 많은 부분을 현실 데이터에 기반해 직접 테스트해 볼 기회를 갖게 됩니다. 이 수업의 궁극적인 목적은 사회학적 상상력이 가진 현실적 유용성을 여러분 스스로가 깨닫도록 하는 것에 있습니다.

이 수업은 여러분의 관련 자격증 취득과도 직접적으로 연관되어 있습니다. 저번 학기 **사회통계학**에서 여러분이 컴퓨터활용능력 2급의 스프레드시트 과목과 사회조사분석사 2급의 사회통계 과목에 해당하는 내용을 배웠다면, 이번 **사회통계연습**에서는 본격적으로 사회조사분석사 2급의 실기 시험에 해당하는 내용을 배우게 됩니다. 내년엔 **사회조사분석**과 **사회조사실습**까지 모두 이수하면 여러분은 사회조사분석사 2급과 컴퓨터활용능력 2급을 위한 핵심 연습을 마치게 됩니다.

이 수업은 사회학과 학부과정 2학년생을 기준으로 설계되었으며 그동안 오로지 하나의 사회통계학 수업만을 이수하였을 것으로 전제합니다. 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).

이 수업은 철저히 실기를 중심으로 구성되어 있습니다. 여러분은 매일매일 빠지지 않고 수업에 참석해야 하고 수업 후에도 연습을 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

사회통계(5663011) 또는 학칙에 따라 이에 준하는 기초통계학 수업을 반드시 이수하여야 합니다.

## 강의 교재

유인물을 배부하므로 필수 교재는 없습니다. 아래는 어디까지나 참고용이지만 **적어도 한 권**은 꼭 구매를 추천합니다. Jaggia의 책은 한 권으로 사회통계연습에서 데이터과학의 주요 토픽까지 모두 커버하는 좋은 책입니다. 분석도구로는 Excel과 R을 사용합니다. 김준호·노성호(2012)는 SPSS와 R을 사용하는 방식을 보여줍니다. 수업시간에는 큰 그림과 논리적 흐름을 강조하므로 세밀한 내용을 모두 다룰 수 없습니다. 그러므로 수업이 끝난 뒤 강의교재를 활용해 개인적으로도 반드시 복습을 해야 합니다.

- Jaggia, Sanjiv, Alison Kelly, Kevin Lertwachara, Leida Chen. 2021. 『비즈니스 애널리틱스』. 한빛아카데미. 그린. (참고)
- 김준호·노성호, 2012. 『사회과학을 위한 통계와 분석: SPSS와 R을 중심으로』. 그린. (참고)

## 학습 보조자료

- YouTube 또는 Google에서 궁금한 주제를 Excel과 연관검색어로 하여 찾아보면 엄청나게 많은 자료가 있습니다. 궁금한 점이 있으면 망설이지 말고 즉시 영어로 검색해보기를 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. 다음 주 수업 시작 전까지 제출해야 하며, (수업이 끝날때 제출하는 등) 이 데드라인을 넘기면 감점됩니다.
- 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 **비대면**으로 진행됩니다.
- 수업 전일에 링크와 유인물을 담은 전체메일이 발송됩니다.
- 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 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

### 토픽1: 기술통계, 교차분석, 그리고 통계적 추정의 논리

1주차	Course Introduction/What is Empirical Science?
2주차	Measures of Central Tendency and Dispersion
3주차	Cross-Tabulations and Tables
4주차	Graphs and Visualization
5주차	Probability and Probability Distribution
6주차	Sampling Distribution and Estimation
7주차	Hypothesis Test
8주차	Midterm

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

### 토픽2: 분석기법의 활용

9주차	Mean Comparison
10주차	Analysis of Variance
11주차	More on Cross-Tabulations and Tables
12주차	Reliability and Correlation Analysis
13주차	Regression Analysis
14주차	More on Regression Analysis
15주차	Final

목표:  $t$ -test/ANOVA, Cronbach's  $\alpha$ , 상관분석/회귀분석을 수행하고 해석할 수 있다.

## 세부 일정

### 1주차 Course Introduction/What is Empirical Science?

- TOPICS 사회통계연습의 필요성; 수업의 구성과 당부사항; 통계분석 패키지 비교; Empirical Science; Methods/Methodology; Statistics; Data; Variables; Observations; Quantification; Measurement; Descriptive Statistics; Nominal, Ordinal, Interval, and Ratio Scales; Continuous and Categorical Scales; Likert Scale
- GOAL 이 수업의 구성과 진행 방식을 파악한다; 왜 사회통계학을 새삼 “연습”할 필요가 있는지 이해한다; 데이터의 개념을 이해한다; 측정도구와 척도의 개념을 이해한다

### 2주차 Measures of Central Tendency and Dispersion

- TOPICS Central Tendency; Mean, Median, and Mode; Frequency distribution table; Dispersion; Range, Variance, and Standard Deviation; 데이터 불러오기와 저장; 레이블; Unit of Analysis
- GOAL 평균, 최빈치, 중위값 개념에서 출발하여 분산 및 표준편차를 복습한다; Excel로 실제 데이터의 기술통계치를 구한다

### 3주차 Cross-Tabulations and Tables

- TOPICS Cross-Sectional Analysis; Hypothesis; Independent and Dependent Variable; 변수간 연관성; Marginals; Mediating, Moderating, and Spurious relationships
- GOAL 교차표(cross-tabulations)의 논리와 해석을 이해한다; 조건부 관계와 허위적 관계를 이해한다

### 4주차 Graphs and Visualization

- TOPICS Scatterplot; Histogram; Boxplot; Pie Chart; Time-Series Analysis; X-axis/Y-axis;
- GOAL 다양한 그래프의 양식을 살펴보고 Excel로 그것들을 구현한다; 그래프의 조작을 둘러싼 트릭을 살펴본다

## 세부 일정 (계속)

<b>5주차</b>	<u>Probability and Probability Distribution</u>
TOPICS	Probability; Marginal Probability; Conditional Probability; Bayes Rule; Random Variable; Probability Distribution; Discrete Random Variable; Continuous Random Variable; Probability Density Function; Bernoulli Distribution; Binomial Distribution; Uniform Distribution; Normal Distribution; Unit Variance; Standardization; z-Score; Standard Normal Distribution
GOAL	확률의 기초적인 개념을 이해한다; 한계확률과 조건부확률을 계산할 수 있다; 확률변수와 확률분포의 아이디어를 이해한다; 대표적인 이산확률분포와 연속확률분포를 학습한다; 정규분포와 표준정규분포의 특성을 이해한다; 원점수의 z-score 변환을 수행할 수 있다
<b>6주차</b>	<u>Sampling Distribution and Estimation</u>
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	표집과정을 좀 더 구체적으로 이해한다; 대수의 법칙과 중심극한정리를 실험적으로 살펴본다; 통계적 추정의 논리를 이해한다
<b>7주차</b>	<u>Hypothesis Test</u>
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	표집분포의 평균과 표집오차 개념을 실습한다; 평균과 비율의 신뢰구간 개념을 실습한다
<b>8주차</b>	<u>Midterm Exam</u>

## 세부 일정 (계속)

### 9주차 Mean Comparison

TOPICS Sampling Distribution of the Differences between Means; Null and Research Hypotheses for Comparing Means; Type I error or  $\alpha$ ;  $z$ -test;  $t$ -distribution;  $t$ -test

GOAL 평균비교의 의미와 목적을 파악한다;  $t$ -test를 수행하고 그 결과를 해석한다.

### 10주차 Analysis of Variance

TOPICS F value; F-Distribution; One-Way ANOVA; Two-Way ANOVA

GOAL 분산분석의 의미와 목적을 파악한다; 간단한 ANOVA를 수행하고 그 결과를 해석한다

### 11주차 More on Cross-tabulations and Tables

TOPICS Chi-square test; Observed frequencies; Expected frequencies

GOAL 카이제곱 검정을 수행하고 그 결과를 해석한다

### 12주차 Reliability and Correlation Analysis

TOPICS Covariance; Cronbach's alpha; Correlation; Pearson's Correlation Coefficient; Correlation and Causation; Visualization of the Correlation; Positive/negative Relations; Non-relationship

GOAL Multiple-item scales을 구성할 수 있다; 신뢰도의 평가측도 중 하나로 Cronbach's alpha를 이해한다; 상관관계의 존재 여부와 강도를 그림으로 살펴본다; Pearson의 상관계수의 논리를 이해하고 실제로 계산한다

## 세부 일정 (계속)

### 13주차 Regression Analysis

TOPICS	Trend Line; Best-Fitting Lin; Error and the Least Square; The Sum of the Squared Error; Slopes and the Intercept of the Regression Equation; Prediction
GOAL	이변량(bivariate) 회귀분석의 아이디어를 복습하고 실행한다; 회귀계수와 표준오차를 해석한다

### 14주차 More on Regression Analysis

TOPICS	Control Variable; Total, Explained, and Unexplained Variations; The Coefficient of Determination; Principal Component Analysis
GOAL	다변량(multivariate) 회귀분석의 아이디어를 복습하고 실행한다; 다변량 맥락에서 회귀계수와 표준오차를 해석한다; 주성분 분석의 아이디어를 간략히 이해하고 실행한다

### 15주차 Final Exam