

사회통계연습

강의 소개

김현우, PhD¹

¹충북대학교 사회학과 부교수



진행 순서

- 1 수업의 구성과 당부사항
- 2 통계 분석 패키지 비교
- 3 증거에 기반한 사회학 연구



수업의 구성과 당부사항



수업의 구성과 당부사항

사회통계학은 두 개의 얼굴을 가지고 있다.

- 과학(science)과 손기술(art)! 그러므로 우리 학과의 수업 역시 “사회통계”와 “연습”으로 나뉘어 있다.
- 손기술은 생각보다 매우 중요하며 연습해두지 않으면 연구나 실무에서 통계 분석을 전혀 쓸 수 없다.
- 물론 컴퓨터활용능력 2급과 사회조사분석사 2급 취득에도 손기술은 필수적이다.
- 투자한 시간이 절대적으로 많아야 한다. 머리 못지 않게 엉덩이로 공부한다!



수업의 구성과 당부사항

이 수업에는 선수과목이 있다.

- 사회통계(5663011)는 필수적으로 이수해야 하며 학점이 D 이하이면 해당 수업을 재수강할 필요가 있다.
- 시스템 미비를 이용해 선수과목을 듣지 않고 들어오는 행위는 매우 바람직하지 않다 (“선수과목을 이수하지 않은 학생을 위한 배려가 부족하다”?).
- 선수과목의 내용은 수업 중에 조금씩 복습하지만 동시에 숙제로도 복습하게 된다.
- 수업 교재는 지난 학기에 사용한 <경영경제 통계학 노트>이다.
- 추가적으로 필요한 교재는 도서관에서 자유롭게 선택할 수 있다.



수업의 구성과 당부사항

온라인 강의와 오프라인 강의를 병행한다.

- eCampus에서 온라인 동영상 강의를 시청하자. 10일 정도 간격으로 출석이 인정된다(놓치지 않게 주의하자).
- 온라인 강의는 매주 수요일에 진행된다. 실습 중심 수업이 될 예정이다.
- 오프라인 강의는 매주 수요일에 진행된다. 이때 숙제를 풀어보이고 질의응답을 받는다.



수업의 구성과 당부사항

강의계획서 내용을 잘 살펴보아야 한다.

- 2번 지각은 1회 결석이다. 3회를 초과한 결석은 낙제로 처리된다.
- 지각, 결석, 과제 불성실 등으로 인한 낙제에 대한 개별적인 안내는 없다.
- 강의계획서는 수업이 진행하면서 계속 업데이트된다. 이후 여러분에게 불리한 방향으로 변하지 않는다(그러나 스스로 규칙적으로 확인할 것!).



수업의 구성과 당부사항

강의실 밖에서도 복습과 퀴즈로 공부가 필요하다.

- 모든 강의안(pdf 파일), 연습용 자료와 퀴즈는 수업 전까지 GitHub에 업로드된다:
<https://github.com/hxk271/SocStatsPrac>
- 단 수업 중 사용할 통계 자료 중 공개배포가 어려운 것들은 eCampus에 업로드된다.
- 수업으로 피로할 수 있지만 복습은 12시간 이내에 하는 편이 좋다(배운 걸 잊어버리기 때문이다).
- 복습을 통해 이 수업이 예전 학기에 배운 내용과 이어진다는 점을 파악하자.



수업의 구성과 당부사항

점수는 노력에 대한 선물이다.

- 이 수업은 (아마도) 제법 힘들다. 나는 여러분이 노력을 거듭해 이 수업에서 많은 지식과 스킬을 배우기를 희망한다. 그 댓가는 관대한 점수이다.
- 약속 하나: 조건없이 여러분 가운데 30%는 A 이상, 40%는 A-B 사이, 30%는 C 이하를 받게 된다. 이것은 학사 시스템에 따른 제약이다. 시스템 권한이 허락하는 한 가장 관대하게 점수를 줄 예정.
- 그러므로 배움에 최대한 집중하면 여러분이 받을 점수에는 섭섭하지 않을 것.
- 통계학 유관 분야의 타과(특히 고학번) 수강생에 대해서는 점수 패널티가 있다 (복수전공자 예외)



통계 분석 패키지 비교



통계 분석 패키지 비교

왜 통계 분석 패키지를 배워야 할까?

- 사회통계(2학년 1학기)에서는 엑셀을 사용하여 기초를 배웠다.
- 이번 학기 사회통계연습(2학년 2학기)에서는 좀 더 본격적인 통계 분석 패키지 사용법을 배운다.
- 사회통계(2학년 1학기)에서는 어디까지나 기초만 배웠다.
- 그러므로 여러분이 “뭔가 배우긴 했는데 실무나 연구에서 해보라고 하면 자신이 없는데...” 라고 느낀다면 지극히 정상이다!

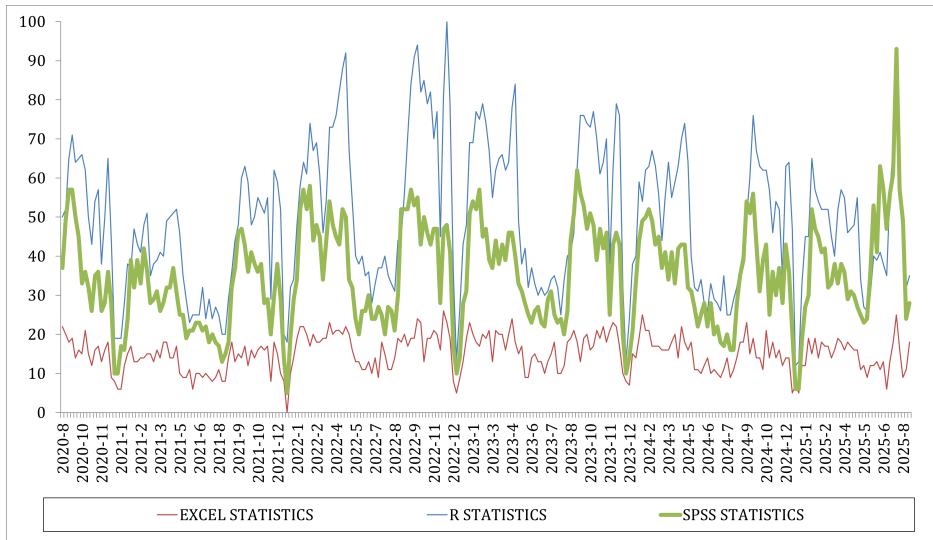


통계 분석 패키지 비교

- 본격적인 분석을 위해 반복 숙달이 필요하다.
- (그렇다고 엑셀을 지루하게 반복하는 것이 아니라) 편리하게 통계 분석을 위한 소프트웨어를 사용하게 된다.
- 연습에서 수학 따위는 전혀 사용하지 않는다(이것은 장인이 공학을 필요로 하지 않는 것과 마찬가지이다).
- 그럼 우리는 왜 지난 학기에 수학을 배웠을까?
- 스스로의 통계 분석에 확신을 갖고 남에게 설명할 수 있기 위해서이다(Why?). 이 부분은 생성형AI의 시대에 그 무엇보다 중요하다.



통계 분석 패키지 비교



소스: Google Trends

통계 분석 패키지 비교

R은 통계학과 데이터사이언스 분야에서 폭넓게 사용된다.

- 유저들의 커뮤니티가 활성화되어 있어 지금 이 순간에도 통계 분석을 위한 여러 새로운 도구가 개발되어 무료로 공개되고 있다.
- 단순한 통계 분석, **시각화(visualization)**, **데이터 사이언스(data science)**에 이르기까지 다양한 분야에서 활용되는 분석툴이다.
- **오픈소스(open source)** 소프트웨어이므로 그 사용이 무료다.
- 배우기 어렵고 수많은 시행착오를 요구하며 (통계학이나 데이터사이언스 자체를 떠나) R 자체를 배우는데 상당한 시간을 써야 한다.
- (약간 주관적이지만) 남이 쓴 코드를 알아보기 어렵다. 즉 코드가 심미적으로 지저분하다.



통계 분석 패키지 비교

- 또다른 대안은 파이썬(Python)이다.
- 전세계의 개발자들 사이에서 R vs. Python은 하나의 밈이다.
- 관심이 있다면 R과 Python을 동영상 강의 등을 통해 각자 스스로 배울 수 있다.
- 우리 학과에서는 소셜데이터사이언스 수업이 개설되고 여기서 Python의 기초를 배울 수 있다.



통계 분석 패키지 비교

엑셀은 그 나름대로 중요한 강점을 가지고 있다.

- R보다 훨씬 배우기 쉽고 졸업 후 실무에서도 훨씬 다양한 분야에서 압도적으로 많이 쓰인다.
- 컴퓨터활용능력시험 자격증에 어차피 엑셀이 필요하다.
- 분석이 이루어지는 과정을 실시간으로 지켜볼 수 있다. 비슷한 도구는 Matlab이 있는데 이쪽은 배우기가 R만큼이나 어렵다.
- 많은 학생들이 이미 가지고 있거나, 설령 없어도 개신누리에서 무료로 다운로드 받을 수 있다
- 엑셀은 “통계 학습의 도구”로는 유용해도 “통계 분석의 도구”로는 상당히 불편하다 (불가능하지 않다).
- (억지로 또다른 단점을 생각해 보면) 엑셀을 너무 잘하면 회사에서 유능한 인간으로 취급받으므로 일거리가 늘어난다.



통계 분석 패키지 비교

SPSS가 우리의 선택이 될 것이다.

- SPSS는 배우고 사용하기 쉽다. 졸업 후 학교 밖 현업 기관에서도 폭넓게 사용된다.
- 이용자 규모가 상당히 크기 때문에 인터넷을 통해 모르는 것을 질문하거나 자료를 찾아보기에 용이하다.
- 치명적인 결함은 (1) 개인이 구매하기에는 너무 비싸고 (2) 고급 통계 분석에는 제약이 있다는 것이다.
- 요즘에는 빅데이터 붐으로 학교에서도 점차 SPSS 대신 R을 가르치자는 유행이 있다. 취지는 이해하지만 **학습 곡선(learning curve)**이 가파르기 때문에 중도포기자가 다수 나온다.
- 사회통계를 배우기도 벅찬데 (R까지 함께 배우는 부담은 좀 줄이고) 일단 SPSS로 시작하는 것도 나름 괜찮은 선택이다(나도 그랬다).



통계 분석 패키지 비교

- 그런데 SPSS 라이선스는 상당한 고가이므로 개인은 물론이고 기관도 구매해주지 않을 수도 있다.
- 이런 경우 (SPSS만큼 성능이 뛰어나지는 못해도) 이와 비슷한 오픈소스 소프트웨어를 사용한다.
- 특히 JASP와 PSPP 그리고 Jamovi가 널리 알려져있다.
- 이에 따라 우리도 동영상 강의에서 Jamovi를 배울 것이다! 무엇보다 온라인 강의를 통해 여러분이 집에서 SPSS를 실습하기 어렵다.
- 매주 수요일 전산실에서 SPSS 실습을 한다(사회조사분석사 자격증에도 도움이 된다).
- Jamovi와 SPSS는 매우 비슷하므로 금방 둘 다 배울 수 있다.



증거에 기반한 사회학 연구



증거에 기반한 사회학 연구

경험과학이란 무엇인가?

- 경험주의(empiricism)에 기초하고 있는 과학이다.
- 경험주의는 감각 경험을 지식의 유일한 원천으로 보는 과학철학의 한 조류이다.
- 경험과학적 연구는 경험적으로 검증가능한 명제(propositions) (e.g., “지구는 달보다 크다”)를 생산하는 것을 주 임무로 삼는다.
- 그 결과는 과학적 지식이 되어 축적된다.
- 이때, 경험에 의한 검증가능성(verifiability)이 생명이다(어원: Verus/참).



증거에 기반한 사회학 연구

무엇이 경험과학적 지식이 아닌가?

- 논리적으로 검증가능한 명제(e.g., “오각형의 내각의 합은 삼각형 세 개의 내각의 합과 같다”)는 경험과학이 지향하는 지식과는 방향성이 약간 다르다.
- 초자연적이고 형이상학적 명제들, 특히 신학적 명제(e.g., “알라는 붓다보다 위대하다”)와는 배척된다.
- 윤리적 명제(e.g., “사람은 약속을 지켜야 한다”)나 미적 명제(e.g., “그의 외모는 아름답다”)와도 상관없다.



증거에 기반한 사회학 연구

다만 위 명제들을 경험적 명제로 뒤바꾸어 경험과학적으로 연구할 수 있다.

- (1) “방과후 교육에 참여한 빈도가 높을수록 오각형의 내각의 합은 삼각형 세 개의 내각의 합과 같다는 명제가 참이라고 응답할 가능성이 높다.”
- (2) “파키스탄 이주민은 이란 이주민보다 알라는 붓다보다 위대하다고 응답할 가능성이 높다.”
- (3) “지난 일주일 사이 연락한 사람의 수가 많을수록 사람은 약속을 지켜야 한다는 명제에 보다 더 동의할 것이다.”
- 이제 우리에게는 위와 같이 검증가능한 명제들이 주어졌다.
- 경험과학으로서 사회학적 연구는 위와 같은 명제들을 증거를 통해 검증가능하는데 주력한다.



증거에 기반한 사회학 연구

경험과학적 주장에는 반드시 증거가 필요하다.

- 주장(argument)에는 증거(evidence)가 필요하고, 증거는 자료(data)를 토대로 한다.
- (너무나도 당연한 말이지만) 자료가 축적되기 위해서는 관찰(observation)이 선행되어야 한다.
- 경험과학에서 정확한 관찰은 너무나도 중요하기 때문에 엄격한 관찰의 과정과 절차에 특별한 의미를 부여한다!
- 측정(measurement)이란 어떤 현상을 관찰한 후 일정한 규칙에 따라 수치를 부여하는 것이다.
- (사회학의 맥락에서 본다면) 이런 행위에는 사회 현상을 계량화(quantify)할 수 있다는 세계관을 내포하고 있다(어떤 사람은 이 관념을 극도로 배격한다).



증거에 기반한 사회학 연구

사회통계학은 과학적 연구에만 중요한가?

- 지금까지의 설명은 결국 과학적 연구에 관한 것인데 모든 사람이 교수·과학자·연구원이 되는 것은 아니다.
- 대학이 인문학의 산실로 여겨지고, 대학생이 지성인을 지망할 때 이런 질문은 아예 제기되지 않았다(기꺼이 이런 “쓸모없는” 지식을 가르치고자 대학이 학생을 받았고, 그런 지식을 배우고자 학생이 대학에 왔다).
- 하지만 이제 교수와 대학생이 대학에서 환금성 높은 지식을 거래하는 것은 평범하게 받아들여진다.
- “이런 시대에 우리는 무얼 위해 사회통계학을 공부해야 하나?”



증거에 기반한 사회학 연구

사회통계는 (과학적 연구 목적 이외에도) 현실 사회 곳곳에서 사용된다.

- 모두가 동의하는 주장에 대해서는 별로 논리나 증거가 필요하지 않다(Why?).
- 그러나 다원화된 현대사회에서 견해가 갈리기 시작하면 그때부터는 설득이 필수적이다.
- 설득에도 여러가지 전략이 있지만 논리와 증거를 사용하는 것이 으뜸이다.
- (흔한 오해나 착각과는 달리) 설득의 순간(!)이 따로 있는 것이 아니다. 여러분의 토크, 발표, 보고 등 매 순간순간에 설득의 언어가 동원된다.



증거에 기반한 사회학 연구

기획·마케팅·홍보에서도 데이터 분석은 필수불가결하다.

- 모든 사업체가 궁금해야 한다: “누가 언제 어디서 우리가 제공하는 상품과 서비스를 구입하는가?”
- 이 상권에서 유동인구와 동종 경쟁업체 수를 고려했을 때 예상 매출액은 얼마인가?
- 요즘 10대에게 어필하려면 어떤 **상품기획(merchandising)**을 실천해야 할까?
- 수익을 극대화하려면 어떤 식으로 **제품 브랜딩(product branding)**을 해야 할까?
- 다른 사업체와의 경쟁에서 살아남으려면 어떤 **시장분할 전략(market segmentation strategy)**을 세워야 할까?
- 이 모든 질문에 감으로 어렵짐직하기 보다는 데이터로 분석하는 편이 훨씬 믿음직하다.



증거에 기반한 사회학 연구

데이터는 이미 이 시대의 원유가 되었다.

- (좋은 싫든) 통계와 데이터 분석은 이미 폭넓게 사용되어 왔고 앞으로도 더욱 확대될 것이다.
- 여러분도 직접 데이터 분석을 수행할 수 있는 능력을 키워야 한다.
- 혹은 적어도 다른 사람의 데이터 분석을 비판적으로 읽어낼 수 있어야 한다.
- 사회통계학은 바로 이러한 목적을 위한 방법이다(물론 유일한 방법은 아니다).



증거에 기반한 사회학 연구

생성형AI는 우리의 학습에 무엇을 의미할까?

- 생성형AI는 여러분을 위해 분석을 대신 해줄 수 있다.
- 그러나 그것이 옳은지 그른지, 그리고 분석 결과를 어떻게 해석할 것인지는 여러분의 몫이다.
- 아예 모르면 생성형AI가 무엇을 할 수 있고 무엇을 할 수 없는지조차 모르기 때문이다 (unknown unknowns).
- 분석 결과를 갖고 다른 사람을 잘 설득하기 위해서, 결국 여러분은 분석 기법을 잘 이해하고 있어야 한다.
- 적어도 무엇을 모르는지 알아야(known unknowns), 생성형AI가 여러분의 조력이 되어줄 수 있다.

