계량분석
Review/Introductory Stata (I)

김현우, PhD[1]

[1] 충북대학교 사회학과 조교수

September 6, 2021

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

진행 순서

1. 선수과목 핵심요약
2. Stata 라이선스, 구동 및 설정
3. 나의 첫 Stata 명령어들

선수과목 핵심요약

선수과목 핵심요약

사회통계(학부 기초통계)에서는 다음의 내용을 다룬다.

- 강의소개(제1주차)
- 기술통계(제2주차)
- 확률론(제3주차)
- 이산/연속확률분포(제4주차)
- 표집(제5주차)
- 통계추정(제6주차)
- 가설검정(제7주차)
- 중간시험(제8주차)
- 평균비교(제9주차)
- 분산비교(제10주차)
- 카이제곱분석(제11주차)
- 상관분석(제12주차)
- 회귀분석(제13주차)
- 회귀분석 조금 더(제14주차)
- 기말시험(제15주차)

* 충북대학교 사회학과 2학년 1학기 전공필수 사회통계 기준으로 주차를 나타냄.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

선수과목 핵심요약

제1주차: 강의소개

- 강의소개
- 경험과학이란 무엇인가?
- 자료유형과 척도
- 엑셀의 설치와 기동

제2주차: 기술통계

- 기술통계란 무엇인가?
- 빈도분포표와 그래프
- 데이터의 요약(I): 중심성향
- 데이터의 요약(II): 산포성향

제3주차: 확률이론

- 확률이론의 기초
- 베이즈 정리
- 확률분포
- 누적확률분포

제4주차: 이산확률분포와 연속확률분포

- 이론적 확률분포들
- 대표적인 이산확률분포: 이항분포
- 대표적인 연속확률분포: 정규분포

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

선수과목 핵심요약

제5주차: 표집

- 모수를 통한 모집단의 추정
- 표본분포와 중심극한정리
- 표본평균의 표본분포
- 표본비율의 표본분포

제6주차: 통계추정

- 신뢰구간과 오차범위
- 모평균의 신뢰구간
- 모비율의 신뢰구간

제7주차: 가설검정

- 가설검정의 논리
- 모평균에 대한 가설검정
- 모비율에 대한 가설검정

제8주차: 중간시험

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

선수과목 핵심요약

제9주차: 평균비교

- t-분포와 자유도
- 단일표본과 2표본
- 독립표본 t-검정
- 쌍체표본 t-검정
- 등분산 가정에 관한 코멘트

제10주차: 분산비교

- 분산분석의 논리
- F-값과 t-값의 관계
- F-분포와 자유도

제11주차: 카이제곱분석

- 교차표
- 카이제곱-분포
- 기대값과 관찰값의 차이

제12주차: 상관분석

- 공분산과 상관계수
- 상관계수는 무엇이고 무엇이 아닌가
- 상관계수 그래프

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

선수과목 핵심요약

제13주차: 회귀분석

- 오차의 제곱의 최소화
- 회귀계수는 무엇이고 무엇이 아닌가
- 추정치의 표준오차
- 모형적합도
- 결과 보고

제14주차: 회귀분석 조금 더

- 다중회귀분석
- 제곱항과 그 해석
- 상호작용항과 그 해석
- 경로분석의 기초
- 회귀분석 가정에 관한 코멘트

제15주차: 기말시험

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

Stata 라이선스, 구동 및 설정

Stata 구동 및 화면 설정

Post Tenebras Lux.

- Stata에도 다양한 버젼(IC, SE, MP 등)이 있지만 일단 신경쓰지 말고 학교 것을 사용할 것.
- 나중에 기관에서 구매를 추진할 일이 있으면 (최소한 SE나) MP를 사도록 유도할 것.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

Stata 구동 및 화면 설정

구동과 화면 설정

- 먼저 여러 windows의 이름과 용도를 파악하면 두려울 것이 없다.
- 편하게 windows를 옮겨서 자신에게 맞도록 설정하자.
- 오래 바라보다 보면 눈알이 터질 것 같으므로 화면 색깔과 폰트 혹은 폰트 사이즈를 변경하자.
- 메뉴도 있다. 의외로 이걸 모르는 사람이 많은데 Stata는 SPSS처럼 point-and-click 분석이 가능하다.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

나의 첫 Stata 명령어들

나의 첫 Stata 명령어들

잘 모를 때는 도움(help)이 필요하다.

- help
- help help

때때로 업데이트(update)를 해야 한다.

- help update

이제 여길 벗어나고(exit) 싶으면,

- exit
- help exit

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

나의 첫 Stata 명령어들

대체 do file이란 무엇이고 왜 중요한가?

- 사실 메뉴에서 point-and-click하거나 명령어 창(command line)을 사용하는 것은 재현(replication)하기 몹시 불편하다.
- do file은 한줄씩 순서대로 실행되는 배치 스크립트(batch scripts)이다.
- 일단 한 번 짜놓으면 실행시킬 때마다 매번 같은 작업을 수행하게 된다.

메뉴와 command line보다는 do file을 적극 활용하자.

- 조금 귀찮을지도 모르지만 특히 기록을 남기는 편이 좋다.
- do file의 백업(backup)을 날짜별로 틈틈이 보관하며 폴더 관리도 체계적으로 해두자.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

나의 첫 Stata 명령어들

일단 Stata를 공부하기로 한 이상, 깔끔한 do 파일을 만드는게 무엇보다 남는 일!

- do 파일에서 코드를 짤 때는 일부러 들여쓰기(indentation) 할 것을 추천!
- 각주달기(annotation)를 철저히! 자신을 단기기억상실증 환자인 것처럼 전제해야.
- 최대한 보기 좋게 꾸며야 한다. 타인의 코드를 읽거나 읽혀야 할 일이 있고, 또 수 주/개월/년 뒤 자신의 코드를 되돌아볼 때도 있다(e.g, R&R 등).

언제나, 언제나, 언제나 자신의 코드를 백업하자!

- 백업을 안해서 후회하는 경우가 생각보다 많다. 해서 후회하지는 않는다.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

나의 첫 Stata 명령어들

언제나 do file의 내용은 심미적으로 잘 관리하자.

- 자신이 그 안에 써넣은 명령어를 모두 기억할 수 있다고 믿어서는 안된다!
- 여백의 미를 충분히 활용할 것!
- 레이블(label)이란 컴퓨터에게는 의미없지만 사람에게는 중요한 노트다.
- 한 줄의 레이블은 * 뒤에, 여러 줄의 레이블은 /* 과 */ 사이에 기록해 둔다.
- 명령어(command)나 command 뭉치마다 레이블을 달아서 자신의 의도를 전달한다.

실행(run)의 단축키는 Ctrl-D이다.

- 전체를 실행시킬 수도 있지만 하이라이트한 부분만 실행할 수도 있다.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

나의 첫 Stata 명령어들

때때로 do file과는 별개로 분석의 결과물(results; outputs)을 기록해 둘 필요가 있다.

- 사실 do file과 data file을 직접 공유하는게 최선일 때가 많다.
- 그러나 보안상 이유로 데이터를 공유할 수 없지만 결과물은 보여줘야 할 때가 있다.
- 때로는 분석 자체에 너무 긴 시간이 걸려 로그(log) 파일을 보여주는 편이 빠를 때도 있다.

로그(log)를 남기려면 log를 활용하자.

- help log
- log using mylog (이때 mylog 대신 다른 파일 이름을 줄 수 있다)
- log close
- log using mylog, replace

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

나의 첫 Stata 명령어들

긴 여행을 떠나기 전에 의식을 치르자.

- 먼저 do file에 나의 의도에 관한 레이블을 기록하자.
- 로그(log)를 켜기 위해 log using mylog를 입력하자.

Stata로 한 번 데이터 파일을 불러보자!

- 먼저 내가 어느 폴더에 있는지 확인하기 위해 cd를 입력하자. cd는 change directory의 머릿글자다.
- 이 폴더에 있는 파일 목록을 살펴보기 위해 dir를 입력하자.
- 특정한 폴더로 이동하기 위해서 cd [folder name]를 활용하자.
- 마음에 드는 파일이 있다면 use [file name]를 통해 불러오자.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

나의 첫 Stata 명령어들

마음에 드는 파일이 없다면 일단 급한대로 적당히 공짜 예제나 보자.

- 당장은 아무런 데이터 파일도 가지고 있지 않으니 차라리 Stata에서 제공하는 예제 파일들 중 하나를 보는 편이 좋을 것 같다.
- 그렇다면 webuse를 사용할 수 있다.
- auto.dta 데이터는 클래식한 미국 자동차 목록을 약간 보여준다. webuse auto를 입력하자.

익숙치 않은 데이터 파일의 시작은 늘 주마간산(走馬看山)이 좋다.

- 일단 edit이라고 쳐보자. 여기서 검은색과 빨간색, 파란색은 각각 의미가 있다.
- 변수 이름을 뒤에 덧붙이면 그것들만 특정해서 볼 수 있다(e.g. edit make weight).
- 변수(variables)가 제법 많아서 이를 간단히 정리된 표로 보고 싶으니 describe를 입력하자.
- 매번 화면이 넘어가기 귀찮으니 set more off, permanently 라고 아예 못을 박아두자.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

나의 첫 Stata 명령어들

변수들을 화면에 나열(list)할 수도 있다.

- list를 입력할 때 화면에 나온 결과는 edit을 통해 본 것과 동일하다.
- 그냥 list를 입력할 수도 있고 변수를 특정할 수도 있다(e.g., list rep78).
- list make in 10/20 같은 명령어는 make 변수의 10에서 20번까지의 관찰값(observations)을 보여준다.

이제 준비되었다. 새로운 변수를 창조하자!

- 새로운 변수를 한 번 만들어 보자. mpg와 headroom을 더한 변수를 만들면 어떨까? generate newvar=mpg+headroom를 입력하자.
- 변수 이름을 바꾸기 위해서는 rename newvar newvar2를 입력하자. newvar는 예전 이름, newvar2는 새 이름이다.
- 새로 만든 변수는 흑역사가 되기 전에 얼른 지우자. drop newvar2를 입력하자.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

나의 첫 Stata 명령어들

아무래도 창조는 쉽지 않으니 기존에 있던 변수를 스리슬쩍 바꿔보자.

- 거 잘보니 rep78에 빈 칸이 조금 보인다. 이 (점이 찍혀 있는) 빈 칸을 결측치 (missing value)라고 부른다.
- 이 결측치 대신 0을 입력하면 뭔가 좀 멋진 것 같다. replace rep78=0 if rep78==. 를 입력하자.

위의 명령어(command)를 잘 보니 =과 ==이 다르다.

- =는 값을 대입(assign)하도록 유도하고, ==는 값이 조건부(conditional)인가를 살펴본다. 이것은 많은 프로그래밍 언어의 공통된 규준이다.

또 색다른 부분은 if 부분이다.

- 이것은 조건문(conditional statement)이라고 불리운다.
- replace rep78=0 는 모든 관측치(observations)에 대해 rep78 변수의 값을 0으로 싹 도배해 버리지만, replace rep78=0 if rep78==. 는 오로지 rep78가 결측치인 조건 아래에서만 rep78을 0으로 바꾼다.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

나의 첫 Stata 명령어들

근데 다시 생각해보니 외제차라면 설령 응답을 안했더라도 수리 한 번 정도는 했을 것 같다.

- 막상 foreign 변수를 보면 파란색, 즉 레이블(label)로 표시되어 있어 실제 입력값을 알기 어렵다.
- 그러므로 command line으로 돌아와 ed, nolabel을 입력하자.

여기서 잘 보면 ed와 nolabel 사이에 쉼표가 붙어있다.

- 쉼표 뒤에 붙어있는 것을 옵션(option)이라고 부른다. , nolabel은 옵션인데 제법 폭넓게 쓰인다.
- 이제 외제차는 1로, 국내차는 0으로 입력된 사실을 깨달았으니 replace rep78=1 if rep78==0 & foreign==1라고 입력하자.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

나의 첫 Stata 명령어들

변수 조작(recoding)은 매우 중요한 문제로 여러분은 이를 능수능란하게 수행할 수 있어야 한다.

예제 1

- length와 weight을 곱한 새로운 변수를 만들자. 이름은 volume으로 하자.

예제 2

- price가 10,000을 넘는 차종을 나타내는 새로운 변수를 만들자. 이름은 luxury가 좋겠다. price가 10,000보다 높으면 1, 그와 같거나 낮으면 0으로 하자.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

나의 첫 Stata 명령어들

첫날부터 제법 고생했으니 이제 슬슬 집에 갈 시간이다.

컴퓨터를 끄기 전에 do file과 데이터 파일을 저장해야 한다.

- 먼저 로그(log)를 닫기 위해 log close 라고 입력하자.
- save [file name]명령어를 통해 데이터 파일을 저장하자.
- 이때 , replace 옵션을 써서 기존에 있던 같은 이름의 파일을 덮어씌울 수 있다.
- 이제 exit 이라고 입력하자.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY