

Sensorveiledning SOS4020

Årets eksamen er en mappeinnlevering som består av to komponenter – et selvstendig empirisk arbeid (Arbeid 1, 4000 ord) og en artikkel som kommenterer en publisert artikkel (Arbeid 2, 2000 ord). Avvik fra ordgrensene på inntil 10% godtas uten negative konsekvenser for karakteren; ordtelling gjøres separat for de to arbeidene. Oppgaveteksten finnes her:

<http://www.uio.no/studier/emner/sv/iss/SOS4020/retningslinjer-for-emneoppgaven.html>

I tillegg til selve besvarelsen leverer studentene log-fil og do-fil (begge for Arbeid 1).

Arbeid 1

Arbeid 1 teller 70%.

Arbeid 1 er et selvstendig empirisk arbeid knyttet til selvvalgt problemstilling. Studentene har levert fem obligatoriske innleveringer som de har fått kommentarer/veiledning på, og som er ment å danne grunnlaget for Arbeid 1. Det er likevel ikke et *krav* at de tidligere innleveringene skal brukes. Arbeid 1 skal framstå som en helhetlig tekst, ikke som to eller flere «sammenklippede» arbeider.

Opgaveteksten for de obligatoriske innleveringene er vedlagt siden den gir en pekepinn på hva som har blitt vektlagt på seminarene.

Studentene har tilgang til tre datasett som alle kan brukes både som tverrsnitt og panel. Datasettene er tilrettelagte versjoner av to amerikanske og ett norsk paneldatasett (PSID, NLSY og NorLAG).

Dokumentasjon av NLSY og NorLAG er gjort tilgjengelig for studentene og er vedlagt her. Når det gjelder PSID, har studentene blitt henvist til undersøkelsens egne nettsider. Samtidig har det også blitt brukt «psid vardoc» (psidtools, fra ssc) i Stata til å slå opp variabler.

Studentene har fått beskjed om å benytte to regresjonsteknikker i oppgaven: En OLS-modell og en mer komplisert teknikk som logit- eller fasteffektanalyse (FE). Studenter som bruker FE, kan velge å gjennomføre hele oppgaven på paneldatasett (altså en pooled OLS og deretter en FE-analyse), eller de kan sammenlikne en OLS fra tverrsnittsdata med FE på paneldata (eller begge deler). Det kan være et pluss hvis BE-estimerer også trekkes inn og kommenteres på en fornuftig måte.

Det er meningen at Arbeid 1 skal være selvforklarende – altså at det ikke skal være nødvendig for sensor å sjekke log- eller do-fil for å forstå hva som er gjort. Log-filen fungerer likevel som en ekstra dokumentasjon på hvordan valg mht. koding og ekskludering av missing etc har fungert. Noen ganger kan pene, håndlagete tabeller skjule underliggende feil i regresjoner, og det kan være nyttig å se log fra originalkjøringene for å vurdere arbeidet. Do-filen skal sensor kun trenge å bruke unntaksvis, hvis det faktisk er tvil om hvordan resultatene i log-filen kan ha kommet fram. (Hvis sensor har behov for selve datasettene, kan disse fås ved henvendelse til instituttet. En taushetserklæring må i så fall underskrives.)

Viktige momenter ved vurderingen av oppgavene

- Det bør være et klart forskningsspørsmål som det er mulig å besvare på en god måte med dataene studentene har tilgjengelig. Man må likevel ta i betraktning av valg av tema gjøres helt i starten av kurset (siden det er lagt opp til at studentene skal bygge på de obligatoriske innleveringene), før f eks problemer med identifisering av kausalsammenhenger er gjennomgått. Mindre gunstige valg av forskningsspørsmål kan i stor grad kompenseres ved gode drøftinger av de begrensninger og problemer dette medfører. Det er fint hvis

forskningsspørsmålet relateres til teori eller tidligere forskning, men også mer common sense-orienterte begrunnelser kan være greit. Man bør videre ta i betraktning at kurset også har studenter fra andre fag enn sosiologi.

- Studentene bør gjøre fornuftige valg mht. avgrensning og bearbeiding av datasettet. Relevant deskriptiv statistikk bør presenteres.
- Regresjonsmodellen bør gjøre et klart skille mellom kontrollvariabler og forklaringsvariabel, der kontroll for bakenforliggende variabler legges til for å rense ut spuriøsitet i sammenhengen mellom forklaringsvariabel og utfall. Det godtas også at mellomliggende variabler (i stedet eller i tillegg) tas inn for å skille mellom direkte og indirekte effekter av forklaringsvariabelen. En klar forståelse av hva man oppnår med å legge til kontrollvariablene også her viktig. Det er en stor styrke hvis relevante problemer med uobserverte konfunderende variabler eller endogen seleksjonsskjevhet (hvis fokus på indirekte effekter) tas opp.
- Det bør være et motivert valg av funksjonsform (særlig for forklaringsvariabel, men empirisk undersøkelse eller tester av funksjonsform for kontrollvariabler er også positivt). Følgende er gjennomgått: lineær sammenheng, polynomer, dummyspesifikasjon, logaritmisk omkodning av avhengig variabel.
- Studenten bør vise forståelse for forskjeller mellom estimerte verdier og generalisering av disse (ved signifikanstester og/eller konfidensintervaller). I forståelsen av tester/konfidensintervaller kan enten en utvalgsbasert eller en modellbasert tilnærming til statistisk inferens legges til grunn. Et motivert valg av tilnærming er selvsagt positivt.
- Studenten bør gjøre adekvat bruk av relevante statistiske tester. I tillegg til t- eller z-tester av enkeltkoeffisienter bør simultane tester av flere koeffisienter (F-test, LR-test eller Wald-test) benyttes der dette er relevant, for eksempel for sett av dummyvariabler eller interaksjonsledd. Test av hele modellen mot en nullmodell uten uavhengige variabler er også positivt men ikke sentralt. BIC og AIC kan gjerne benyttes, men er ikke gjennomgått på forelesningene. Generelt er det aller viktigste at kandidaten viser god forståelse av hvordan resultatene av testene kan tolkes.
- For studenter som bruker logit-modeller, er det viktig at det er gode tekniske og substansielle tolkninger av enten oddsrater eller marginaleffekter. Adekvat forståelse av begge typer tolkninger teller imidlertid positivt. Det kan også telle positivt at kandidaten er klar over problemer med å tolke forskjeller i oddsrater (eller koeffisienter) mellom modeller, men dette er ikke en del av pensum (selv om det er omtalt i forelesning) og manglende behandling av dette skal slik ikke trekke ned.
- Det er positivt at kandidatene viser forståelse av begrensninger ved bruk av kontrollvariabler (og faste effekter hvis dette brukes) for å trekke slutninger om årsakssammenhenger eller eventuelt skille mellom direkte og indirekte effekter. Merk at det ikke kreves på noe karakternivå at en evner å identifisere kausaleffekter på en tilfredsstillende måte, men gode diskusjon av begrensninger ved analysene vil være en vesentlig styrke.
- Det bør vektlegges om det framgår klart hvordan data er bearbeidet og hvordan forskjellige analyser er spesifisert.

Arbeid 2

Arbeid 2 teller 30%.

I arbeid 2 skal studentene kommentere en publisert artikkel som bruker en av metodene som dekkes av pensum, men ikke gjennomgås på seminarene. Artikkelen skal være publisert i et tidsskrift på nivå 1 eller 2. Studentene har fått beskjed om at de kan undersøke om tidsskriftet de har valgt, oppfyller kriteriet her: <https://dbh.nsd.uib.no/publiseringskanaler/Forside>.

I Retningslinjene står det også at tidsskriftet skal være definert som «samfunnsvitenskapelig» av NSD. Det er ikke presisert hvilke fagfelt dette skal omfatte og man bør være liberal her, også slik at man godtar artikler i tidsskrifter som vanskelig kan defineres som samfunnsvitenskapelige, dersom artikkelen selv kan sies å ha et samfunnsvitenskapelig relevant tema.

På seminarene er følgende metoder gjennomgått:

- OLS
- Binær logistisk regresjon
- Regresjon (OLS) med faste individeffekter (FE)

Det vil si at enhver regresjonsbasert analyse som ikke er en av disse (noe som bør begrunnes hvis det ikke er opplagt), kan brukes i Arbeid 2 (dersom den er på pensum). Sentrale eksempler er (ikke komplett liste):

- Instrumentvariabler (IV)
- Regresjon-diskontinuitet (RD)
- Forskjell-i-forskjell (DID)
- Flernivåanalyse (ikke FE)
- Analyse av paneldata med Random Effects (RE)
- Eksperiment
- Ordinal logistisk regresjon
- Multinomisk logistisk regresjon

Det er ikke spesifisert i retningslinjene at artikler på pensum ikke kan benyttes, men det er nevnt på forelesningene at det er uheldig (og det bør også være selvsagt). Hvis noen mot formodning skulle ha valgt å kommentere en pensumartikkel, bør vurderingen av besvarelsen hovedsakelig baseres på selvstendige kommentarer (dvs. det som går utover det som er tatt opp på forelesningene). Forelesninger som har gjennomgått artikler, vedlegges.

Dette er en krevende oppgave, og for mange kandidater vil det å vise forståelse av metoden og hvordan den er anvendt være det som er realistisk. For «et godt arbeid» er det også et krav å «drøfte[r] styrker og begrensninger ved både regresjonsteknikken og den konkrete empiriske anvendelsen». Kurset legger i hovedsak opp til substansielle innvendinger («er det dokumentert at pre-reform-trenden i kontroll- og intervensjonsgruppa er lik?»). Innvendinger av mer teknisk art («her er standardfeilene klynget på feil nivå») går utover det som forventes på kurset.

Vedlegg:

- Dokumentasjon av tilrettelagte datasett (2 dokumenter)

- Obligatoriske oppgaver SOS4020 2020V (1 dokument)
- Forelesninger der pensumartikler gjennomgås (3 dokumenter)

Sensor kan ved forespørsel få tilgang til datasettene som er brukt av SOS4020-studentene.