

# UTSATT SKOLEEKSAMEN I

## SOS4020 - KVANTITATIV METODE

**8. april 2016**

**(4 timer)**

Tillatte hjelpemidler:           Alle skriftlige hjelpemidler og kalkulator.

Sensur for eksamen faller 29. april klokken 14.00. Sensuren publiseres i Studentweb cirka klokken 15.00 samme dag.

Sensuren regnes som mottatt av studentene når den publiseres i Studentweb. Vi minner om at kandidater som ønsker begrunnelse må be om dette **senest 1 uke etter sensur er falt**. Frist for å klage på karakteren er tre uker etter sensur er falt. Informasjon om rutiner for begrunnelser og klager ved Institutt for sosiologi og samfunnsgeografi finnes på emnesiden

Oppgavesettet består av **4 sider** inkludert denne.

Kandidaten skal levere både originalen og kopien av besvarelsen. Kladd skal ikke leveres.

NB! Skriv hardt nok til at kopien blir leselig. Tusj penn kan ikke brukes.

Husk å notere deg kandidatnummeret ditt.

**LYKKE TIL!**

OPPGAVESETTET BESTÅR AV 13 DELOPPGAVER. HUSK Å BESVARE ALLE OPPGAVENE!

SENSORER HAR IKKE TILGANG TIL DINE SKRIFTLIGE HJELPEMIDLER. FØLGELIG BØR DU IKKE HENVISE TIL DISSE.

En forsker ønsker å undersøke hvordan det å ha bli siktet for lovbrudd påvirker lønnsinntekt. Hun får tilgang til data fra administrative registre for alle norske menn og kvinner født i Norge i 1980, og som har vært bosatt sammenhengende i Norge til og med 2015. Dataene inneholder både informasjon om antall siktelser hvert år fra 20-34 års alder, og lønnsinntekt hvert år fra 20-35 års alder. Til sammen er det 56 864 personer i utvalget, hvorav 9 437 har minst en siktelse ved 34 års alder.

Gjennomsnittlig lønnsinntekt ved 35 års alder (standardavvik i parentes) er 303 416 (20 481) for de som har minst en siktelse ved 34 års alder, og 340 801 (10 213) for de som ikke har noen siktelser.

1. Konstruer et 99% konfidensintervaller rundt hvert av de to gjennomsnittene.
2. Er forskjellen mellom de to gjennomsnittene statistisk signifikant? Gjør en t-test. Gi en substansiell tolkning av resultatet av t-testen.
3. Skyldes forskjellen i gjennomsnitt mellom de to gruppene (utelukkende) at siktelser påvirker lønnsinntekt? Diskuter.

For å bedre kunne sammenlikne likt med likt, estimerer forskeren en lineær regresjonsmodell (OLS) med kontrollvariable. Lønnsinntekt ved 35 års alder er utfallsvariabel, og antall siktelser ved 34 års alder er forklaringsvariabel (Modell 1). Modellen bygges trinnvis, og er vist i tabell 1.

Modell 1a inneholder kun antall siktelser (*SIKTELSE*), en kontinuerlig variabel med verdier fra 0 til 15.<sup>1</sup>

I Modell 1b legges det i tillegg til informasjon om respondentens kjønn (*KJONN*), som en dummyvariabel der 1=kvinne og 0=mann.

I modell 1c inkluderes i tillegg variabelen *FORELDRE\_KRIM*, en dikotom variabel som tar 1 hvis minst en av foreldene har vært fengslet før respondenten fylte 18, ellers 0.

---

<sup>1</sup>For personer med flere enn 15 siktelser settes antall siktelser til 15.

Table 1: Lineær regresjonsanalyse med lønnsinntekt ved 35 års alder som avhengig variabel. Standardfeil i parantes.

	Modell 1a	Modell 1b	Modell 1c
Konstantledd	400 112 (1 225)	420 928 (752)	425 523 (791)
<i>SIKTELSE</i> R	-12 295 (2 482)	-15 342 (1 932)	-13 132 (1 424)
<i>KJONN</i> (ref=Mann)			
Kvinne		- 49 435 (967)	- 58 392 (842)
<i>FORELDRE_KRIM</i> (ref=Nei)			
Ja			-28 135 (3 952)
N	56 864	56 864	56 864
$R^2$	0.2	0.15	0.16

- Gjør en signifikanstest av konstantleddet og estimatet for *SIKTELSE*R i Modell 1a. Gi en substansiell fortolkning av begge.
- Forklar kort hvordan variabelen *FORELDRE\_KRIM* kan konfundere sammenhengen mellom respondentens siktelsler og lønnsinntekt. Tegn gjerne et kausaldiagram.
- Hvordan endrer estimatet for *SIKTELSE*R seg når variabelen *KJONN* legges til? Gi en substansiell fortolkning.
- Gitt sammenhengen mellom *KJONN* og lønnsinntekt i Modell 1b, og endringen i estimatet for *SIKTELSE*R når variabelen for *KJONN* legges til, hvilken retning må sammenhengen mellom *KJONN* og *SIKTELSE*R ha?
- Beskriv hvordan modellens forklaringskraft endrer seg etter hvert som flere kontrollvariable legges til.
- Av modell 1a, 1b og 1c, hvilken kommer nærmest å estimere effekten av siktelsler på lønnsinntekt? Begrunn svaret.

Table 2: Lineær regresjonsanalyse med lønnsinntekt som avhengig variabel. Individfasteffekter. Standardfeil i parantes.

<b>Modell 2</b>	
<i>SIKT_OKNING</i>	-1 214 (206)
N (personår)	141 555

For å komme nærmere et svar på om siktelser har en effekt på lønnsinntekt, bestemmer forskeren seg for å estimere en individfasteffekt-modell, der endring i lønnsinntekt (sammenliknet med året før) er avhengig variabel. For å estimere denne modellen begrenser forskeren utvalget til menn og kvinner som har minst en siktelse ved 34 års alder (totalt 9 437 personer). Basert på informasjon om disse personene lager forskeren et paneldatasett. I paneldatasettet inngår det en observasjon av hver person for hvert år fra de er 21 år til og med 35 år (til sammen 15 person-år per person). Paneldatasettet består av til sammen 141 555 person-år (15 personår per person  $\times$  9 437 personer).

For hvert personår måles endring i lønnsinntekt ved alder  $a$  (relativt til året før), og antall nye siktelser ved alder  $a-1$ . Den uavhengige variabelen er *SIKT\_OKNING* er altså antallet nye siktelser året før lønnsinntekt ble målt. For et personår der lønnsinntekt er målt ved 33 års alder, gir variabelen *SIKT\_OKNING* antall nye siktelser det året personen fylte 32. Resultater fra fasteffektmodellen er vist i tabell 2.

10. Gjør en signifikanstest av estimatet for *SIKT\_OKNING* i Modell 2. Gi en substansiell tolkning.
11. Hva kan vi lære av å sammenlikne estimatet for *SIKTELSER* i Modell 1c med estimatet for *SIKT\_OKNING* i Modell 2?
12. Hvorfor kan variablene *KJONN* og *FORELDRE\_KRIM* ikke inkluderes i en modell med individfasteffekter?
13. Under hvilke forutsetning vil variabelen *SIKTELSER* fange opp den kausale effekten av siktelser på lønnsinntekt?