

SKOLEEKSAMEN I

SOS4020 - KVANTITATIV METODE

29. februar 2016

(4 timer)

Tillatte hjelpemidler: Alle skriftlige hjelpemidler og kalkulator.

Sensur for eksamen faller 21. mars 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 **6 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 12 DELOPPGAVER. HUSK Å BESVARE ALLE OPPGAVENE!

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

I forbindelse med undersøkelsen NORFERT besvarte 3025 kvinner et spørreskjema i 2015. Kvinnene er bosatt i Norge i 2015, og født i Norge enten 1960 eller 1965.¹ Gjennomsnittlig antall barn blant kvinnene i utvalget er 1.93. Variansen rundt gjennomsnittet er 0.07.

1. Konstruer et 95% konfidensintervall rundt gjennomsnittet. Gi en substansiell fortolkning av konfidensintervallet.
2. Hva vil de si at estimatoren for gjennomsnitt er forventningsrett?
3. I samme utvalg er andelen trebarnsmødre 0.33 blant kvinner som har fullført høyere utdanning, og 0.38 blant kvinner som ikke har fullført høyere utdanning. 1204 av kvinnene har fullført høyere utdanning. Tyder disse utvalgsdataene på at andelen trebarnsmødre varierer med utdanningsnivå i populasjonen?

En forsker ønsker å undersøke om utdanningsnivå påvirker antall barn kvinner får i løpet av livet ved å bruke NORFERT-dataene. Forskeren estimerer en regresjonsmodell der antall barn er avhengig variabel, og utdanningsnivå er uavhengig variabel. Resultatene vises i Tabell 1. Modell 1a i tabell 1 viser sammenhengen mellom utdanningslengde i år (*UTDAAR*) og antall barn.

4. Prediker gjennomsnittlig antall barn for en kvinne med ni års utdanning (grunnskole) fra Modell 1a.
5. Figur 1 viser et plott av antall års utdanning og antall barn. Hvilke standardfeil bør forskeren benytte? Begrunn svaret.

I modell 1b (tabell 1) legger forskeren til kontroller for to mulige bakenforliggende utelatte variable.

¹Utypisk nok har alle kvinnene som sender inn spørreskjemaet besvart alle spørsmål i skjemaet.

Figure 1: Scatterplott av antall års utdanning og antall barn. Data fra spørreundersøkelsen NORFERT, norskfødte kvinner født 1960 og 1965.

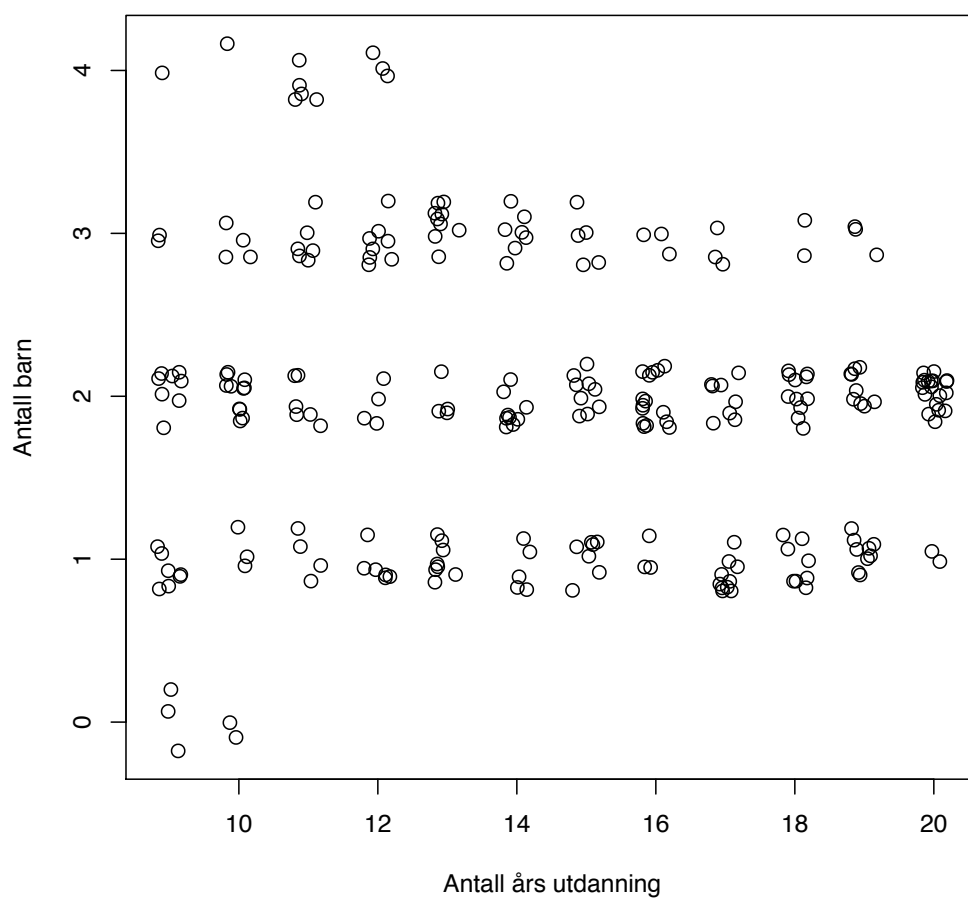


Table 1: Lineær regresjonsanalyse med antall barn som avhengig variabel. Standardfeil i parentes.

	Modell 1a	Modell 1b
Konstantledd	2.32 (0.04)	2.42 (0.03)
UTDAAR	-0.03 (0.01)	-0.02 (0.01)
FAAR (ref=1960)		
1965		-0.05 (0.02)
MORUTD (ref= Vgs./grunnskole)		
Høyere utdanning		-0.11 (0.06)
N	3025	3025
R^2	0.04	0.07
SSE	290	163

- *FAAR*: Dummyvariabel for kvinnens fødselsår. Kodet 0 for kvinner født i 1960, og 1 for kvinner født i 1965.
 - *MORUTD*: En dummyvariabel for kvinnens mors høyeste fullførte utdanningsnivå. Kodet 1 for høyere utdanning, 0 for grunnskole eller videregående skole.
6. Gi en substansiell fortolkning av endringen i parameterestimatet for utdanningsnivå når kontrollvariablene legges til.
 7. Forbedres modellens forklaringskraft signifikant når kontrollvariablene legges til? Gjør en F-test av modell 1a mot modell 1b.
 8. NORFERT inkluderer et spørsmål om timelønn. Selv om lønn henger sammen med utdanningsnivå, og kanskje også påvirker antall barn, har forskeren valgt å ikke inkludere lønn som kontrollvariabel. Skisser en motiasjon for å ikke inkludere lønn som forklaringsvariabel. (Tegn gjerne opp et kausaldiagram.)
 9. Gir estimatene i modell 1b den kausale effekten av utdanningsnivå på antall barn? Hvorfor/hvorfor ikke?

Table 2: Lineær regresjonsanalyse med antall barn som avhengig variabel. Standardfeil i parantes.

	Modell 2
Konstantledd	2.32 (0.04)
REFORMKOMM (ref=Nei)	
Ja	0.10 (0.02)
FAAR (ref=1960)	
1965	-0.12 (0.04)
FAAR* REFORMKOMM	-0.08 (0.03)
N	3025

På grunn av bekymring for lavt utdanningsnivå i distriktene, vedtok Stortinget i 1964 at alle barn født 1965 eller seinere skulle ha rett til tre års høyere utdanning med 100% stipend fra Lånekassen, hvis og bare hvis de var født i en distriktskommune. Tidligere studier har vist at skolereformen av 1964 virket – andelen med høyere utdanning vokste raskere blant unge født i reformkommuner enn blant unge født i andre kommuner. Forskeren bestemmer seg for å utnytte skolereformen av 1964 til en forskjell-i-forskjells-analyse. Tabell 2 inneholder resultatene fra en slik analyse. De inkluderte variablene i Modell 2 er:

- *REFORMKOMM*: Dummyvariabel som tar 1 for kvinner født i en reformkommune, 0 kvinner som er født i en kommune som ikke er omfattet av reformen.
 - *FAAR*: Dummyvariabel som tar 1 for kvinner født i 1965, 0 for kvinner født i 1960.
 - En interaksjon mellom *FAAR* og *REFORMKOMM* (*FAAR*REFORMKOMM*).
10. Hvilke kvinner får verdien 1 på variabelen *FAAR*REFORMKOMM*? Gi en substansiell tolkning av parameterestimatet for denne variabelen.

11. Er effekten av $FAAR*REFORMKOMM$ statistisk signifikant? Gjør en t-test.
12. Under hvilke forutsetninger kan vi tolke parameterestimatet for $FAAR*REFORMKOMM$ som effekten av økt andel høyere utdannede kvinner på fruktbarhet? Er det noen trusler mot kausal identifikasjon i denne modellen?

LYKKE TIL!