

# **SENSORVEILEDNING FOR SKOLEEKSAMEN I**

## **SOS4020 - KVANTITATIV METODE**

**11. mars 2015**

**(4 timer)**

Tillatte hjelpemidler:           Alle skriftlige hjelpemidler og kalkulator.

Sensur for eksamen faller 7. 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 10 DELOPPGAVER. HUSK Å BESVARE ALLE OPPGAVENE!**

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

En forskergruppe arbeider med problemstillinger knyttet til innvandring og integrering. De samlet i år 2010 inn data om et enkelt tilfeldig utvalg av norskfødte personer i yrkesaktiv alder fra befolkningen i Norge. Datasettet inneholder en rekke variabler:

- *Holdning* er respondentens holdning til innvandrere. Den er hentet fra en faktoranalyse av ulike spørsmål om innvandring og innvandrere. Variabelen har verdier fra 0 til 4, hvor 0 angir en sterk negativ holdning og 4 angir en sterk positiv holdning.
- *Alder* er en kontinuerlig variabel, omkodet slik at nullpunktet er 20-åringer.
- *Urban* er en dummyvariabel som er kodet 1 om respondenten bor i et urbant område, 0 ellers.
- *Utdanning* er en dummyvariabel som er 1 om respondentens har utdanning utover videregående skole.
- *Kjønn* er kodet 0 for kvinner og 1 for menn.

Tabell 1 viser resultater fra flere regresjonsanalyser hvor *Holdning* er avhengig variabel. Alder, utdanning og sysselsetting er uavhengige variabler i modellen. Det er tatt med et annengradsledd for alder.

**Tabell 1. Regresjonsmodeller av holdning til innvandrere i 2010. N = 1023.**

	<i>Modell 1</i>		<i>Modell 2</i>	
	<i>Lin. Regresjon</i>		<i>Lin. Regresjon</i>	
	B	SE	B	SE
Konstant	2,4	0,2	2,2	0,2
Utdanning	0,15	0,05	0,10	0,02
Alder			0,02	0,01
Alder—kvadrert			-0,001	0,0002
Urban			0,3	0,01
$R^2$	19%		25%	

- a) Gi substansielle tolkninger av konstantleddet og regresjonskoeffisienten i modell 1.

En substansiell tolkning forteller oss hva resultatene sier om verden, ikke om regresjonskoeffisienter. Konstantleddet i modell 1 sier at norskfødte i yrkesaktiv alder bosatt i Norge i 2010 uten utdanning utover grunnskole i gjennomsnittet skårer 2,4 på

variabelen holdning til innvandrere. De med utdanning utover grunnskole skårer noe høyere enn de uten slik utdanning. Noen vil også påpeke at forskjellen er signifikant ( $t = 3$ ).

- b) Et medlem i forskergruppen hevder at en persons holdning til innvandrere endres om vedkommende tar høyere utdanning. Gir modell 1 grunnlag for en slik konklusjon?

Første skritt her vil være å teste (om det ikke allerede er gjort) om forskjellen er målbar forskjellig fra null. Neste skritt vil være å evaluere den kausale tolkningen av koeffisienten som påstanden innebærer. Det kan gjøres på flere måter. En måte er å påpeke at det ikke foreligger longitudinelle data, og at påstanden handler om en prosess. En annen måte er å argumentere med at det ikke er tilfeldig hvem som har høyere utdanning og at det er sannsynlig at det finnes en korrelasjon mellom X og restleddet. Svaret er selvfølgelig nei.

- c) I modell 2 er det kontrollert for alder. Beskriv kort aldersmønsteret i holdning til innvandrere.

Dette er en grei, halvt teknisk oppgave hvor studenten skal vise at de forstår hva et annengradsledd innebærer. En mulig løsning er å begynne med nullpunktet på Alder-variabelen (20-åringer). Forventet nivå på holdning til innvandrere er for 20-åringer (uten utdanning og bosatt i rurale strøk) er 2,2. Med økende alder blir holdningen til innvandrere mer positiv, men endringen per år blir mindre frem mot 40 års alder (ved  $x=20$ ).  $X = 20$  er med andre ord et toppunkt for aldersmønsteret. Her snur utviklingen og med økende alder blir holdningen til innvandrere stadig mindre positiv.

- d) Konstruer et konfidensintervall rundt koeffisienten for variabelen *Urban* fra modell 2. Hva forteller dette intervallet oss?

Et konfidensintervall rundt koeffisienten for variabelen *Urban* kan lages ved å velge konfidensgrad og beregne øvre og nedre grenser:  $b \pm t \cdot SE(b)$

Med 95% konfidens blir  $t$  1,96 (det er svært mange frihetsgrader) og grensene vil da bli  $[0,3 - 0,0196 \quad 0,3 + 0,0196]$  som er  $[0,28 \quad 0,32]$ . Intervallet har følgende tolkning: Det sanne regresjonskoeffisienten for *Urban* i modell 2 ligger med 95% sannsynlighet innenfor intervallet  $[0,28 \quad 0,32]$ . Siden *Urban* er en dummyvariabel forteller intervallet oss at forskjellen i holdning til innvandrere mellom de som bor i rurale strøk og de som bor i urbane strøk ligger innenfor intervallet, når det er kontrollert for alder og utdanning.

Forskerne har tidligere år gjennomført sammenlignbare undersøkelser av den samme populasjonen. For å studere endringer over tid i holdninger til innvandrere slår de sammen datasettene for en rekke undersøkelser de siste tiårene.

En ny variabel, *År*, forteller hvilket år observasjonen er hentet fra. Den har verdien 0 for året 1985 og øker med 1 for hvert år. (Den har dog ikke verdier for alle årene mellom 1985 og 2010).

Tabell 2 rapporterer en regresjonsanalyse av det samlede datasettet. Forskerne er interessert i om forskjellen mellom sosiale grupper øker eller minker i løpet av perioden de studerer. I modellene er både *Urban* og *Utdanning* med, og det er samspill mellom variabelen *År* og disse to uavhengige variablene. I tillegg er det kontrollert for alder og kjønn i modellene.

**Tabell 2. Regresjonsmodeller av sosial endring i holdning til innvandrere. N = 5219.**

	B	SE
Konstant	1,5	0,1
År	0,05	0,02
Utdanning	0,60	0,15
Utdanning*År	-0,025	0,01
Urban	-0,1	0,1
Urban*År	0,02	0,01

NB: I tillegg er det kontrollert for alder og kjønn i modellene.

- e) Hvem utgjør referansegruppen i modellen i Tabell 2 og hvordan er utviklingen over tid i holdning til innvandrere for denne gruppen?

Referansegruppen er de med null på alle variable. Denne gruppen er de som i 1985 ikke har utdanning utover grunnskole og bor i rurale strøk. Utviklingen over tid for de som ikke har utdanning utover grunnskole og bor i rurale strøk kan leses av koeffisienten for år. Koeffisienten er 0,05 (SE 0,02;  $p < 0.05$ ) er innebærer at for hvert år som går etter 1985 blir holdningene til innvandrere i denne gruppen noe mer positive.

- f) Forskergruppen konkluderer med at den viktigste skillelinjen i holdning til innvandrere over tid har gått fra å være sosial (etter utdanningsnivå) til å bli geografisk (etter bosted). Har vedkommende rett i dette? Begrunn svaret.

Denne oppgaven krever forståelse av samspill med T i repeterte tverrsnittsundersøkelser. Den er vanskelig og gode forsøk bør premieres.

Forskergruppens påstand handler om hvilken dimensjon som representerer «den viktigste skillelinjen i holdning til innvandrere». La oss si man her tenker på «viktig skillelinje» som å innebære «størst forskjell i holdning til innvandrere». Da kan man se på utviklingen over tid i forskjellen på de med utdanning utover grunnskolen og de uten utdanning utover grunnskolen og den tilsvarende utviklingen over tid i forskjellen på de som bor i rurale strøk og de som bor i urbane strøk.

I 1985 er forskjellen på utdanning==1 og utdanning==0 0,60. For hvert år som går blir denne forskjellen noe mindre. Koeffisienten for samspill mellom år og utdanning er estimert til -0,025. Fra 1985(år == 0) til 2010 (år == 25) har det gått 25 år. I 2010 vil forskjellen på de med og utdanning utover grunnskole være  $0,6+(25*-0,025) = -0,025$ .

Tilsvarende beregning gjøres også for forskjellen etter urban/rural bostedsstrøk. I 1985 er forskjellen kun -0,1, hvilket betyr at de som bor i urbane strøk har så vidt mer negative holdninger til innvandrere enn de som bor i rurale strøk. I 2010 er denne forskjellen blitt  $-0,1 + (25 * 0,02) = -1 + 0,5 = 0,4$ . Forskjellen på de som bor i urbane strøk og i rurale strøk er nå 0,4, slik at de som bor i urbane strøk skårer 0,4 poeng høyere på holdning til innvandrere-skalaen (og dermed har en mer positiv holdning til innvandrere).

Fra 1985, hvor utdanningsforskjellen var stor og urban/rural-forskjellen var liten, har situasjonen blitt omvendt. En forskjell på 0,4 er nødvendigvis mye større enn -0,025. Konklusjonen ser ut til å være riktig.

- g) Modellen antar at eventuell endring i befolkningens holdning til innvandrere er den samme hvert år. Om denne antagelsen ikke er rimelig, hvordan kunne modellen vært endret for å unngå å gjøre en slik antagelse?

*Hint: Lineær historieskriving.*

Her spørres det implisitt om linearitetsforutsetningen. All utvikling over tid er tenkt å være lineær i perioden 1985 til 2010. Dette er kan hende en urimelig antagelse (men det kan helt sikkert argumenteres godt for begge standpunkt). Om man vil gjøre noe med den kunne man latt utviklingen følge en annengradskurve eller gjøre om År-variabelen til et sett av dummyvariable. Begge disse utvidelsene vil la forskerne slippe å gjøre denne antakelsen.

Opgaven er uvanlig ettersom konteksten er en endringsmodell og ikke en ordinær OLS-modell av ett tverrsnitt. Den er følgelig ganske vanskelig.

Senere ble undersøkelsen også fulgt opp med en mindre panelundersøkelse. Et mindre utvalg av de som ble intervjuet i 2010 ble intervjuet på nytt i 2015, og de samme variablene som ble målt i 2010 er målt i 2015. Nå er forskerne interessert i å vurdere om endringer i livssituasjon påvirker ens holdning til innvandrere. Tabell 3 rapporterer to separate analyser av paneldataene.

**Tabell 3. Paneldatamodeller av data fra 2010-2015. N = 385**

	<i>Modell 1</i>	<i>Modell 2</i>
	<i>First-difference-modell</i>	<i>First-difference-modell</i>
	<i>MENN</i>	<i>KVINNER</i>

	B	SE	B	SE
Konstant	-0,15	0,1	0,15	0,1
Utdanning	0,25	0,1	0,07	0,12
Urban	0,10	0,3	0,08	0,3

- h) Tolk konstantleddet og koeffisienten for utdanning i Modell 1 (MENN).

Konstantleddet måler endringen fra 2010 til 2015. Dette er en alders- eller periodeeffekt. Alle har fått litt mer negative holdninger til innvandrere i perioden ( $b = -0,15$ ).

Koeffisienten for utdanning er 0,25. Denne innebærer at om man går fra å ikke ha utdanning over grunnskole til å ha utdanning over grunnskole så øker ens holdning til innvandrere (mot mer positive holdninger) med 0,25 poeng, kontrollert for urbanitet og alle stabile kjennetegn ved personene.

- i) Forskergruppen ønsker å publisere et arbeid hvor de konkluderer med at flytting til sentrale strøk påvirker ens holdning til innvandrere. Ville du anbefalt en slik uforbeholden konklusjon? Hvorfor / hvorfor ikke?

Koeffisienten er 0,10 ( $SE = 0,3$ ;  $t > 3$ ;  $p < 5\%$ ) hvilket innebærer at å gå fra ruralt til urbant bosted er assosiert med en 0,10 poengs økning i holdning til innvandrere (i positiv retning). Ved første øyekast kan det se ut som det er en effekt av flytting til sentrale strøk på holdninger til innvandrere.

For å tolke denne koeffisienten som uttrykk for en årsakssammenheng kan det ikke foreligge endringer i øvrige uavhengige variable (stabilitetsforutsetningen) og de uavhengige variablene må ha konstante effekter. Et annet forbehold gjelder årsaksretningen: holdning til innvandrere kan påvirke valg av bosted

Studenten bør ikke anbefale en slik uforbeholden konklusjon.

- j) Et relevant spørsmål er om menn og kvinner, uavhengig av deres utdanning og bosted, har en ulik utvikling over livsløpet i holdninger til innvandrere. Kan resultatene i Tabell 3 brukes til å informere et slikt spørsmål? Hvordan?

Konstantleddet i modellen for menn og modellen for kvinner er hhv. -0,15 og 0,15. Disse parameterestimatene anslår (som beskrevet i oppgave h) hva endringen har vært i den avhengige variabelen fra  $t_1$  (2010) til  $t_2$  (2015).

Endringen for menn har vært -0,15, en liten endring i retning av mer skepsis mot innvandrere. Endringen for kvinner er like stor, men går i motsatt retning.

Disse endringene reflekterer alders- eller periodeeffekter fra 2010 til 2015. Det er umulig å skille disse fra hverandre, men om man har god grunn til å tro at det ikke er periodeeffekter (f.eks. fra andre data, repeterte tverrsnitt o.l.) vil disse endringene representere alderseffekter. Endringer over livsløpet som ikke skyldes endringer i personens situasjon må kunne sies å være alderseffekter.

☺ LYKKE TIL! ☺