

**UTSATT EKSAMEN I  
SOS4020 – KVANTITATIV METODE**

**SKOLEEKSAMEN  
8. januar 2008  
(4 timer)**

Ikke-programmerbar kalkulator er tillatt under eksamen. Ingen andre hjelpemidler er tillatt.

Sensuren faller fredag 29. januar kl. 14.00. Sensuren blir slått opp på tavlen utenfor Aud. 7, 1. etasje i Eilert Sundts hus. Sensuren kan også hentes på Studentweb fra ca kl. 14.30 samme dag.

Sensuren blir regnet som mottatt av studentene når den blir slått opp, med mindre det kan dokumenteres gyldig fravær. Vi minner om at kandidater som ønsker begrunnelse må søke til instituttet om dette **senest 1 uke** etter at sensur er mottatt. Frist for å klage på karakteren er tre uker etter at sensuren har falt.

Oppgaven er på **5 sider** inkludert denne.

Kandidaten skal levere både originalen og kopien av besvarelsen.

Husk å notere deg kandidatnummeret ditt.

**LYKKE TIL!**

**Oppgavesettet består av 14 spørsmål. Alle spørsmålene skal besvares. Det siste arket inneholder formler som kan være til nytte ved besvarelsen av oppgaven.**

**NB: En sensor skal lese kopien av oppgaven din. Pass derfor på at gjennomslagskopien blir tydelig.**

Nedenfor er en rekke regresjonsanalyser hvor den avhengige variabelen er hvor mye boliggjeld husholdningen har. De uavhengige variablene er:

- **Barn**, antall barn i husholdet under 16 år
- **Par**, hvor 1=gifte, samboende eller partnere og 0=enslige
- **Alder**, det vil si hvor gammel den som har høyest inntekt i husholdet er
- **Urbaniseringsgrad**, hvor 1=spredtbygd strøk, 2=tettbygd strøk med 200-1 999 bosatte, 3=tettbygd strøk med 2 000-19 999 innbyggere, 4=tettbygd strøk med 20 000-99 999 innbyggere og 5=tettbygd strøk med 100 000 eller flere bosatte.
- **Husholdsinntekt**, det vil si samlet inntekt i husholdet etter skatt

Det primære siktemålet med analysene er å undersøke betydningen av antall barn i husholdet på boliggjelden.

Spm 1: Utform en årsakshypotese om sammenhengen mellom antall barn i husholdet og boliggjeld og illustrer denne med en årsaksmodell. Gi substansielle begrunnelser for hypotesen.

Spm 2: Hva er forutsetningen for å inkludere variabelen *urbaniseringsgrad* slik som den er kodet ovenfor?

Hvordan ville du kode denne variabelen dersom denne forutsetningen ikke er rimelig?

**Tabell 1: Boliggjeld avhengig av antall barn under 16 år i husholdningen. Ordinær regresjonsanalyse. 2001. N=1051.**

	B	SE(b)
Konstant	255 000	17 775
Barn	108 000	12 000

Justert  $R^2$ : 0,076

Spm 3: Tolk konstant, regresjonskoeffisient og  $R^2$  i tabell 1. Kommenter også statistisk signifikans.

**Tabell 2: Boliggjeld avhengig av antall barn under 16 år i husholdningen, familietype (par vs. enslig), urbaniseringsgrad og husholdningsinntekt. Ordinær regresjonsanalyse. 2001. N=1051.**

	B	SE(b)
Konstant	-54 410	39 680
Barn	86 120	11 700
Par	149 071	31 100
Urbaniseringsgrad	46 815	8 870
Husholdsinntekt	0,20	0,04

Justert R<sup>2</sup>: 0,125

Spm 4: Tolk konstant, regresjonskoeffisienter og R<sup>2</sup> i tabell 2. Kommenter også statistisk signifikans.

**Tabell 3: Boliggjeld avhengig av antall barn under 16 år i husholdningen, familietype (par vs. enslig), urbaniseringsgrad, husholdningsinntekt og alder. Ordinær regresjonsanalyse. 2001. N=1051.**

	B	SE(b)
Konstant	-245 420	135 825
Barn	45 870	12 780
Par	156 580	30 460
Urbaniseringsgrad	40 715	8 750
Husholdsinntekt	0,17	0,04
Alder	18 050	5 850
Alder kvadrert	-252	60

Justert R<sup>2</sup>: 0,181

Spm 5: Gjør rede for sammenhengen mellom alder og boliggjeld basert på analysen i tabell 3. Kommenter også statistisk signifikans.

Gi en substansiell fortolkning av sammenhengen.

Gjør rede for en annen måte å modellere denne sammenhengen på.

Spm 6: Gjør rede for og forklar endringene fra tabell 2 til tabell 3 når det gjelder variabelen barn.

Spm 7: Avgjør ut fra analysen i tabell 3 hvilke sosiale grupper som har henholdsvis minst og mest boliggjeld.

Gjør rede mekanismene som påvirker boliggjeld ut fra analysen.

**Tabell 4: Boliggjeld avhengig av antall barn under 16 år i husholdningen, familietype (par vs. enslig), urbaniseringsgrad, husholdningsinntekt og alder. Ordinær regresjonsanalyse. 2001. N=1051.**

	B	SE(b)
Konstant	-236 230	136 290
Barn	21 115	32 180
Par	141 500	35 380
Urbaniseringsgrad	40 745	8 750
Husholdsinntekt	0,16	0,04
Alder	18 180	5 855
Alder kvadrert	-254	61
Barn * Par	28 300	33 755

Justert R<sup>2</sup>: 0,181

I tabell 4 er det inkludert en samspillsvariabel (produktledd) mellom variablene *par* og *barn*.

Spm 8: Vis hvordan man konstruerer en slik samspillsvariabel.

Gjør rede for hvorfor man inkluderer en samspillsvariabel i en regresjonsanalyse.

Spm 9: Utform en forskningshypotese om samspill knyttet til samspillsanalysen i tabell 4. Gi substansielle begrunnelser for hypotesen.

Tolk koeffisientene for variablene *par*, *barn* og *barn\*par*.

Gir analysen støtte for at det er samspill? Begrunn svaret.

Spm 10: Hva er intern validitet? Pek på momenter av betydning for den interne validiteten knyttet til analysene i tabellene ovenfor.

Spm 11: Gjør rede for hvilket problem som kan oppstå når vi bruker tidsserier til å teste hypoteser om årsakssammenhenger.

Gjør også rede for hvordan dette problemet kan utbedres.

Tabell 5 viser resultatet av en logistisk regresjonsanalyse hvor den avhengige variabelen boliggjeld er kodet om til en dummyvariabel med verdien 0 dersom husholdningen ikke har boliggjeld og verdien 1 dersom husholdningen har boliggjeld.

**Tabell 5: Logistisk regresjon av boliggjeld avhengig av antall barn under 16 år i husholdningen og familiestatus. 2001. N=1051.**

	B	S.E.	Wald	Exp(B)
Konstant	-0,5	0,10	25	0,6
Barn	0,6	0,06	100	1,8
Par	0,6	0,15	16	1,8

Spm 12: Tolk odds og oddsratio for konstanten og variablene *barn* og *par*.

Spm 13: Beregn sannsynlighet for å ha boliggjeld for en enslig uten barn og et gift par med to barn.

Hosmer-Lemeshow testen for analysen i tabell 5 har signifikanssannsynlighet lik 0,006.

Spm 14: Forklar hva vi mener med signifikanssannsynlighet.

Gjør rede for hva Hosmer-Lemeshow testen går ut på.

Sett opp hypoteser og konkluder ut fra testen.

Hvordan kan et utilfredstillende resultat av Hosmer-Lemeshow testen håndteres?

**Formler:**

T-test for regresjonskoeffisient:

$$t = \frac{\hat{b}_1}{SE(\hat{b}_1)}$$

Beregning av sannsynlighet basert på odds:

$$\tilde{Y} = \frac{e^{(b_0+b_1 \cdot X)}}{1 + e^{(b_0+b_1 \cdot X)}} = \frac{Odds(\tilde{Y})}{1 + Odds(\tilde{Y})}$$

**LYKKE TIL!**