

**EKSAMEN I  
SOS4020  
KVANTITATIV METODE  
20. mars 2009**

**(4 timer)**

Tillatte hjelpemidler: Ikke-programmerbar kalkulator

Liste med matematiske uttrykk/andeler i fordelinger (bakerst i oppgavesettet)

Sensur på eksamen faller onsdag 14. april 2009. Sensuren kan hentes på Studentweb fra ca kl. 14.30 samme dag.

Sensuren regnes som mottatt av studentene når den offentliggjøres. Vi minner om at kandidater som ønsker begrunnelse må søke om dette til instituttet senest 1 uke etter at sensur har falt. Klagefristen er tre uker fra sensuren er falt, eller tre uker fra begrunnelse er mottatt.

Oppgavesettet består av **5 sider** inkludert denne.

Kandidaten skal levere både originalen og kopien av besvarelsen.

Husk å notere deg kandidatnummeret ditt.

**LYKKE TIL!**

**Oppgavesettet består av 11 delspørsmål. 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.**

Det er gjort en spørreundersøkelse blant kvinner som ved intervju tidspunktet var samboende eller gift. Utvalget var et enkelt sannsynlighetsutvalg av befolkningen i den aktuelle aldersgruppen. Utvalgets størrelse var  $N=857$ . Formålet med undersøkelsen er å kartlegge sammenhenger mellom graden av forpliktelse individer har til sine samliv og utvalgte andre faktorer.

I undersøkelsen inngikk en rekke spørsmål som omhandlet forpliktelse og samliv. Fra disse spørsmålene ble det så konstruert en indeks som skal være et mål på graden av forpliktelse personen føler for det aktuelle samlivet vedkommende er i. Indeksen kan ta verdier fra 0, som representerer ingen forpliktelse, til 21, som representerer en svært høy grad av forpliktelse. I tillegg til indeksvariabelen er det målt flere andre variable. Disse er:

- *Familiebakgrunn* er en dummyvariabel hvor verdien 0 betyr at kvinnens foreldre er gift, og verdien 1 betyr at kvinnens foreldre ikke er gift.
- *Samlivstatus* er en dummyvariabel hvor verdien 0 betyr at kvinnen er gift og 1 betyr at personen lever i et samboerforhold.
- *Antall barn* er en kategorisk variabel som har tre verdier:
  - ”A” som indikerer at kvinnen ikke har noen barn med sin nåværende partner.
  - ”B” indikerer at kvinnen har 1 eller 2 barn med sin nåværende partner.
  - ”C” som indikerer at kvinnen har 3 eller flere barn med sin nåværende partner.
- *Varighet* som er en kontinuerlig variabel som måler antall år kvinnen har levd i sitt nåværende samliv. Om hun for eksempel flyttet sammen med sin partner for nøyaktig tre år siden har variabelen verdien 3. Om hun flyttet sammen med vedkommende for 10 år siden er verdien 10.

a) Beskriv kort forskjellen på et longitudinelt design og et tverrsnittsdesign. Hva er fordelene og ulempene med de to typene design?

Resultatene fra tre regresjonsmodeller hvor forpliktelsesindeksen er den avhengige variabelen er gjengitt i tabell 1 (under).

**Tabell 1. Resultater fra tre regresjonsanalyser av forpliktelse til samliv.**

Parameter	Modell 1		Modell 2		Modell 3	
	B	SE(b)	B	SE(b)	b	SE(b)
Konstant	18,2	0,25	17,7	0,40	14,8	0,89
Familiebakgrunn	-2,72	0,40	-0,93	0,72	-0,66	0,89
Samlivstatus			-1,93	0,45	-1,25	0,48
Antall barn ( <i>Referansegruppe er 0</i> )						
<i>Dummyvariabel for 1-2 barn</i>			2,22	0,53	1,75	0,43
<i>Dummyvariabel for 3+ barn</i>			2,51	0,61	2,03	0,64
Forholdets varighet					0,56	0,21
Forholdets varighet kvadrert					-0,04	0,01
R <sup>2</sup>	0,04		0,17		0,22	

- b) I modell 1 er det kun den uavhengige variabelen familiebakgrunn med i regresjonsmodellen. Er regresjonskoeffisienten for denne variabelen statistisk signifikant? Vis signifikanstest. Hva betyr det at en regresjonskoeffisient er statistisk signifikant? Gi en substansiell tolkning av koeffisienten.
- c) Hva er et konfidensintervall? Et 99% konfidensintervall for regresjonskoeffisienten for familiebakgrunn i modell 1 går fra -3,75 til -1,69. Gi en substansiell tolkning av konfidensintervallet.
- d) I modell 2 er ytterligere to variable tatt med i modellen. På grunnlag av resultatene i modell 2, vil du si at vi kan trekke følgende konklusjon: *"Det er en signifikant forskjell mellom kvinner som har 1-2 barn og kvinner som har 3 barn i graden av forpliktelse de føler for sitt samliv."* Begrunn svaret.
- e) Anta at vi lurer på om sammenhengen mellom samlivsstatus og forpliktelse er sterkere for de som ikke har barn. Gir modell 2 et tilfredstillende svar på dette spørsmålet? Hvis nei, hvilke forbedringer ville du foreslått?
- f) I modell 3 er også forholdets varighet tatt med i analysen, både som en ordinær uavhengig variabel og som et annengradsledd. Hva er hensikten med å ha med både varighet og varighet kvadrert i analysen?
- g) På grunnlag av resultatene i modell 3: Beregn predikert (forventet) verdi på den avhengige variabelen for kvinner hvis nåværende samliv har vart i henholdsvis 2, 5 og 10 år. Ta utgangspunkt i at kvinnene tilhører referansegruppen på de andre variablene. Illustrer med en figur sammenhengen mellom varighet og forpliktelse. Hvordan kan denne sammenhengen forklares?
- h) Er det grunnlag for å si at forskjellen mellom samboende og gifte kvinner som kommer til uttrykk i koeffisienten for samlivsstatus er en *kausal* sammenheng? Forklar hvorfor/hvorfor ikke.

Det er også gjort en logistisk regresjonsanalyse med variabelen *Vurdert samlivsbrudd* som avhengig variabel. Det er to uavhengige variable i modellen, samt et samspillsledd mellom de. Definisjonene av den avhengige og de uavhengige variablene i modellen er som følger:

- *Kvinnens sosiale bakgrunn* er en dummyvariabel hvor 0 representerer lavere sosiale sjikt og 1 representerer høyere sosiale sjikt.
- *Partnerens sosiale bakgrunn* er en dummyvariabel som er målt med samme verdier som kvinnens egen sosiale bakgrunn.
- *Vurdert samlivsbrudd* er en dummyvariabel. Den har verdien 1 om kvinnen oppgir at hun har vurdert å gå fra partneren sin det siste året, og 0 ellers.

**Tabell 2. Resultater fra logistisk regresjonsanalyse av å ha vurdert samlivsbrudd siste år.**

Parameter	b	SE(b)	Wald	Sig.	Exp(b)
Konstant	-1.0	0,10	100	0,001	0,37
Kvinnens sosiale bakgrunn	0,30	0,05	36	0,001	1,35
Partnerens sosiale bakgrunn	0,30	0,05	36	0,001	1,35
Kvinnens sosiale bakgrunn*Partnerens sosiale bakgrunn	-0,20	0,01	4	0,046	0,82

- Hva er referansegruppen i den logistiske regresjonsanalysen i tabell 2? Hvilken odds har de for å ha vurdert å gå fra partneren sin det siste året?
- Beregn oddsen og sannsynligheten for at kvinner som tilhører det lave sosiale sjiktet og lever med en partner fra det øvre sosiale sjiktet skal ha vurdert å gå fra hverandre.
- Hvilke grupper har størst og minst sannsynlighet for å ha vurdert samlivsbrudd? Kan du gi en substansiell forklaring på dette mønsteret?

**Liste over matematiske uttrykk som kan være nyttige**

$$t = \frac{\hat{m}}{SE(\hat{m})}$$

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

$$t = \frac{\hat{m}_1 - \hat{m}_2}{SE(\hat{m}_1 - \hat{m}_2)}$$

$$t = \frac{\hat{b}_1 - \hat{b}_2}{SE(\hat{b}_1 - \hat{b}_2)}$$

$$SE(\hat{m}) = \frac{s}{\sqrt{N}} \approx \frac{\hat{s}}{\sqrt{N}}$$

$$SE(\hat{m}_1 - \hat{m}_2) = \sqrt{[SE(\hat{m}_1)]^2 + [SE(\hat{m}_2)]^2}$$

$$SE(\hat{b}_1 - \hat{b}_2) = \sqrt{[SE(\hat{b}_1)]^2 + [SE(\hat{b}_2)]^2}$$

$$K.I.: \hat{m} \pm t \cdot SE(\hat{m})$$

$$K.I.: \hat{b} \pm t \cdot SE(\hat{b})$$

$$Andel = \frac{Odds}{1 + Odds}$$

$$Odds = \frac{p}{1 - p}$$

**Kritiske verdier i statistiske fordelinger:**

Normalfordelingen / Student's t  
med svært mange frihetsgrader:

<u>Andel</u>	<u>z/t</u>
95%	±1,96
99%	±2,58

Kji-kvadratfordelingen (med en frihetsgrad):

<u>Andel</u>	<u><math>\chi^2</math></u>
95%	3,84
99%	6,63