

**EKSAMEN I
SOS4020 KVANTITATIV METODE
(MASTER)
14. MAI 2004
(4 timer)**

Bruk av ikke-programmerbar kalkulator er tillatt under eksamen. Utover det er ingen hjelpemidler tillatt.

Sensur faller mandag 7. juni kl. 14.00, se oppslag på tavla utenfor Aud. 7, Eilert Sundts hus. Sensuren kan også hentes på Studentweb fra ca. kl. 14.30 samme dag.

Sensuren regnes som mottatt av studenten når den blir hengt opp, dersom ikke gyldig fravær kan dokumenteres. Vi minner om at kandidater som ønsker begrunnelse må søke om dette til instituttet senest 1 uke etter sensur har falt. Frist for å klage på karakteren er tre uker etter sensur har falt.

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

Kandidaten skal levere både originalen og kopien av besvarelsen.

Husk å skrive ned kandidatnummeret ditt et sted så du finner det igjen.

LYKKE TIL!!!

Alle oppgaver skal besvares

I en ungdomsundersøkelse har en blant annet stilt en del spørsmål som skal belyse ensomhetsproblemer og sosial isolasjon blant informantene. En har på grunnlag av disse spørsmålene konstruert en indeks ved å summere skårene på flere spørsmål som belyser dette fenomenet. Vi skal derfor oppfatte denne indeksen som et mål på grad av sosial isolasjon. Den har verdier som går fra 0 til 9, og lave verdier betyr liten grad av isolasjon, mens høye verdier betyr stor grad av sosial isolasjon. Utvalgets størrelse er på $N=1420$ personer og ungdommene er i alderen 13-18 år.

Gjennomsnittlig indeksverdi i utvalget er 2.77, standardavviket er lik 2.08 og standardfeilen til gjennomsnittet er 0.055. Utvalgets fordeling på indeksen er gjengitt i tabell 1.

Tabell 1. Fordeling (prosent) av utvalget på en indeks som måler sosial isolasjon.

Indeksverdi	Prosent
0	14.0
1	17.9
2	19.6
3	15.5
4	12.8
5	7.6
6	6.5
7	4.4
8	1.1
9	0.6
Totalt	100.0

a) Gi en kort beskrivelse av denne utvalgsfordelingen. Avviker den fra normalfordelingen? (Lag gjerne en skisse.) Konstruer et 95 prosents konfidensintervall for gjennomsnittet i utvalget, og forklar hva et konfidensintervall er.

b) Ovennevnte indeks er basert på informantenes svar på tre spørsmål. Informantene er bedt om å ta standpunkt til følgende tre utsagn, der svaralternativene er "aldri", "sjelden", "av og til" og "ofte":

- Jeg har følelsen av at ingen kjenner meg særlig godt
- Jeg føler meg ensom
- Jeg savner å ha en god venn

I forbindelse med indekskonstruksjonen har en undersøkt i hvilken grad det er korrelasjon mellom informantenes svar på spørsmålene. En krysstabell for to av spørsmålene er gjengitt i tabell 2.

Tabell 2. Sammenheng mellom svar på to spørsmål om sosial isolasjon. Antall personer.

Savner en god venn	Føler seg ensom				Totalt
	Aldri (0)	Sjelden (1)	Av og til (2)	Ofte (3)	
Aldri (0)	354	292	68	4	718
Sjelden (1)	80	176	84	8	348
Av og til (2)	30	118	104	20	272
Ofte (3)	6	18	40	18	82
Totalt	470	604	296	50	1420

En har beregnet kjikvadratobservatoren for denne tabellen, og fått $kji\text{-kvadrat} = 364.7$. Hva måler denne observatoren og hva er den grunnleggende logikken bak dette målet? Hvor mange frihetsgrader er det i dette tilfellet? På 5 prosentnivået er kritisk $kji\text{-kvadratverdi}$ 16.9. Hva betyr dette, og hva blir konklusjonen av denne testen?

c) Når man koder svarene aldri = 0, sjelden = 1, av og til = 2 og ofte = 3 kan man beregne korrelasjonskoeffisienten for de to variablene "savner en god venn" og "føler seg ensom". Denne korrelasjonskoeffisienten blir 0.47. Standardfeilen til korrelasjonskoeffisienten kan beregnes ut fra formelen

$$SE(\hat{r}) = \frac{1}{\sqrt{N-1}}$$

Bruk dette til å teste nullhypotesen om at korrelasjonskoeffisienten er lik null. Velg signifikansnivå 5 prosent.

d) Da ovennevnte indeks ble konstruert, ble alle de tre ovennevnte variablene kodet som angitt. Indeksen er lik summen av de tre variablene. Cronbachs alpha-koeffisient for denne indeksen er 0.71. Hva er alpha-koeffisientene et mål på? Hvis man hadde konstruert indeksen på grunnlag av flere beslektede spørsmål, skulle man da vente at alpha-koeffisienten hadde blitt annerledes? Begrunn svaret.

e) I tabell 3 gjøres det en sammenligning av gutter og jenter med hensyn til verdier på nevnte indeks for sosial isolasjon. Beskriv forskjellen mellom de to kjønn. Er forskjellen i gjennomsnittsverdi statistisk signifikant på 5 prosentnivået? Standardfeilen til differansen mellom to gjennomsnitt kan beregnes ut fra formelen

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

Tabell 3. Kjønnforskjeller på indeksen som måler sosial isolasjon

	Gjennomsnitt	Standardavvik	Standardfeil
Jenter	3.08	2.14	0.114
Gutter	2.45	1.97	0.105

I tabell 4 gjengis regresjonsanalyser med indeksen for sosial isolasjon som avhengig variabel. I de to første modellene er gutter og jenter analysert hver for seg, mens hele materialet er analysert under ett i de tre siste modellene.

- Variabelen **alder** er kodet med 13 år som nullpunkt, slik at 13-åringer har verdien 0, 14-åringer verdien 1, osv.
- Variabelen **kjønn** er kodet med verdien 1 for jenter og 0 for gutter.
- Variabelen **kjønn x alder** er produktet av de to angjeldende variablene.
- Dernest er det en dummyvariabel som angir om informanten **bor sammen med begge foreldre** (kodet 1) eller ikke (kodet 0).
- Dessuten er det konstruert en indeks som måler hvor god **foreldrekontakt** informanten har. Denne variabelen har fire forskjellige verdier, der verdien 0 indikerer dårlig kontakt og største verdi 3 indikerer meget god kontakt. Mellomliggende verdier indikerer mellomliggende grader av kontakt.

Tabell 4. Regresjonsanalyser med sosial isolasjonsindeks som avhengig variabel

	Bare jenter Modell #1		Bare gutter Modell #2		Begge kjønn Modell #3		Begge kjønn Modell #4		Begge kjønn Modell #5	
	b	SE(b)	b	SE(b)	b	SE(b)	b	SE(b)	b	SE(b)
Konstant	3.38	0.13	2.28	0.14	2.28	0.15	2.49	0.18	3.52	0.225
Alder (over 13)	-0.120	0.057	0.096	0.057	0.096	0.060	0.084	0.060	0.043	0.061
Kjønn (jenter =1)	-	-	-	-	1.005	0.195	0.988	0.195	1.146	0.198
Kjønn x Alder	-	-	-	-	-0.215	0.081	-0.214	0.081	-0.257	0.082
Bor med begge foreldre (ja =1)	-	-	-	-	-	-	-0.259	0.106	-0.023	0.108
Foreldrekontakt	-	-	-	-	-	-	-	-	-0.541	0.066
Justert R ²	0.005		0.003		0.023		0.025		0.086	

f) I modell #1 og #2 har en undersøkt om det er alderforskjeller med hensyn til sosial isolasjon hos de to kjønn. Gi en fortolkning av analyseresultat med særlig vekt på kjønnsforskjellene. Er forskjellen mellom regresjonskoeffisientene for alder hos gutter og jenter statistisk signifikant på 5 prosentnivået? (Hint: Standardfeilen til differansen mellom to regresjonskoeffisienter kan beregnes på samme måte som standardfeilen til differansen mellom to gjennomsnitt, se ovenfor.) Hva forteller resultatet oss?

g) I modell #3 er materialet analysert under ett og en har i tillegg til kjønn og alder tatt med en variabel som er lik produktet av kjønn og alder. Hva er hensikten med dette siste leddet? Forklar hvorledes de tre regresjonskoeffisientene tolkes.

h) Beregn predikert (forventet) isolasjonsindeks for gutter på henholdsvis 13 og 18 år, og for jenter i samme aldergrupper med utgangspunkt i modell #3. Lag på grunnlag av dette et diagram med alder langs den horisontale akse og isolasjonsindeks langs den vertikale akse, som viser regresjonslinjene for jenter og gutter. Gi en karakteristik av sammenhengene.

i) I modell #4 er også dummyvariabelen "bor sammen med begge foreldre" trukket inn i analysen. Hva er en dummyvariabel? Er sammenhengen statistisk signifikant på 5 prosentnivået? Gi en substansiell fortolkning av resultatet.

j) I modell #5 er også indeksen "foreldrekontakt" trukket inn i analysen. Gi en substansiell tolkning av resultatet. Hvordan kan man tolke det forhold at parameterverdien for variabelen "bor med begge foreldre" har endret seg?

k) Analysen i modell #5 kunne også vært gjort ved å representere variabelen "foreldrekontakt" med dummyvariable. Hvor mange dummyvariable ville en da ha konstruert og hva ville fordelene med en slik fremgangsmåte være?