

**EKSAMEN I
SOS4020 – KVANTITATIV METODE**

**SKOLEEKSAMEN
2. november 2007
(4 timer)**

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

Sensuren faller fredag 23. november 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 **6 sider** inkludert denne.

Kandidaten skal levere både originalen og kopien av besvarelsen.

Husk å notere deg kandidatnummeret ditt.

LYKKE TIL!

Oppgavesettet består av 12 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.

Oppgave 1:

Høsten 2005, 2006 og 2007 har det blitt gjennomført en spørreundersøkelse blant studenter som tar bacheloremnet i kvantitativ metode (SOS1120). Disse studentene består i hovedsak av sosiologistudenter (rundt 60 prosent) og kriminologistudenter (rundt 35 prosent). De øvrige studentene kommer hovedsakelig fra andre fagområder på SV-fakultetet. Frafallet er svært lite, det vil si at nesten samtlige av studentene på emnet har besvart undersøkelsen.

Ett av spørsmålene i denne undersøkelsen er:

Tenk tilbake på siste uke, det vil de siste 7 dagene. Hvor mange ganger har du shoppet (vært i butikker og kanskje kjøpt noe)?

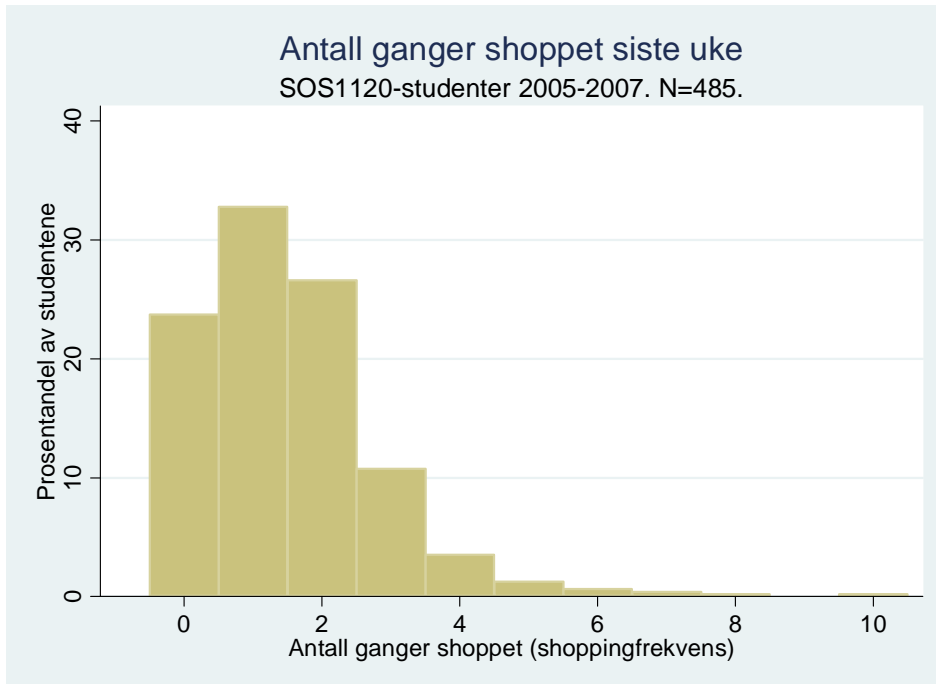
Det er dette spørsmålet som representerer den avhengige variabelen i de videre analysene. Vi vil kalle denne variabelen for *shopping* i oppgaven.

Spm 1: Gjør først generelt rede for seleksjonsproblemet innenfor forskningsmetode. Pek deretter på sannsynlige seleksjonsprosesser som medfører at utvalget ovenfor ikke er et representativt utvalg for populasjonen av samtlige universitetsstudenter i Norge. Trekk inn begrepene *begrepsvaliditet* og *ekstern validitet*.

Vi vil i dette oppgavesettet betrakte det nevnte utvalget som et sannsynlighetsutvalg. Begrunnelsen er at utvalget kan ses på som et tilfeldig utvalg fra populasjonen av (fortidige, nåtidige og framtidige) årskull som tar bacheloremnet SOS1120.

Spm 2: Pek på styrker og svakheter ved denne begrunnelsen.

Fordelingen på variabelen *shopping* framgår av følgende stolpediagram:



Gjennomsnittet på variabelen er 1,49 og standardavviket er 1,34.

Spm 3: Hva viser frekvensfordelingen? Konstruer deretter et 95 prosent konfidensintervall for gjennomsnittet og gjør rede for hvordan dette skal fortolkes.

En sammenlikning mellom mannlige og kvinnelige studenter viser følgende:

Group Statistics

		N	Mean	Std. Deviation	Std. Error Mean
Hvor mange ganger shoppet	Mann	129	1,01	1,093	,096
	Kvinne	351	1,67	1,375	,073

Spm 4: Sett først opp hypoteser og undersøk om den observerte forskjellen skyldes tilfeldigheter eller en reell forskjell i populasjonen. (Levines test av variansen i de to gruppene er statistisk signifikant og tyder på at variansen er ulik i de to gruppene.) Pek deretter på mulige substansielle forklaringer på den observerte forskjellen mellom mannlige og kvinnelige studenters shoppingfrekvens.

Nedenfor vises resultatene fra tre regresjonsanalyser med shopping som avhengig variabel. De uavhengige variablene er:

- *Kjønn* med verdien 0 for mannlige studenter og 1 for kvinnelige studenter.
- *Alder* kodet som dummyvariabler. Referanse kategorien er inntil 20 år.
- *Arbeidstimer* – antall timer lønnet arbeid man har hatt siste uke.
- *Månedslønn* – hvor mye man har hatt i lønnsinntekt siste uke. Målt i antall tusen kroner (1500 kroner vil for eksempel være kodet som 1,5).
- *Venner* – antall ganger man har vært sammen med venner siste uke.
- *Boform* – om man bor hjemme hos foreldre (kodet 0) eller leier/eier bolig (kodet 1).
- *Politisk interesse* – grad av politisk interesse målt på en skala fra 0 (ingen politisk interesse) til 4 (svært politisk interessert)

Regresjonsanalyser av shoppingfrekvens siste uke. SOS1120-studenter 2005-2007.

	Modell 1		Modell 2		Modell 3	
	Koeff.	St.feil	Koeff.	St.feil	Koeff.	St.feil
Kjønn	0,625	0,138	0,656	0,139	1,813	0,475
Alder 20-24 år	0,430	0,416	0,431	0,415	0,444	0,412
Alder 25-29 år	0,512	0,436	0,509	0,435	0,490	0,431
Alder 30 år eller mer	0,552	0,449	0,513	0,448	0,492	0,444
Venner	0,105	0,033	0,106	0,033	0,104	0,033
Boform	-0,491	0,209	-0,521	0,209	-0,480	0,208
Politisk interesse	-0,261	0,073	-0,257	0,073	0,073	0,144
Kjønn*Politisk interesse					-0,423	0,167
Arbeidstid	0,013	0,006	0,004	0,007		
Månedslønn (1000 kr)			0,030	0,011	0,033	0,009
Konstant	1,307	0,485	1,219	0,485	0,279	0,608
N	441		441		441	
R ²	0,108		0,122		0,128	

Spm 5: Fortolk koeffisientene i modell 1. Undersøk også om koeffisientene er statistisk signifikante.

Spm 6: Gir modell 1 støtte for en antakelse om at de kvinnelige studentene har shoppet mer enn de mannlige fordi de har mer omgang med venner? Begrunn svaret.

Spm 7: Gjør rede for endringene for arbeidstidsvariabelen fra modell 1 til modell 2 og forklar endringene. Hvorfor tror du arbeidstidsvariabelen er tatt ut i modell 3?

I modell 3 er det inkludert et interaksjonsledd (samspillsledd) mellom variablene *kjønn* og *politisk interesse*.

Spm 8: Hva er hensikten med å inkludere et interaksjonsledd i regresjonsanalyse? Gir analysen støtte for at det er samspill mellom kjønn og politisk interesse? Gi både en teknisk og substansiell fortolkning av koeffisientene for kjønn, politisk interesse og samspillsvariabelen.

Spm 9: Hva tror du man ønsker å undersøke gjennom å kode aldersvariabelen som dummyvariabler i de tre analysene? Hvilken konklusjon gir analysen i så måte? Gjør rede for eventuelle andre måter å undersøke dette på.

Hvis vi koder om den avhengige variabelen shopping til en dummyvariabel med verdien 0 for dem som ikke har shoppet og verdien 1 for dem som har shoppet minst 1 gang, kan vi undersøke hvilke variabler som påvirker hvorvidt studentene har shoppet siste uke eller ikke. Andelen som har shoppet er 0,6 for de mannlige studentene og 0,8 for de kvinnelige.

Spm 10: Forklar begrepene odds og oddsratio. Beregn oddsen for å shoppe for henholdsvis mannlige og kvinnelige studenter. Beregn også oddsratioen mellom de to gruppene. Hva forteller disse målene i dette tilfellet?

En logistisk regresjonsanalyse mellom den todelte shoppingvariabelen og de uavhengige variablene *kjønn* og *månedslønn* gir følgende resultat.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	Kjønn	1,156	,235	24,291	1	,000	3,177
	Månedslønn	,039	,020	4,059	1	,044	1,040
	Konstant	,198	,214	,852	1	,356	1,218

a. Variable(s) entered on step 1: kjønn_31, månedslønn.

Spm 11: Fortolk oddsen for konstanten og oddsratioene for variablene *kjønn* og *månedslønn*. Beregn deretter sannsynligheten for å ha shoppet for henholdsvis en mannlig og en kvinnelig student som begge har tjent 2 000 kroner forrige måned.

Spm 12: I læreboken hevder Skog at en variabel med en høy oddsratio ikke nødvendigvis er en viktig forklaringsvariabel. Gjør rede for hvorfor dette kan være tilfelle.

Formler:

Standardfeil for et gjennomsnitt: $SE(\hat{m}) = \frac{\hat{s}}{\sqrt{N}}$

Konfidensintervall for et gjennomsnitt: $KI = \hat{m} \pm t \cdot SE(\hat{m})$

Standardfeil for forskjell mellom to gjennomsnitt: $SE(\hat{m}_1 - \hat{m}_2) = \sqrt{[SE(\hat{m}_1)]^2 + [SE(\hat{m}_2)]^2}$

T-test for forskjell mellom to gjennomsnitt: $t = \frac{\hat{m}_1 - \hat{m}_2}{SE(\hat{m}_1 - \hat{m}_2)}$

T-test for regresjonskoeffisient: $t = \frac{\hat{b}_1}{SE(\hat{b}_1)}$

Forholdet mellom sannsynlighet og 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!