

**EKSAMEN I
SOS1120 KVANTITATIV METODE
24. NOVEMBER 2009
(4 timer)**

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

Sensur faller 15. desember 2009 kl. 14.00, og kan hentes på Studentweb ca. en time senere.

Sensuren regnes som mottatt av studenten når den blir publisert.

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 at sensur har falt.

Oppgavesettet består av **4 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!

Oppgavesettet består av tre oppgaver. Alle oppgavene skal besvares og teller like mye.

NB! En sensor skal lese kopien av besvarelsen også. Skriv slik at den blir leselig.

Oppgave 1

Et år ble foreldreklagene i Oslo-skolene undersøkt. Alle foreldre av barn i barne- og ungdomsskolene ble registrert og nummerert fra 1 og oppover. Hver 10'ende forelder fikk en invitasjon til å delta i spørreundersøkelsen, og noe under halvparten sa seg villige til å delta. Foreldrene ble bedt om å svare ja eller nei på om de hadde fremmet skriftlige eller muntlige klager på skolens undervisningstilbud, disiplinærtiltak, fritidstilbud, sikkerhet i forbindelse med barnas skolevei, og/eller lærernes kvalitet og opptreden. Svarene ble til slutt slått sammen til én variabel hvor alle som hadde fremmet minst én klage fikk verdien ”ja” og alle andre verdien ”nei”.

Samtidig ble egenskaper ved foreldrene og området skolene ligger i registrert. Bl.a. ble det skilt mellom foreldre som hadde samlet inntekt over 600.000 kroner — høy inntekt — og de med lavere inntekt. Skolene ble gruppert etter hvorvidt de lå i områder hvor de var en overvekt av foreldre med høy sosial status eller ikke.

1-1) Nevn et eksempel på et sannsynlighetsutvalg og et ikke-sannsynlighetsutvalg. Hva er fordeler og ulemper ved hver av dem? Er utvalget i undersøkelsen som er beskrevet ovenfor et sannsynlighetsutvalg? Begrunn svaret.

Tabell 1 Klager etter skolens statusområde og foreldrenes inntekt. Prosent.

Klage	Høystatusområde		Lavstatusområde		Totalt
	Høy inntekt	Lavere inntekt	Høy inntekt	Lavere inntekt	
Ja	43	30	23	8	30
Nei	57	70	87	92	70
Totalt	100	100	100	100	100
N	300	700	500	1000	2500

1-2) Hva er enheter, verdier, avhengig og uavhengige variable i tabell 1? Tegn en årsaksmodell/ kausaldiagram over sammenhengene mellom variablene. Gi en begrunnelse for hvorfor du plasserer variablene i diagrammet slik du gjør.

1-3) Med utgangspunkt i tabellen er følgende fem påstander fremmet. Noen av dem er korrekte mens andre er gale. Vurder hver enkelt av dem, og begrunn hvorfor du mener de er enten korrekte eller gale.

- ”Foreldre med høy inntekt klager sterkere enn foreldre med lavere inntekt”
- ”Foreldre med lav inntekt er minst skeptiske til skolens undervisningsopplegg”
- ”43% av de som klager har høy inntekt og samtidig barn på skole i et høystatusområde”

- d) ”Skolens statusområde betyr mer for tendensen til å klage enn hvilken inntekt foreldrene har”
- e) ”Inntektens betydning for å klage er større i høystatusområder enn lavstatusområder”

Oppgave 2

I en landsrepresentativ spørreundersøkelse ble respondentene spurt om hvor godt de behersket en rekke ulike datatjenester. Deretter utviklet man et samlet mål for disse ferdighetene kalt ”*datakunnskap*”. Denne variabelen varierer i utgangspunktet mellom 0 og 30, hvor 0 betyr ingen ferdigheter på noen av ferdighetsområdene, mens 30 innebærer at man har maksimale kunnskaper på samtlige felt.

Tabell 2 inneholder en grafisk framstilling og noen nøkkeltall om fordelingen på variabelen ”*datakunnskap*”:

Statistikk	Skåre	Grafisk framstilling av fordelingen
Gjennomsnitt	18,4	
Median	19	
Minimum	0	
Maksimum	28	
Varians	39,06	
Standardavvik (s)	6,25	
Standardfeil ($s_{\bar{x}}$)	0,22	
1. kvartil	14	
2. kvartil	19	
3. kvartil	23	
N	807	

2-1) Gi en kort beskrivelse av fordelingen på variabelen ”*datakunnskap*” med utgangspunkt i histogrammet og utvalgte nøkkeltall i tabell 2. Er variabelen normalfordelt? Hvordan vurderer du det generelle kunnskapsnivået på området? Bruk posisjonsmålene og målene på sentraltendens til å underbygge vurderingen.

2-2) I tabell 2 oppgis varians (s^2) standardavviket (s) og standardfeilen ($s_{\bar{x}}$). Hva slags informasjon gir disse målene?

2-3) Hvorfor er det ofte ønskelig å operere med intervaller istedenfor punkttestimater? Beregn et 95% konfidensintervall for gjennomsnittet på variabelen ”*datakunnskap*”. Hvordan vil du fortolke dette intervallet? Forklar logikken som ligger til grunn for denne fortolkningen.

Oppgave 3

En regresjonsanalyse av fenomenet datakunnskap ble utført for å få mer innsikt i hvordan slik kunnskap fordeler seg i befolkningen. Ved siden av ”datakunnskap” som avhengig variabel ble tre uavhengige variable trukket inn i analysen:

Kjønn: 0 = mann, 1 = kvinne

Alder: 0 = under 60 år, 1 = 60 år eller mer

Utdanning: 0 = lav (grunnskole, videregående), 1 = høy (universitet, høyskole)

Tabell 3: Datakunnskap etter kjønn, alder og utdanning. Lineær regresjonsanalyse. 2009.

	b	St.feil s_b	t	p
Konstant	24,7	.38	65,0	$p < .001$
Kjønn	- 1,2	.40	-3,0	$p < .01$
Alder	- 4,5	.41	- 11,0	$p < .001$
Utdanning	2,4	.40	6,0	$p < .001$

$R^2 = .174$

3-1) Gi en fortolkning av konstantleddet og koeffisientene for kjønn, alder og utdanning. Hva slags informasjon gir R^2 ?

3-2) Hvilken sosial gruppe har lavest datakunnskap ifølge modellen? Hvilken sosial gruppe har mest kunnskap? Beregn kunnskapsnivået for lavutdannede menn over 60 år.

3-3) Tabell 3 rapporterer også signifikanstester for hver av variablene i modellen. Sett opp nullhypotese og alternativ hypotese for testen knyttet til kjønnsvariabelen. Vis testen. Hva betyr det at $t = -3$ og $p < .01$? Hvilken konklusjon kan trekkes på grunnlag av testen?

LYKKE TIL!