

**EKSAMEN I SOSIOLOGI (MASTER)
SOS4020 - KVANTITATIV METODE**

**SKOLEEKSAMEN
11. mai 2005
(4 timer)**

Tillatt hjelpemiddel: Ikke-programmerbar kalkulator.

Oppgavesettet består av 6 sider inkludert denne.

Kandidaten skal levere både originalen og kopien av besvarelsen.

Sensur på eksamen faller innen onsdag 1. juni kl. 14.00. Sensuren slås opp på tavlen utenfor Aud. 7, i 1. etasje i Eilert Sundts hus. Sensuren kan også hentes på studentweb ca. en time senere.

Sensuren regnes som mottatt av studentene når den slås opp, med mindre gyldig fravær kan dokumenteres.

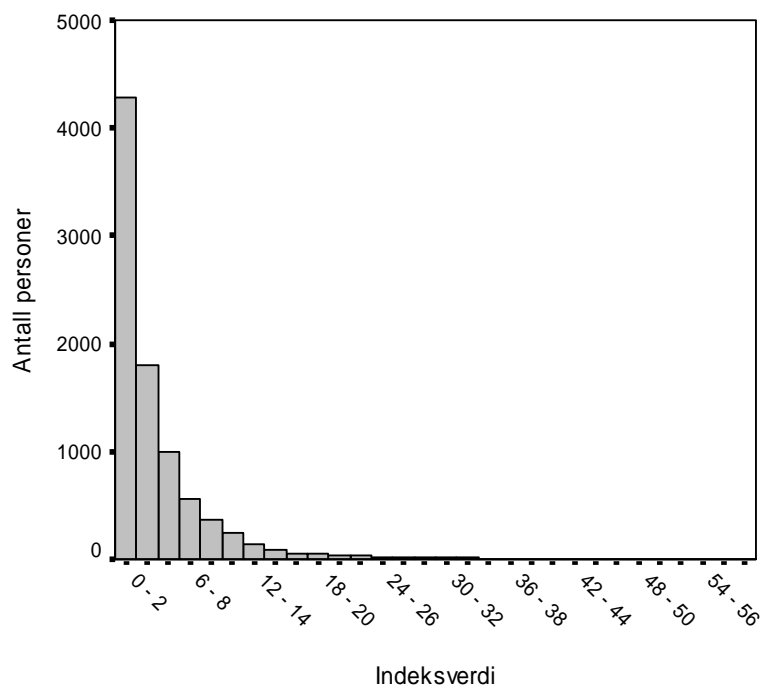
Husk å notere deg kandidatnummeret ditt.

LYKKE TIL!

Alle spørsmål skal besvares og teller like mye.

I en spørreskjemaundersøkelse blant skoleungdom i alderen 13-17 år har en stilt en del spørsmål om ulovlige og asosiale handlinger. På grunnlag av 12 slike spørsmål har en konstruert en indeks som skal måle i hvilken grad svarpersonene har begått slike handlinger. Indeksen går fra 0 til 60 poeng og små verdier betyr at respondenten rapporterer få ulovlige og antisosiale handlinger. Utvalgets størrelse er $N=8777$.

a) Figur 1 viser fordelingen av utvalget etter indeksverdi. Gjennomsnittlig indeksverdi er 3.4. Av utvalget har 35.9% indeksverdi null (ingen antisosiale handlinger) og 12.8% har indeksverdi lik én (kun én slik handling). 7.3% av utvalget har indeksverdi på 10 indekspoeng eller mer. Gi en kort beskrivelse av fordelingsform. Har fordelingsform noen konsekvenser når en skal beregne konfidensintervall for gjennomsnittet?



Figur 1. Fordelingen av respondentene etter indeksverdi.

b) I tabell 1 gjengis gjennomsnittskårer, standardavvik og standardfeil til gjennomsnittet for henholdsvis gutter og jenter. Hva viser resultatet? Still opp en nullhypotese og en alternativ hypotese og test forskjellen mellom gutter og jenters gjennomsnittsskåre med signifikansnivå 5%. (Hint: 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 1. Gjennomsnittlig indeksverdi for henholdsvis gutter og jenter, samt standardavvik og standardfeil til gjennomsnittet.

	Gjennomsnitt	Standardavvik	Standardfeil	N
Gutter	4.08	6.45	0.098	4294
Jenter	2.63	3.93	0.059	4397

c) I tabell 2 gjengis gjennomsnittskårer, standardavvik og standardfeil til gjennomsnittet for ulike aldersgrupper. Hva viser resultatet? Drøft kort mulige substansielle forklaringer på alderforskjellene. Kan det tenkes at resultatet i noen grad har sin forklaring i metodeproblemer?

Tabell 2. Gjennomsnittlig indeksverdi i ulike aldergrupper,, samt standardavvik og standardfeil til gjennomsnittet.

Alder i år	Gjennomsnitt	Standardavvik	Standardfeil	N
13	2.98	4.92	0.129	1456
14	3.80	5.93	0.141	1766
15	3.85	5.74	0.134	1839
16	3.15	5.03	0.120	1767
17	2.95	5.06	0.115	1949

d) En ønsker å undersøke om forskjellene mellom de 5 aldersgruppene kan være et resultat av tilfeldigheter og vil derfor teste en nullhypotese som sier at alle aldersgrupper egentlig har samme gjennomsnitt. I tabell 3 gjengis resultatene av Fishers F-test. Hva blir konklusjonen på testen? Forklar kort hva de viktigste størrelsene i tabellen betyr.

Tabell 3. Variansanalyse og F-test av alderforskjellene i tabell 2.

	Kvadratsum (sum of squares)	<i>df</i>	Varians (mean square)	<i>F</i>	<i>p</i>
Mellom grupper	1388.8	4	347.2	12.06	<0.001
Innenfor grupper	252614.1	8772	28.8		
Totalt	254002.9	8776			

Det er også gjort lineære regresjonsanalyser med indeksen for antisosial atferd som avhengig variabel. I tillegg til kjønn og alder har en villet undersøke om forekomsten av antisosial atferd er avhengig av hjemstedets størrelse og om informanten bor med begge foreldre. Dessuten ønsker man å undersøke om andregenerasjons innvandrerungdom avviker fra øvrig ungdom med hensyn til antisosial atferd. I regresjonsanalysene nedenfor er variablene kodet som følger:

Kjønn: gutt = 0, jente = 1

Alder: Antall år over 13 (dvs. 13 år som nullpunkt)

Alder kvadrert = Variabelen alder multiplisert med seg selv

Bor med begge foreldre: Nei = 0, Ja = 1

Hjemstedets størrelse: Kodet fra 1 til 6, der storby (over 100 000 innbyggere) er kodet 1, mellomstor by er kodet 2, osv. ned til bygd eller grend kodet 6

Andregenerasjons innvandrer: Kodet 1 hvis far ikke er født i Norge, mens informanten selv er født i Norge, ellers kodet 0.

e) I modell nr. 1 er kjønn, alder og alder kvadrert inkludert i en lineær regresjonsanalyse. Forklar hva analysen viser og sammenlign resultatet med tabellanalysen ovenfor.

f) I modell nr. 2 og 3 er gutter og jenter analysert hver for seg, og en har i tillegg tatt med variablene Bor med begge foreldre, Hjemstedets størrelse og Andregenerasjons innvandrer. Gi en substansiell tolkning av de to analysene.

g) Sammenlign parameterverdiene for variabelen Andregenerasjons innvandrer for gutter og jenter. Test nullhypotesen om at de to kjønn har samme parameterverdi. Hva forteller resultatet oss? (Hint: Standardfeilen til differansen mellom to parameterestimer kan beregnes på samme måte som i pkt. b ovenfor.)

Tabell 4. Lineære regresjonsanalyser med indeks for antisosial atferd som avhengig variabel.

	Modell nr. 1		Modell nr. 2: Gutter		Modell nr. 3: Jenter	
	b	SE(b)	b	SE(b)	b	SE(b)
Konstant	3.89	0.14	5.07	0.35	3.62	0.20
Kjønn	-1.46 ^{***}	0.11	-	-	-	-
Alder	0.653 ^{***}	0.147	0.878 ^{**}	0.255	0.502 ^{**}	0.152
Alder kvadrert	-0.184 ^{***}	0.034	-0.206 ^{**}	0.059	-0.196 ^{***}	0.036
Bor med begge foreldre	-	-	-1.72 ^{***}	0.22	-0.83 ^{***}	0.13
Hjemstedets størrelse	-	-	-0.10 [*]	0.05	-0.05 ^{NS}	0.035
Andre generasjons innvandrere	-	-	0.84 [*]	0.42	-0.47 ^{NS}	0.28
Justert R ²	0.022		0.018		0.026	

^{NS} $p > 0.05$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Det er også konstruert en dikotom variabel på grunnlag av indeksen, slik at de med indeksverdi 10 eller høyere har fått verdien 1 og de øvrige har fått verdien 0. Denne dikotome variabelen er brukt som en indikator på alvorlige antisosiale handlinger. Det er gjort en logistisk regresjonsanalyse med denne dikotome variabelen som avhengig variabel og med de samme uavhengige variabler som ovenfor. I tabell 5 gjengis resultatet av analysen for gutter.

h) Forklar hva størrelsene i de ulike kolonnene i tabellen måler og gi en tolkning av analyseresultatet. Er det noen vesentlig forskjell, sammenlignet med den lineære regresjonsanalysen?

i) Hva er Hosmer-Lemeshow en test på, og hva forteller testen oss i dette tilfellet?

Tabell 5. Logistisk regresjonsanalyse med dikotomisert indeks for antisosial atferd som avhengig variabel. Bare gutter.

	b	SE(b)	Wald	sig	Exp(b)
Konstant	-1.980	0.171	134.7	<0.001	0.138
Alder	0.385	0.130	8.7	0.003	1.469
Alder kvadrert	-0.076	0.030	6.6	0.010	0.926
Bor med begge foreldre	-0.569	0.099	33.0	<0.001	0.566
Hjemstedets størrelse	-0.015	0.028	0.3	0.588	0.985
Andregenerasjons innvandrere	0.292	0.200	2.1	0.144	1.340
Hosmer-Lemeshow	10.1, $df = 8$, $p = 0.261$				