

SKOLEEKSAMEN I

SOS1120 Kvantitativ metode

2. juni 2015
4 timer

Følgende hjelpemidler er tillatt på eksamen:

- alle pensumbøker (inkl. kompendiet fra unipub)
- programmerbar kalkulator
- 2 X A4 ark, egne notater (på begge sider): i alt fire sider
- egne notater i bøkene er lov

Sensur for eksamen faller 23. juni kl. 14.00. Sensuren publiseres i StudentWeb ca. kl. 15.00 samme dag.

Sensuren regnes som mottatt av studentene når den publiseres i StudentWeb. Vi minner om at kandidater som ønsker begrunnelse må be om dette **senest 1 uke etter at sensur er falt**. Frist for å klage på karakteren er tre uker etter at sensur er falt. Informasjon om rutiner for begrunnelser og klager ved ISS finnes på emnesiden.

Oppgavesettet består av **9 sider** inkludert denne.

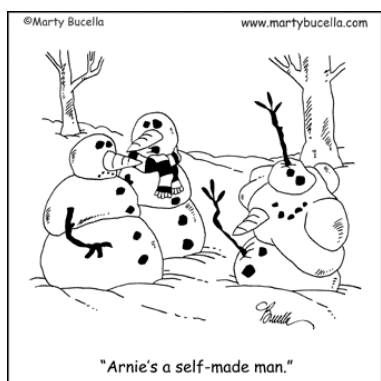
Kandidaten skal levere både originalen og kopien av besvarelsen. Kladd skal ikke leveres.

NB! Skriv hardt nok til at kopien blir leselig. Tusj penn kan ikke brukes.

Husk å notere deg kandidatnummeret ditt.

Flere av spørsmålene nedenfor har flervalgsalternativer («multiple choice»). Alle svar må skrives på gjennomslagsarkene. Følg derfor instruksjonene ved disse spørsmålene for hvordan svarene føres på gjennomslagsarkene.

«The self made man»: sosial ulikhet, arbeidstid og inntekt i USA



Noen sosiologer hevder at vi stadig blir friere fordi vi stadig får flere valgmuligheter i moderne samfunn, samtidig som at slike samfunn har større grad av meritokrati. Litt forenklet, betyr meritokrati at dine livssjanser i hovedsak bestemmes av hvor produktiv og flink du er. Ifølge resonnetet er vi i en forstand vår egen lykkes smed, som skaper våre muligheter gjennom hardt arbeid og utdanningsprestasjoner («talenter»), og dessuten hva vi velger å jobbe som («preferanser»).

Men samtidig styres mange valgmuligheter av hvor mye ting koster. Ikke alle har råd til å «velge» en flunkende ny miljøvennlig el-bil, eller å hyre inn en privatlærer til barna sine. Familieøkonomi og inntekt kan derfor spille en sentral rolle i den nye valgfriheten. Hvor mye du jobber (arbeidstid) og hva du jobber som (yrke) er igjen viktige for hva du tjener (inntekt).

Noen teorier om meritokrati og sosial ulikhet har sin opprinnelse blant amerikanske sosiologer på sekstitallet. Vi bruker derfor det amerikanske arbeidsmarkedet som et eksempel på et moderne samfunn. Her bruker vi data fra General Social Survey (GSS) i årene 1988 til 2010. Fra disse dataene har vi plukket ut personer som er mellom 30 og 70 år. Vi undersøker flere viktige aspekter som demografiske forhold, arbeidstid, sosial ulikhet og inntekt.

Variabelbeskrivelse (gjelder alle oppgaver i denne eksamen):

Variabelnavn	Betydning	Koding
region	Landsdeler i USA. Inndelt i 4 kategorier	1 = «east coast» 2= «central states» 3= «southern states» 4= «west coast»
sei50	Yrkets sosiale prestisjeskåre. Variabelen går fra 0 til 100. Men den er sentrert på prestisjeskåren 50	0 betyr prestisjeskåre «50» 1 betyr prestisjeskåre «51», osv
parsei50	Yrkets sosiale prestisjeskåre til foreldrene. Variabelen går fra 0 til 100, og er sentrert på skåren 50	0 betyr prestisjeskåre «50» 1 betyr prestisjeskåre «51», osv
sex	Respondentens kjønn	Mann = 0, kvinne =1
educ13year	Respondentens høyeste utdanning målt i antall år. Variabelen er sentrert på 13 års skolegang	Målt i antall år. Tallkode 0 betyr «13 år med skole», tallkode 1 betyr 14 år, osv)
degree	Respondentens høyeste utdanning målt i kategorier	1 = grunnskole («lt high school») 2 = videregående (« high school») 3 = påbygning til vgs (« junior college») 4 = universitet, lavere grad («bachelor») 5 = universitet, høyere grad («master/phd»)
year2000	Det aktuelle inntektsåret. Kodet som en dummyvariabel	0 = «før år 2000», 1 = «2000 og senere»
Parsei50Xyear2000	Samspillsvariabel mellom årsvariabelen (year2000) og foreldrenes sosiale yrkesstatus (parsei50)	samspillsledd

coninc	Inntekter fra arbeid	Målt i faktiske US-dollar (\$)
Age	Faktisk oppgitt alder	Tallkode 0 betyr 0 år gammel, osv
age30	Alder er sentrert på 30 år	Tallkode 0 betyr 30 år, tallkode 1 betyr 31 år, osv
hrs1	Antall arbeidstimer per uke.	Målt i antall arbeidstimer. Tallkoden 0 betyr 0 timer, etc..

Oppgave 1

- Hvilke målenivå befinner variablene «degree», «coninc», «region», «hrs1» seg på? (1 poeng)
- Begrunn kort hvorfor du mener «region» befinner seg på det målenivået (1 poeng)
- Hva ville sentraltendensmålet *modus* ha fortalt oss om du hadde brukt dette målet på variabelen «region»? (0.5 poeng)
- Kan vi bruke *median* på variabelen for region? Ja / Nei (0.5 poeng)

Tabell 1: Frekvenstabell av utdanningskategorier. N=26039

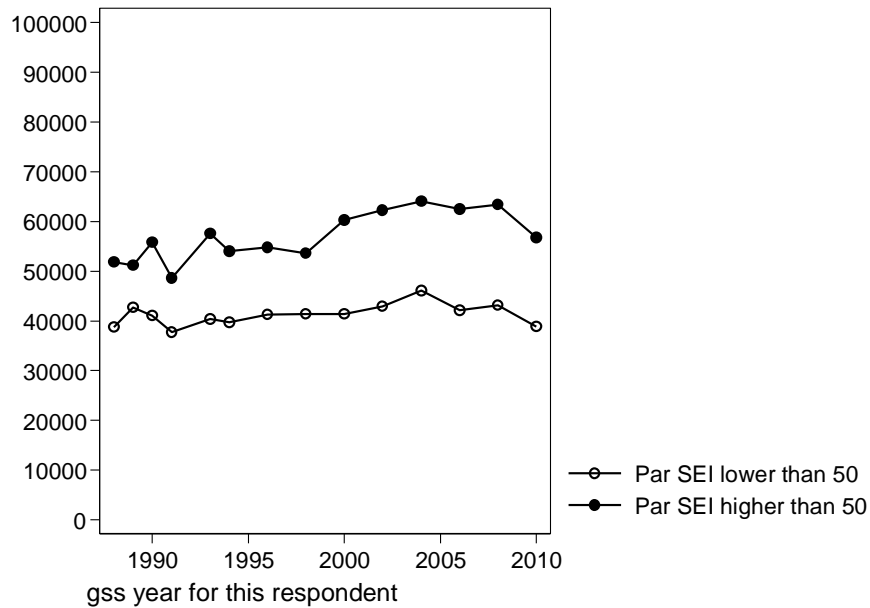
rs highest degree	Freq.	Percent	Cum.
lt high school	2,687	10.32	10.32
high school	13,820	53.07	63.39
junior college	1,832	7.04	70.43
bachelor	5,084	19.52	89.95
graduate	2,616	10.05	100.00
Total	26,039	100.00	

- Hvilken utdanningskategori er modus i tabell 1? (1 poeng)
- Tolk medianen i tabell 1 (1 poeng)
- Hvor stor prosentandel av utvalget har fullført bachelorgrad eller høyere ifølge denne tabellen? (1 poeng)
- Tegn et histogram basert på tabell 1. Marker stolpene med kategorinavn. Bruk antall observasjoner på y-aksen. (1 poeng) (Rund av til nærmeste tusen).

Oppgave 2

Anta at figuren (figur 1) nedenfor er representativ for den arbeidsføre delen av USA. *Beskriv kort* hva linjene i figuren nedenfor forteller oss om inntektsutviklingen avhengig om de har fra foreldre med høy eller lav yrkesposisjon. (2 poeng)

FIGUR1: Gjennomsnittlig årlig inntekt (i dollar) for personer med foreldre som har lav og høy yrkesstatus («parents' sei»: Par SEI). Personer mellom 30-70 år.



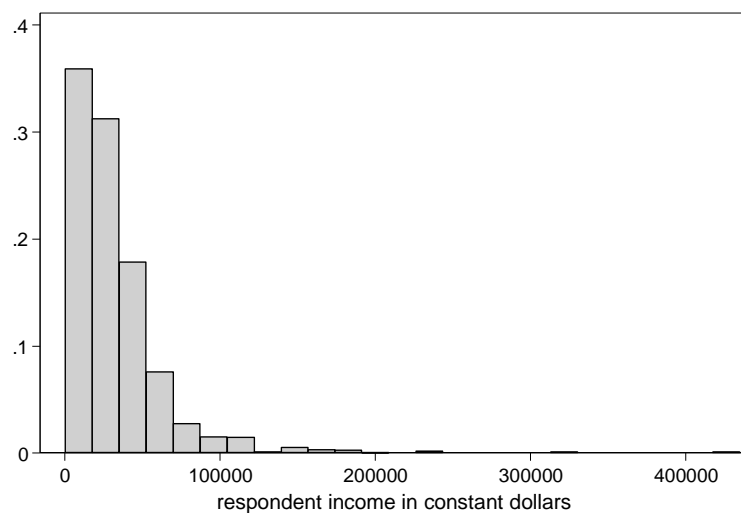
Oppgave 3

Tenk deg at en forsker stiller seg opp foran en kino klokken 13.00, en helt tilfeldig valgt dag. Kinoen har nettopp vist en barnefilm om Snurre Sprett. Folk strømmer ut av kinoen. Han spør de ti tilfeldig første personene som kommer ut av kinoen om de har lyst til å være med i en spørreundersøkelse. Anta at alle sier ja til å være med.

- Er dette et sannsynlighetsutvalg hvor analyseresultatene kan generaliseres til resten av befolkningen? Ja/nei (1 poeng)
- Begrunn kort svaret ditt i a. (2 poeng)

Oppgave 4

FIGUR 2: Histogram av inntektsfordeling i USA. Personer 30-70 år.



4.1 Inntektsfordelingen i figuren over er (1 poeng) :

- a. høyeskjev
- b. venstreskjev
- c. normalfordelt
- d. t-fordelt
- e. uniform

4.2 Hvis medianen i denne inntektsfordelingen er 35602 \$, hva er det mest sannsynlig gjennomsnittet for inntekt av følgende alternativer (1 poeng) :

- a. 44680 \$
- b. 35602 \$
- c. 12101 \$

Oppgave 5

Tabell 2: Krysstabell mellom inntekt (incDV) og utdanningsnivå (degreeDV). Dette er variabler som er omkodet til to kategorier basert på inntektsvariabelen «coninc» og utdanningsvariabelen «degree».

```
. ta degreeDV incDV , row v chi nof
```

RECODE of degree (rs highest degree)	RECODE of rincome (respondents income)		Total
	<25000	>=25000	
low educ	70.99	29.01	100.00
high educ	37.81	62.19	100.00
Total	62.37	37.63	100.00

Pearson chi2(1) = 3.1e+03 Pr = 0.000
Cramér's V = 0.3004

- 5.1 Anta at utdanning påvirker inntekten. Regn ut prosentdifferansen i forhold hva som er antatt årsak. Vis utregningen.(1 poeng)
- 5.2 Gi en kort tolkning av denne prosentdifferansen (1 poeng)
- 5.3 Tolk korrelasjonsmålet Cramers V (1 poeng)
- 5.4 Formuler en nullhypotese og en forskningshypotese for sammenhengen mellom utdanning og inntekt. (1 poeng)
- 5.5 Utfør en moderne kjikvadrattest. La signifikansnivået (alpha) være 0,05. Signifikanssannsynligheten (p-verdien) er oppgitt som «Pr =0,000». (2 poeng)

Oppgave 6

Nedenfor ser du at tabellen oppgir beskrivende statistikk for prestisjeskåre for arbeidstimer per uke.

Tabell 3: Deskriptiv statistikk for antall arbeidstimer per uke (hrs1).
Mean = gjennomsnitt, SE(mean) = standardfeil, N = antall observasjoner.

. tabstat hrs1 , stat(mean sem N)

variable	mean	se (mean)	N
hrs1	42.57777	.1117816	15835

- Finn kritisk t-verdi i vedleggstabellen for et signifikansnivå på 99,9% (alfa= 0.001). Frihetsgrader er gitt ved $df = N-1$. (1 poeng)
- Beregn (estimér) et 99,9% konfidensintervall for gjennomsnittet av arbeidstimer i utvalget basert på formelen (1 poeng):

$$KI_{99,9\%} = mean \pm t*SE$$

- Gi en kort tolkning av konfidensintervallet (1 poeng)

Oppgave 7

Nedenfor ser du en korrelasjonstabell mellom flere av variablene. Alle tallene viser pearsons-r som er korrelasjonsmål for to kontinuerlige variabler av gangen.

Tabell 4: korrelasjonstabell. Pearsons-r

. corr coninc sei parsei region
 (obs=24191)

	coninc	sei	parsei	region
coninc	1.0000			
sei	0.3915	1.0000		
parsei	0.2230	0.2956	1.0000	
region	-0.0230	0.0169	0.0444	1.0000

- Tolk korrelasjonen mellom barnas sosiale yrkesstatus(sei) og inntekt (coninc). (1 poeng)
- Hvorfor viser alle korrelasjonene langs diagonalen i tabellen $r=1$? (1 poeng)
- Hvilke(n) av variablene hører ikke hjemme rent statistisk sett i denne korrelasjonstabellen? (2 poeng)

Oppgave 8 (2 poeng)

Noen hevder at meritokrati i moderne samfunn fremmer sosial reproduksjon på tvers av generasjoner i moderne samfunn. Ifølge dette resonnementet overføres talent, utdanningspreferanser og jobbpreferanser, enten sosialt eller genetisk fra foreldregenerasjon til barna. I et «perfekt» meritokrati vil familiebakgrunn dermed ha en fullstendig indirekte

effekt på barnas inntekt. Familiebakgrunn vil virke gjennom at barn av foreldre med høy status oftere får høy utdanning og dermed høyere lønn enn barn av foreldre med lav status. Rent statistisk sett betyr dette at etter kontroll for barnas yrkesposisjon og utdanning, vil familiebakgrunn dermed ikke ha noen gjenværende direkte effekt på barnas inntekt.

Tegn et årsak-virkningsdiagram som viser hvordan foreldres sosiale posisjon (parsei50) antas å påvirke inntekt (coninc) *fullstendig* gjennom egen utdanning (educ13) og yrkesprestisje (sei50).

Oppgave 9

Vi ønsker her å undersøke sammenhengen mellom forelderens sosiale posisjon (parsei) og barnas inntekt oppgitt i US-dollars (\$\$\$). Vi inkluderer kontrollvariabler i alle modellene.

Modell 1:

```
. reg coninc parsei50 age30, beta
```

Source	SS	df	MS		
Model	1.9270e+12	2	9.6350e+11	Number of obs =	25337
Residual	3.5858e+13	25334	1.4154e+09	F(2, 25334) =	680.72
Total	3.7785e+13	25336	1.4914e+09	Prob > F =	0.0000
				R-squared =	0.0510
				Adj R-squared =	0.0509
				Root MSE =	37622

coninc	Coef.	Std. Err.	t	P> t	Beta
parsei50	458.3265	12.52584	36.59	0.000	.2273166
age30	22.77181	14.28682	1.59	0.111	.0099021
_cons	48303.74	321.8364	150.09	0.000	.

9.1 Tolk konstantleddet i modell 1 (1 poeng)

9.2 Tolk regresjonskoeffisienten for alder (age30) (1 poeng)

Modell 2

```
. reg coninc parsei50 age30 educ13years sei50, beta
```

Source	SS	df	MS		
Model	6.9663e+12	4	1.7416e+12	Number of obs =	24133
Residual	2.9021e+13	24128	1.2028e+09	F(4, 24128) =	1447.93
Total	3.5987e+13	24132	1.4913e+09	Prob > F =	0.0000
				R-squared =	0.1936
				Adj R-squared =	0.1934
				Root MSE =	34681

coninc	Coef.	Std. Err.	t	P> t	Beta
parsei50	144.1102	12.7465	11.31	0.000	.0714973
age30	26.82744	13.9323	1.93	0.054	.0114513
educ13years	2839.341	97.82583	29.02	0.000	.2174337
sei50	483.5553	14.53315	33.27	0.000	.2415356
_cons	47670.97	321.2127	148.41	0.000	.

- 9.3 Fortolk den (ujusterte) R^2 i modell2 (1 poeng)
- 9.4 Hvilken av variablene i modell 2 ser ut til å være minst viktig for barnas inntekt? (Oppgi tallet du vurderer dette ut i fra) (1 poeng)
- 9.5 Begrunn kort hvorfor regresjonskoeffisienten til foreldrenes sosiale yrkestatus (parsei50) synker i modell 2 i forhold til modell 1? (2 poeng)
- 9.6 Predikér årlig inntekt for en person med 14 års utdanning, som er 45 år gammel, som har gjennomsnittlig prestisjeskåre (SEI=50), og som har foreldre med prestisjeskåre PARSEI= 51. Vis utregning (2 poeng).

Hint: Husk å kikke på variabelbeskrivelsen.

Oppgave 10

Modell 3:

```
. reg coninc parsei50 age30 educ13years sei50 year2000 parsei50Xyear2000
```

Source	SS	df	MS	Number of obs =	24133
Model	7.0587e+12	6	1.1765e+12	F(6, 24126) =	981.14
Residual	2.8929e+13	24126	1.1991e+09	Prob > F =	0.0000
Total	3.5987e+13	24132	1.4913e+09	R-squared =	0.1961
				Adj R-squared =	0.1959
				Root MSE =	34627

coninc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
parsei50	63.55626	17.69913	3.59	0.000	28.86487 98.24765
age30	20.9517	13.93013	1.50	0.133	-6.352225 48.25562
educ13years	2822.469	97.74828	28.87	0.000	2630.876 3014.061
sei50	481.5806	14.51235	33.18	0.000	453.1355 510.0257
year2000	2917.458	448.6771	6.50	0.000	2038.023 3796.893
parsei50Xyear2000	148.1438	23.3335	6.35	0.000	102.4087 193.879
_cons	46245.7	383.8942	120.46	0.000	45493.25 46998.16

I modell 3 kontrolleres det for tid (før og etter 2000). Det er dessuten inkludert et samspillsledd mellom år og foreldrenes yrkesposisjon (parsei50Xyear2000).

- a. Hvor mye mer betyr foreldrenes sosiale yrkes status for barnas inntekt etter år 2000, selv etter vi har kontrollert for barnas utdanning og yrkesstatus? (2 poeng).
- b. Ta utgangspunkt i likningen:

$$\widehat{\text{Inntekt}} = 46245.7 + 63,6 * \text{parsei50} + 2917,5 * \text{year2000} + 148 * \text{parsei50Xyear2000}$$

Utled to regresjonslikninger basert på dummyvariabelen for tid (year2000): før og etter år 2000.(3 poeng).

VEDLEGG:

T-fordeling (forkortet tabell)

Df	Signifikansnivåer for tohalet t-test (alfa)					
	0.2	0.1	0.05	0.02	0.01	0.001
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.599
3	1.638	2.353	3.182	4.541	5.841	12.924
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.869
10	1.372	1.812	2.228	2.764	3.169	4.587
50	1.299	1.676	2.009	2.403	2.678	3.496
60	1.296	1.671	2.000	2.390	2.660	3.460
80	1.292	1.664	1.990	2.374	2.639	3.416
120	1.289	1.658	1.980	2.358	2.617	3.373
>120	1.282	1.645	1.960	2.326	2.576	3.291