

## Eksamensoppgave SOSGEO1120 våren 2020

Hele eksamen er lagt opp med randomiserte oppgaver innenfor bolker. Ingen kandidater skal da ha fått helt lik eksamen. Fullstendig eksamen legges ikke ut.

På grunn av omstendighetene dette året var det ikke basert på å gjøre kjøringene selv med statistikksoftware, men tolke output og kunne kodene.

Det var ikke lagt opp til å skulle legge ved håndregning på grunn av usikkerhet rundt mulige tekniske problemer.

- Den første bolken er flervalgsoppgaver trukket tilfeldig fra en oppgavebank. (Legges ikke ut)
- En oppgave trukket tilfeldig fra oppgavebank om oppslag i normalfordelingen. (Legges ikke ut)
- To oppgaver trukket tilfeldig fra oppgavebank om anvendelse av fordelinger. (Legges ikke ut)
- En større oppgave med flere deler er basert på softwaren STATA. Her var det flere oppgavesett med lik tekst, men ulike underutvalg av datasettet slik at svarene ikke nødvendigvis var like. Det er kun lagt ut en variant av denne oppgaven.
- En oppgave om STATA-kode (lik for alle)
- Fire oppgaver i tilfeldig rekkefølge (men lik for alle)
- To oppgaver med tekstsvaer (lik for alle)

SOSGEO1120 V2020

### i SOSGEO1120 Info eksamen 2020

#### Informasjon om oppgavetyper

Opgavesettet består av en rekke enkeltoppgaver av ulik type. En god del av oppgavene vil være ett svar på hver side, mens en del andre vil ha flere spørsmål på hver side.

For alle enkelt spørsmål der du skal krysse av for riktig svar (dvs. flervalgsoppgave) får du 1 poeng hvis riktig, men blir trukket 0.25 poeng for hvert feil svar. Ubesvarte gir null poeng. Med andre ord: Du bør altså ikke gjette hvis du ikke er ganske sikker på svaret.

Andre oppgavetyper er å skrive inn riktig tall eller skrive noe tekst. For disse oppgaven får man ikke trekk hvis feil.

For alle typer oppgaver får du 1 poeng per riktige svar hvis ikke annet er angitt.

For oppgavetyper der du skal fylle inn riktig tall er det viktig at du svarer med riktig antall desimaler. Galt antall desimaler kan medføre at svaret blir regnet som feil.

Noen oppgaver krever håndregning før svaret avgis. Det er IKKE lagt opp til at håndregning skal leveres inn.

Merk: Du kan gå frem og tilbake i oppgavesettet og justere svarene senere.

## 17 SOSGEO1120 STATA\_a - ESS tillitt

Du er interessert i sammenhengen mellom politisk ståsted og tillitt til andre mennesker blant bosatte personer i Norge. For å undersøke dette tar du utgangspunkt i European Social Survey; en representativ spørreundersøkelse som inneholder spørsmål om bl.a. tillitt, holdninger og sosioøkonomisk bakgrunn.

8/44

---

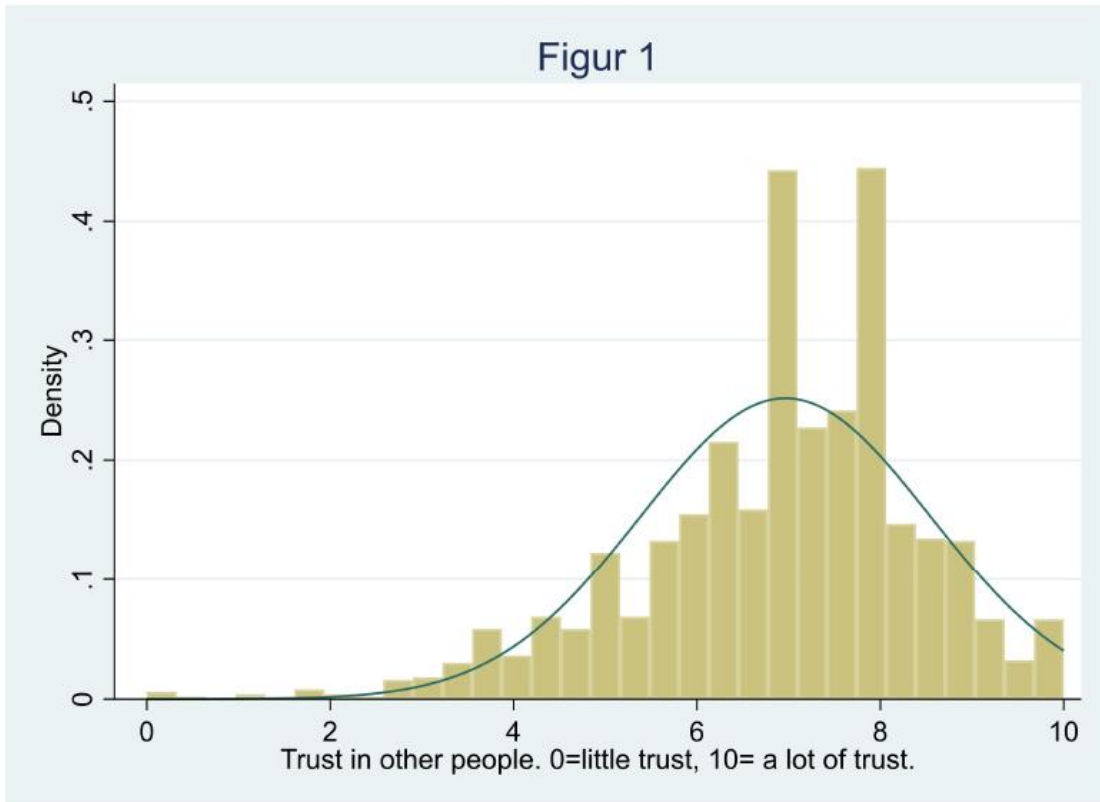
SOSGEO1120 V2020

### Oppgave 1: Deskriptiv statistikk

Som avhengig variabel bruker du variabelen "trustpeople", som er basert på følgende spørsmål:

*Using this card, generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people? Please tell me on a score of 0 to 10, where 0 means you can't be too careful and 10 means that most people can be trusted.*

I det følgende vil denne variabelen bli omtalt som «tillitt». Figur 1 viser et histogram over fordelingen på tillitt-variabelen i utvalget, med en normalfordelingskurve tegnet inn. Output 1 viser resultatet fra summarize-kommandoen (m. detail-oppsjonen) i STATA.



**Output 1:**

```
. summarize trustpeople, detail
```

```
Trust in other people. 0=little trust, 10= a lot
of trust.
```

Percentiles		Smallest		
1%	2.666667	0		
5%	4	0		
10%	4.666667	0	Obs	1,541
25%	6	.3333333	Sum of Wgt.	1,541
50%	7		Mean	6.96539
		Largest	Std. Dev.	1.585909
75%	8	10		
90%	8.666667	10	Variance	2.515107
95%	9.333333	10	Skewness	-.794873
99%	10	10	Kurtosis	4.061154

```
.
```

**Oppgave a)**

Gi en 5-talls-oppsummering av fordelingen av tillitt-variabelen.

9/44

---

SOSGEO1120 V2020

**Velg ett alternativ**

- 0 - 6 - 7 - 8 - 10
- 0 - 0.33 - 7 - 10 - 10
- 1541 - 6.96 - 7 - 1.59 - 10
- 0 - 4.6 - 8 - 9.33 - 10

**Oppgave b)**

Hvordan vil du beskrive fordelingen?

**Velg ett alternativ**

- Symetrisk skjev
- Symetrisk
- Høyreskjev (hale til venstre)
- Venstreskjev (hale til høyre)

**Oppgave c)**

Hvis du rangerer alle verdiene på tillitt-variabelen fra lavest til høyest, hvilken verdi deler de 25% med høyeste verdier på tillitt-variabelen fra de resterende 75%?

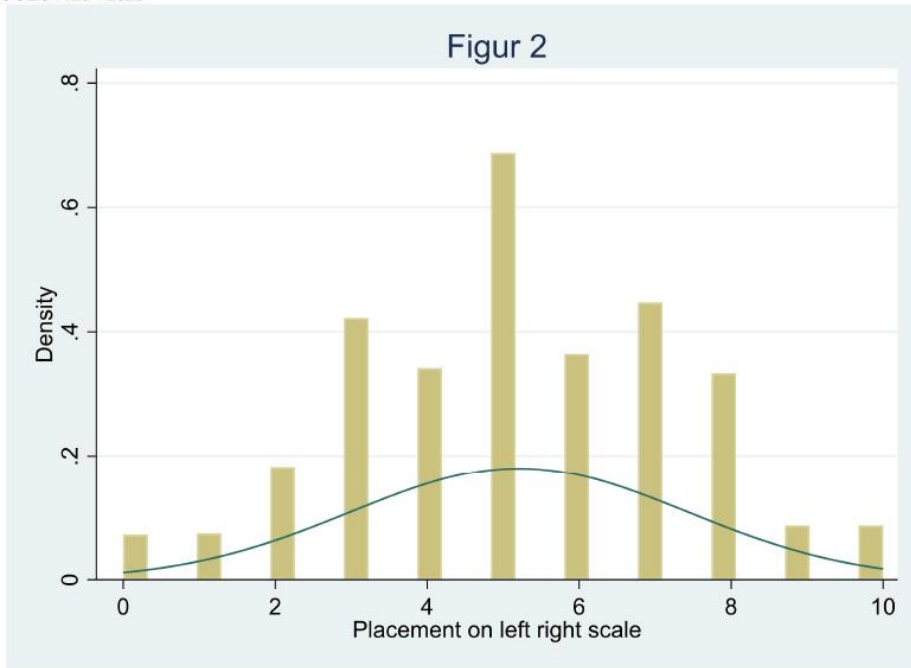
Skriv inn riktig tall med en desimal her:

Som den sentrale uavhengige variabelen bruker du variabelen «left\_rigt\_scale», som er basert på følgende spørsmål:

*In politics people sometimes talk of 'left' and 'right'. Using this card, where would you place yourself on this scale, where 0 means the left and 10 means the right?*

I det følgende vil denne variabelen omtales som «politisk ståsted».

Figur 2 viser et histogram over fordelingen på variabelen for politisk ståsted i utvalget, med en normalfordelingskurve tegnet inn. Output 2 viser resultatet fra summarize-kommandoen (m. detail-oppsjonen) i STATA for variabelen politisk ståsted.



**Output 2:**

```
. summarize left_right_scale, detail
```

Placement on left right scale				
Percentiles	Smallest			
1%	0	0		
5%	2	0		
10%	2	0	Obs	1,497
25%	4	0	Sum of wgt.	1,497
50%	5		Mean	5.209085
		Largest	Std. Dev.	2.219499
75%	7	10		
90%	8	10	Variance	4.926174
95%	9	10	Skewness	-.07946
99%	10	10	Kurtosis	2.604055

d) Hvor mange personer har svart på dette spørsmålet om politisk ståsted?

SVAR:

e) Hva er den vanligste verdien å svare på spørsmålet om politisk ståsted?

SVAR:

**Oppgave 2: Korrelasjon**

Du undersøker først korrelasjonen mellom politisk ståsted og tillitt ved å bruke corr-kommandoen i STATA. Output vises nedenfor.

**Output 3:**

```

. corr left_right_scale trustpeople
(obs=1,494)

              left_right_scale  trustpeople
-----
left_right_scale      1.0000
trustpeople          -0.0884      1.0000

```

a) Hva er korrelasjonskoeffisienten? Bruk fire desimaler i svaret:

b) Hva viser denne korrelasjonskoeffisienten?

**Velg ett alternativ**

- At mer venstreorientert politisk typisk går sammen med høyere verdier på tillitt.
- At mer venstreorientert politisk typisk går sammen med lavere verdier på tillitt.
- At mer høyreorientert politisk typisk går sammen med høyere verdier på tillitt.

At verdier på politisk ståsted (målt på en skala fra venstre til høyre) typisk går sammen med verdier på tillitt.

**Oppgave 3**

Du går nå videre ved å undersøke sammenhengen mellom politisk ståsted og tillitt i en enkel, lineær regresjonsmodell. Resultatet ses i Output 4.

**Output 4:**

```

. reg trustpeople left_right_scale

Source |           SS           df           MS       Number of obs   =       1,494
-----+-----+-----+-----+-----+-----
Model   |   28.7875122           1   28.7875122   F(1, 1492)       =       11.75
Residual|  3656.43838         1,492   2.45069597   Prob > F         =       0.0006
Total   |  3685.22589         1,493   2.46833516   R-squared        =       0.0078
                                           Adj R-squared    =       0.0071
                                           Root MSE       =       1.5655

-----+-----
trustpeople |           Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----+-----+-----+-----+-----+-----
left_right_scale |  -.0625965   .0182639   -3.43   0.001   -0.0984222   -0.0267709
               _cons |   7.29779   .1034748   70.53   0.000   7.094818    7.500761

```

a) Hva er gjennomsnittsverdien på tillitsvariabelen for personer som oppgir å være helt til venstre på den politiske aksene? Oppgi svaret med fire desimaler.

SVAR:

b) Hva er stigningstallet/regresjonskoeffisienten for politisk ståsted? Oppgi svaret med fire desimaler.

c) Beskriv med én setning hva stigningstallet for politisk ståsted betyr:

d) Hva er p-verdien til stigningstallet for politisk ståsted?

e) Beskriv med én setning hva denne p-verdien betyr

f) Hva er  $R^2$ ? Oppgi svaret i prosent med tre desimaler.

#### Oppgave 4 Multipl regressjon

Du legger nå til fire sosiodemografiske variable du mistenker at kan drive sammenhengen mellom politisk ståsted og tillitt til andre mennesker:

«age»: respondentens alder (i år) – heretter omtalt som «alder»

«female»: respondentens kjønn (kodet som en dummy satt til 1 for kvinner og 0 for menn) – heretter omtalt som «kjønn»

«education»: respondentens høyeste fullførte utdanningsnivå (numerisk variabel med verdier for:

1. Grunnskole, folkehøgskole eller lavere

2. Videregående

3. Universitet/høgskole, kort

4. Universitet/høgskole, lang

5. Doktorgrad) – heretter omtalt som «utdanningsnivå»

«hh\_inc\_decile»: respondentens samlede husholdningsinntekt (numerisk variabel kodet i desiler (dvs. hvilket inntektsdesil respondenten tilhører)) – heretter omtalt som «inntekt».

En oversikt over disse variablene (basert på codebook-kommandoen), samt en tabell over utdanningsvariablen (basert på tabulate-kommandoen), ser du i Output 5.

#### Output 5:

```
. codebook age female education hh_inc_decile, compact
```

Variable	Obs	Unique	Mean	Min	Max	Label
age	1545	82	46.96311	15	98	Age of respondent, calculated
female	1545	2	.4527832	0	1	
education	1539	5	2.459389	1	5	Highest level of education
hh_inc_decile	1465	10	5.262799	1	10	Household's total net income, all sources

```
. tab education
```

Highest level of education	Freq.	Percent	Cum.
Grunnskole, folkehøgskole eller lavere	277	18.00	18.00
Videregående	558	36.26	54.26
Universitet/høgskole, kort	447	29.04	83.30
Universitet/høgskole, lang	234	15.20	98.51
Doktorgrad	23	1.49	100.00
Total	1,539	100.00	

Resultatet fra den multiple regressjonsmodellen ser du i Output 6.

#### Output 6:

```
. reg trustpeople left_right_scale age female i.education hh_inc_decile
```

Source	SS	df	MS	Number of obs	=	1,420
Model	261.53262	8	32.6915775	F(8, 1411)	=	14.57
Residual	3165.94276	1,411	2.24375816	Prob > F	=	0.0060
Total	3427.47538	1,419	2.41541606	R-squared	=	0.0763
				Adj R-squared	=	0.0711
				Root MSE	=	1.4979

trustpeople	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
left_right_scale	-.0611682	.0183432	-3.33	0.001	-.0971511 -.0251853
age	.0121481	.0022063	5.51	0.000	.0078201 .0164761
female	.060384	.0814931	0.74	0.459	-.0954768 .2202447
education					
Videregående	.1931181	.121421	1.59	0.112	-.0450669 .4313031
Universitet/høyskole, kort	.6328811	.1272013	4.98	0.000	.3833571 .8824051
Universitet/høyskole, lang	.6970326	.1458814	4.78	0.000	.4168649 .9832003
Doktorgrad	.5635116	.345705	1.63	0.103	-.1146393 1.241663
hh_inc_decile	.0596275	.0155773	3.83	0.000	.0250703 .0901847
_cons	6.009567	.1966344	30.56	0.000	5.623839 6.395294

a) Hva er stigningstallet for politisk ståsted i den multiple modellen? Oppgi svaret med fire desimaler.

SVAR:

b) Hva viser resultatet om mistanken du hadde om at sammenhengen mellom politisk ståsted og tillitt til andre mennesker du så i Output 4 var spurios? Skriv 1-2 setninger.

SVAR:

c) Er forskjellen i tillitt til andre mennesker mellom personer med doktorgrad og grunnskoleutdanning statistisk signifikant på 1%-nivå?

Velg et alternativ

- Nei  
 Ja

d) Hvilke av følgende alternativ er rimelig tolknig av  $R^2$  fra denne modellen?

Velg ett alternativ

- Forklart varians er lav, men over det som regnes som kritisk grense for å regresjon.  
 Modellen vil ikke være særlig pålitelig hvis den skal brukes til å predikere tillitt verken for grupper eller individer.  
 Modellen har så liten forklaringskraft at den ikke kan brukes til noen ting.  
 Modellen vil ikke være særlig pålitelig for å predikere tillitt for individer, men på gruppenivå vil den være ganske treffsikker.

Du har lest flere tidligere studier som fant at sammenhengen mellom politisk ståsted og tillitt varierte med



utdanningsnivå. For å undersøke dette i dine data koder du først om variabelen «education» til en dummy («high\_ed») som tar verdien 1 for de som har fullført kort høyere utdanning, lang høyere utdanning eller doktorgrad, og 0 for de som ikke har det). Deretter legger du inn et samspillsledd mellom politisk ståsted og utdanningsnivå i modellen. Resultatet fra denne modellen ser du i Output 7.

#### Output 7:

```
. reg trustpeople c.left_right_scale#c.high_ed age female hh_inc_decile
```

Source	SS	df	MS	Number of obs =	1,420
Model	264.580402	6	44.0967336	F(6, 1413)	= 19.70
Residual	3162.89498	1,413	2.23842532	Prob > F	= 0.0000
				R-squared	= 0.0772
				Adj R-squared	= 0.0733
Total	3427.47538	1,419	2.41541606	Root MSE	= 1.4961

trustpeople	Coeff.	Std. Err.	t	P> t	[95% Conf. Interval]
left_right_scale	-.0255028	.0253558	-1.01	0.315	-.0752419 .0242362
high_ed	.9001106	.2048384	4.39	0.000	.4982906 1.301931
c.left_right_scale#c.high_ed	-.07392	.0358948	-2.06	0.040	-.1443328 -.0035072
age	.0117632	.002186	5.38	0.000	.0074751 .0160514
female	.0442854	.0812606	0.54	0.586	-.115119 .2036898
hh_inc_decile	.0621036	.0154738	4.01	0.000	.0317495 .0924577
_cons	5.963574	.1972065	30.24	0.000	5.576725 6.350423

a) Hva er stigningstallet politisk orientering for de med høyere utdanning?

#### Velg ett alternativ

- 0.0256
- 0.0484
- 0.9
- 0.0484
- .0994

b) Hva er en rimelig tolkning av samspillsleddet?

#### Velg ett alternativ

- For de med lavere utdanning vil politisk orientering ha mer å si for tillitt enn det har å si for de med høyere utdanning
- For de med høyere utdanning vil politisk orientering ha mer å si for tillitt enn det har å si for de uten høyere utdanning
- For de med høyere utdanning vil politisk orientering ha mindre å si for tillitt enn det har å si for de uten høyere utdanning
- Utdanningsnivå har ingenting å si for sammenhengen mellom politisk orientering og tillitt .

c) Gitt at du hadde en hypotese om at sammenhengen mellom politisk orientering og tillitt varierte med utdanning. Hvordan ville du da konkludert i dette tilfellet? Ta utgangspunkt i et 5% signifikansnivå.

-----  
**Velg ett alternativ**

- Det er ikke et slikt samspill. Det estimatet vi fikk her er ganske sannsynlig å få ved en tilfeldighet.
- Det er et slikt samspill. Det er lite sannsynlig å ikke få et slikt estimatet ved en tilfeldighet
- Det er et slikt samspill. Det estimatet vi fikk her er lite sannsynlig at vi ville fått ved en tilfeldighet.

**21 SOSGEO1120 STATA FV\_a scatter**

Datasettet har følgende variable:

kvinne - indikatorvariabel for kvinne (dummy) 0 = mann, 1 = kvinne

alder - alder per 1. januar

gjengang - indikatorvariabel (dummy) for om vedkommende har tidligere straffedommer 0=ingen, 1 = mer enn 1

vinning - antall tidligere dommer for vinningslovbrudd

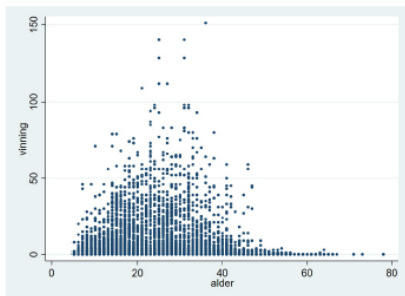
vold - antall tidligere dommer for voldslovbrudd

sivilstand - sivilstand per 1. januar

residiv - om vedkommende har fått en ny dom i løpet av oppfølgingstiden

**Oppgave a)**

Hvilke kode er brukt til å lage dette plottet?

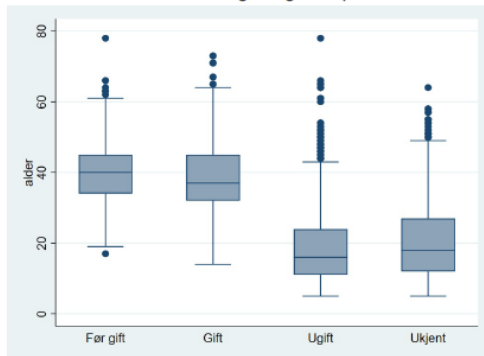


**Velg ett alternativ**

- scatter vinning alder
- scatter vold alder
- graph scatter vinning, by(alder)
- twoway (vinning ) ///  
scatter( alder )
- twoway (scatter vinning alder) ///  
(fit vinning alder)

### Oppgave b)

Hvilken kode er brukt til å lage følgende plot?



### Velg ett alternativ

- graph bar alder sivilstand
- graph box sivilstand, over(alder)
- graph box alder , over(sivilstand)
- graph bar alder , over(sivilstand)

### Oppgave c)

Hvilke kode er brukt til å lage følgende tabell?

residiv	kvinne		Total
	0	1	
0	8,028	1,935	9,963
1	7,774	1,261	9,035
Total	15,802	3,196	18,998

Pearson  $\chi^2(1) = 101.1328$  Pr = 0.000

### Velg ett alternativ:

- tab residiv kvinne, chi2
- prtest residiv, by(kvinne)
- tabstat residiv, by(kvinne)
- tab kvinne \* residiv , chi2 col row
- ttest residiv, by(kvinne)
- tab residiv kvinne, col row

### Oppgave d)

Hvilke kode er brukt til å lage følgende output?

Source	SS	df	MS	Number of obs	=	18,998
Model	557.920118	5	111.584024	F(5, 18992)	=	506.96
Residual	4180.24732	18,992	.220105693	Prob > F	=	0.0000
				R-squared	=	0.1178
				Adj R-squared	=	0.1175
Total	4738.16744	18,997	.249416615	Root MSE	=	.46915

residiv	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
alder	-.0077301	.0002979	-25.95	0.000	-.0083141 -.0071462
kvinne	-.0773636	.0095401	-8.11	0.000	-.096063 -.0586642
vinning	.0108548	.0003704	29.30	0.000	.0101288 .0115808
vold	.0331524	.0016714	19.83	0.000	.0298763 .0364286
c.kvinne#c.vold	.0423975	.0082532	5.14	0.000	.0262205 .0585745
_cons	.5885114	.0075331	78.12	0.000	.5737458 .6032771

**Velg ett alternativ**

- reg residiv alder kvinne vinning vold
- reg residiv alder vinning vold, by(kvinne)
- reg residiv alder kvinne vinning vold c.kvinne#c.vinning
- reg residiv alder kvinne vinning vold kvinne#i.vold
- reg residiv alder kvinne vinning vold c.kvinne#c.vold

## 26 SOSGEO1120 TXT Ferie og skolekarakterer

En gruppe forskere lurer på om det å dra på ferie til andre land kan forbedre barnas språk- og læringsevne gjennom eksponering for andre språk og kulturer. De har derfor undersøkt sammenhengen mellom hvor mye penger foreldre bruker på feriereiser for barna sine og karakterene barna får på skolen. Dette har de gjort på følgende måte:

De har trukket et tilfeldig utvalg på 1000 familier der eldste barn nettopp har fullført 10. klasse på skoler i Oslo vest, Asker og Bærum. En forelder i hver av disse familiene har blitt spurt «Hvor mye penger brukte familien på feriereiser til utlandet i fjor?», der svarene er oppgitt i kroner, og «Hvilken karakter fikk ditt eldste barn i norsk hovedmål, skriftlig ved eksamen i 10.klasse?», der svarene er oppgitt i tallkarakter, 1-6. I tillegg ble foreldrene spurt om kjønnnet på sitt eldste barn. 437 familier svarte på undersøkelsen.

Forskerne gjør en multipl regressjonsanalyse der  $y$  er karakterer,  $x$  er penger brukt på feriereiser og  $z$  er en dummyvariabel for kjønn (gutt = 0, jente = 1):

$$y = a + b_1x + b_2z$$

Resultatet ble:

	Koeffisient	P-verdi
Penger brukt på feriereiser	0.00005	0.047
Kjønn	0.32104	0.001
Konstant	3.34591	0.000

$R^2 = 0.09$   
 $F = 0.00$

Forskerne konkluderer med at:

«Feriereiser i utlandet har en signifikant effekt på norske ungdommers karakterer. Å bruke 20 000 kr mer på ferieturen kan forbedre karakterene med et helt poeng. Dette skyldes at det å eksponeres for andre språk og kulturer bidrar til å utvikle språk og læringsevner. Effekten er større for jenter enn for gutter.»

Hva er galt med denne konklusjonen? Nevn opptil 3 ting. (Det kan være flere feil, men det gis max 3 poeng).

Skriv ditt svar her...

## 27 SOSGEO1120 TXT Covid

Folkehelseinstituttet gjennomførte en studie der 397 personer ble testet for om de har vært smittet av koronaviruset. Det ble estimert at det var under 2 prosent som har vært smittet.

Dataene er et underutvalg av den store nasjonale helseundersøkelsen "Mor og barn" (MOBA). Utvalget er 100 000 kvinner som var gravide mellom 1999 og 2006, som så har blitt fulgt fremover i tid. Testen for koronavirus ble gjort på et underutvalg som bor i Oslo. Du kan gå ut fra at den medisinske testen er pålitelig.

I et intervju i NRK (7. mai 2020) uttaler en forsker: "Dette forteller at den norske befolkningen i liten grad viser tegn på at de har hatt infeksjon".

Er det god dekning for å si det på dette grunnlaget? Kommenter kort.

Skriv ditt svar her...

