

UTSATT SKOLEEKSAMEN I

SOS4020 - KVANTITATIV METODE

29. Mars 2017

(4 timer)

Tillatte hjelpemidler: Alle skriftlige hjelpemidler og kalkulator.

Sensur for eksamen faller 19. April klokken 14.00. Sensuren publiseres i Studentweb cirka klokken 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 sensur er falt**. Frist for å klage på karakteren er tre uker etter sensur er falt. Informasjon om rutiner for begrunnelser og klager ved Institutt for sosiologi og samfunnsgeografi finnes på emnesiden

Oppgavesettet består av **6 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.

LYKKE TIL!

OPPGAVESETTET BESTÅR AV 15 DELOPPGAVER. HUSK Å BESVARE ALLE OPPGAVENE!

SENSORER HAR IKKE TILGANG TIL DINE SKRIFTLIGE HJELPEMIDLER. FØLGELIG BØR DU IKKE HENVISE TIL DISSE.

En forsker ønsker å undersøke om utdanningsnivå har en effekt på holdninger til alternativ medisin. Han bruker data fra spørreundersøkelsen General Social Survey, samlet inn i USA i 2012 (<http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/35478/>).

I analysene bruker han følgende variable:

- **AGE:** Alder i år på intervju tidspunktet.
- **WOMAN:** Dummyvariabel som tar verdien 1 hvis respondenten er kvinne, 0 for menn.
- **EDUC:** Respondentens utdanning i antall år.
- **ALTMED:** Respondentens rating av følgende utsagn: "Alternative medicine provides better solutions". Variabelen har følgende verdier:
 1. Strongly agree
 2. Agree
 3. Neither agree nor disagree
 4. Disagree
 5. Strongly disagree

Bare respondenter som har gyldige verdier alle disse fire variablene er inkludert i analyseutvalget. Analyseutvalget teller 1 376 personer.

Tabell 1 viser relevant deskriptiv statistikk.

Stata-output 1: Deskriptiv statistikk.

```
. summarize EDUC WOMAN AGE
```

Variable	Obs	Mean	Std. Dev.	Min	Max
EDUC	1,376	13.92442	2.962571	0	20
WOMAN	1,376	.556686	.4969569	0	1
AGE	1,376	49.25291	16.3925	20	88

```
. ta ALTMED
```

ALTERNATIVE MEDICINE PROVIDES BETTER SOLUTIONS	Freq.	Percent	Cum.
Strongly agree	86	6.25	6.25
Agree	323	23.47	29.72
Neither agree nor disagree	547	39.75	69.48
Disagree	359	26.09	95.57
Strongly disagree	61	4.43	100.00
Total	1,376	100.00	

Forskeren ønsker å bruke kontrollvariabelmetoden for å teste en teori om at høyere utdanning gjør at folk blir mer negative til alternativ medisin.

1. Tegn et kausaldiagram for denne hypotesen. Skisser kort to mekanismer som kan gjøre at høyere utdanning påvirker holdninger til alternativ medisin.

Stata-output 2: Modell 1a

```
. reg ALTMED EDUC
```

Source	SS	df	MS	Number of obs	=	1,376
Model	34.6808859	1	34.6808859	F(1, 1374)	=	38.58
Residual	1235.17667	1,374	.898964099	Prob > F	=	0.0000
				R-squared	=	0.0273
				Adj R-squared	=	0.0266
Total	1269.85756	1,375	.92353277	Root MSE	=	.94814

ALTMED	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
EDUC	.0536074	.0086308	6.21	0.000	.0366765 .0705384
_cons	2.243373	.1228669	18.26	0.000	2.002346 2.4844

2. Hvilket målenivå er variabelen ALTMED på? Hva antar forskeren ved å bruke denne som avhengig variabel i en lineær regresjon?

3. Beskriv hvordan holdningene til alternativ medisin endrer seg når utdanning øker med ett år. Hvor sannsynlig er det å få dette resultatet hvis det ikke er noen sammenheng mellom utdanningsnivå og holdninger til alternativ medisin i populasjonen?

4. Hva er den gjennomsnittlige holdningen til alternativ medisin blant respondenter som har 12 års utdannelse?

5. Diskuter fordeler og ulemper med å inkludere utdanning målt i antall år på denne måten. Foreslå minst en annen måte å kode (parameterisere) utdanning på.

Stata-output 3: Modell 1b

```
. reg ALTMED EDUC WOMAN
```

Source	SS	df	MS	Number of obs	=	1,376
Model	42.6621187	2	21.3310593	F(2, 1373)	=	23.87
Residual	1227.19544	1,373	.893805855	Prob > F	=	0.0000
				R-squared	=	0.0336
				Adj R-squared	=	0.0322
Total	1269.85756	1,375	.92353277	Root MSE	=	.94541

ALTMED	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
EDUC	.0538094	.0086063	6.25	0.000	.0369265 .0706922
WOMAN	-.1533128	.0513056	-2.99	0.003	-.2539586 -.0526669
_cons	2.325909	.1255887	18.52	0.000	2.079542 2.572275

6. Hva er motivasjonen for å legge til dummyvariabelen "WOMAN" i Modell 1b, gitt forskerens problemstilling?

7. Endrer sammenhengen mellom utdanningsnivå og holdninger til alterna-

tiv medisin seg når en holder kjønn konstant? Gjør en signifikanstest.

8. Forklart kort hva R^2 er. Hvordan endrer R^2 seg fra modell 1a til modell 1b? Kan endringen i R^2 fra Modell 1a til Modell 1b i seg selv tolkes som at Modell 1b er en bedre modell, gitt forskerens problemstilling?

Stata-output 4: Modell 1c og deskriptiv statistikk for AGE_CAT

```
. ta AGE_CAT
```

AGE_CAT	Freq.	Percent	Cum.
30	217	15.77	15.77
40	249	18.10	33.87
50	253	18.39	52.25
60	279	20.28	72.53
70	233	16.93	89.46
80	145	10.54	100.00
Total	1,376	100.00	

```
. reg ALTMED EDUC WOMAN i.AGE_CAT
```

Source	SS	df	MS	Number of obs	=	1,376
Model	62.5258241	7	8.93226059	F(7, 1368)	=	10.12
Residual	1207.33173	1,368	.882552437	Prob > F	=	0.0000
Total	1269.85756	1,375	.92353277	R-squared	=	0.0492
				Adj R-squared	=	0.0444
				Root MSE	=	.93944

ALTMED	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
EDUC	.0539843	.0085872	6.29	0.000	.0371388 .0708297
WOMAN	-.1584786	.0510731	-3.10	0.002	-.2586686 -.0582887
AGE_CAT					
40	.1969806	.0872608	2.26	0.024	.0258011 .3681601
50	.2389352	.0870366	2.75	0.006	.0681955 .4096748
60	.1473569	.0850494	1.73	0.083	-.0194845 .3141984
70	.3006393	.0886706	3.39	0.001	.1266942 .4745844
80	.4307377	.1008642	4.27	0.000	.2328724 .6286029
_cons	2.120595	.1374972	15.42	0.000	1.850867 2.390323

9. Til slutt legger forskeren til en kontroll for alder. Forklar kort hvordan forskjeller i alder kan konfundere sammenhengen mellom utdanning og holdninger til alternativ medisin. Er en modell som inkluderer settet av dummyvariable for alder signifikant bedre enn en modell som utelater dette settet av 5 dummyvariable?

Stata-output 5: Modell 1d

```
. reg ALTMED EDUC WOMAN AGE
```

Source	SS	df	MS	Number of obs	=	1,376
Model	56.885504	3	18.9618347	F(3, 1372)	=	21.45
Residual	1212.97205	1,372	.884090418	Prob > F	=	0.0000
				R-squared	=	0.0448
				Adj R-squared	=	0.0427
Total	1269.85756	1,375	.92353277	Root MSE	=	.94026

ALTMED	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
EDUC	.0542347	.00856	6.34	0.000	.0374426 .0710269
WOMAN	-.1543244	.0510266	-3.02	0.003	-.2544231 -.0542257
AGE	.006205	.001547	4.01	0.000	.0031703 .0092398
_cons	2.014934	.1470104	13.71	0.000	1.726544 2.303323

10. I modell 1d har forskeren byttet ut AGE_CAT med AGE (alder i år). Forklar hvorfor de to modellene ikke kan sammenliknes ved hjelp av en F-test.

11. Forskeren konkluderer med at Modell 1d viser en kausal sammenheng mellom utdanningsnivå og holdninger til alternativ medisin. Er du enig i konklusjonen? Hvis ikke, kan resultatene likevel gi oss nyttig informasjon om samfunnet?

Stata-output 6: BIC og AIC for Modell 1c og Modell 1d

```
. di "Modell 1c"
Modell 1c
. qui: reg ALTMED EDUC WOMAN i.AGE_CAT
. estat ic
Akaike's information criterion and Bayesian information criterion
```

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	1,376	-1897.23	-1862.491	8	3740.982	3782.798

Note: N=Obs used in calculating BIC; see [R] BIC note.

```
. di "Modell 1d"
Modell 1d
. qui: reg ALTMED EDUC WOMAN AGE AGE_SQ
. estat ic
Akaike's information criterion and Bayesian information criterion
```

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	1,376	-1897.23	-1865.698	4	3739.395	3760.303

Note: N=Obs used in calculating BIC; see [R] BIC note.

12. Tabell 5 viser BIC og AIC for Modell 1c og d. Hvilken spesifikkasjon

(koding) av alder vil du råde forskeren til å velge? Hvorfor er en best mulig spesifisering av kontrollvariabelen alder viktig hvis vi ønsker å fange opp effekten av antall års utdanning på holdninger til alternativ medisin?

Forskeren lager så en dummyvariabel for å være positiv til alternativ medisin, kodet 1 for respondenter som er positive eller svært positive til alternativ medisin. I Modell 2 er denne brukt som avhengig variabel. De uavhengige variablene er de samme som i Modell 1d.

Stata-output 7: Modell 2

```
.
. gen PRO_ALT_DUM = ALTMED == 1 | ALTMED == 2
.
. reg PRO_ALT_DUM EDUC WOMAN i.AGE_CAT, robust
Linear regression              Number of obs   =       1,376
                              F(7, 1368)         =         7.70
                              Prob > F           =         0.0000
                              R-squared          =         0.0354
                              Root MSE       =         .45018
```

PRO_ALT_DUM	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
EDUC	-.0235983	.0040641	-5.81	0.000	-.0315708	-.0156258
WOMAN	.0404053	.0242502	1.67	0.096	-.0071663	.0879769
AGE_CAT						
40	-.1078924	.0433403	-2.49	0.013	-.1929132	-.0228717
50	-.0931792	.0428105	-2.18	0.030	-.1771606	-.0091979
60	-.0512203	.0427919	-1.20	0.232	-.1351651	.0327245
70	-.1085798	.0434448	-2.50	0.013	-.1938055	-.0233541
80	-.1664942	.0471893	-3.53	0.000	-.2590655	-.0739229
_cons	.6863109	.067847	10.12	0.000	.5532155	.8194062

13. Forskeren har nå valgt å bruke robuste standardfeil. Forklar hvorfor dette nå er nødvendig.

14. Hva slags regresjonsmodell er Modell 2? Forklar kort fordeler og ulemper ved å bruke en slik modell for et dikotomt utfall, sammenliknet med logistisk regresjon.

15. Tolk og signifikanstest koeffisienten for EDUC. Forklar hvordan tolkningen av denne avviker fra tolkningen av EDUC i Modell 1d.