

# SKOLEEKSAMEN I

## SOS4020 - KVANTITATIV METODE

**27. februar 2017**

**(4 timer)**

Tillatte hjelpemidler:           Alle skriftlige hjelpemidler og kalkulator.

Sensur for eksamen faller 20. mars 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 **8 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 14 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 sammenhengen mellom familiesituasjon og småkriminalitet blant unge. Nærmere bestemt vil hun se på hvorvidt foreldres skilsmisse påvirker hvorvidt unge nasker, og hvor mye de nasker.

Forskeren bruker data fra spørreundersøkelsen “Second International Self-Reported Delinquency Study, 2005-2007” – offentlig tilgjengelig gjennom ICPSR (<https://www.icpsr.umich.edu/icpsrweb/>). Dette er en spørreundersøkelse om avvik blant ungdom i alderen 12-15 år, med respondenter fra 31 land, hovedsakelig i Europa, Nord- og Sør-Amerika. Dataene ble samlet inn i perioden 2005-2007, og enheten i utvalget er individer. Dette er et tverrsnittdatasett.

Forskeren benytter følgende variable:

- **SHOPLYP:** Kodet 1 hvis respondenten noen ganger har nasket, ellers 0.
- **SHOPLYI:** Verdien av varene respondenten har nasket siste år. Inntekten fra nasking er målt i USD.
- **DIVORCED:** Kodet 1 hvis respondentens foreldre var skilt/separert da spørreskjemaet ble besvart, 0 hvis foreldrene var gift.
- **WEEKDINNERS:** Antall middager respondenten spiser med familien sin hver uke.
- **FEMALE:** En dummyvariabel kodet 1 for kvinnelige respondenter, ellers 0.

I undersøkelsen inkluderes kun individer som har informasjon (ikke missing) på alle de relevante variablene.

Under ser du relevant deskriptiv statistikk.

#### Stata-output 1: Deskriptiv statistikk for hovedutvalg.

```
. su SHOPLYP DIVORCED FEMALE WEEKDINNERS
```

Variable	Obs	Mean	Std. Dev.	Min	Max
SHOPLYP	65,871	.0515553	.2211292	0	1
DIVORCED	65,871	.2046728	.4034654	0	1
FEMALE	65,871	.5099816	.4999042	0	1
WEEKDINNERS	65,871	5.53201	2.266301	0	7

Forskeren begynner prosjektet med å estimere en logistisk regresjonsanalyse der SHOPLYP er avhengig variabel og DIVORCED forklaringsvariabel. Resultatene vises i Stata-output 2.

1. Hva er oddsen for å naske blant de som har skilte foreldre, i følge Modell 1 (Stata-output 2)?

### Stata-output 2: Modell 1

```
. logit SHOPLYP DIVORCED
Iteration 0:  log likelihood = -13376.394
Iteration 1:  log likelihood = -13274.498
Iteration 2:  log likelihood = -13269.862
Iteration 3:  log likelihood = -13269.855
Iteration 4:  log likelihood = -13269.855

Logistic regression              Number of obs   =    65,871
                                LR chi2(1)      =    213.08
                                Prob > chi2     =    0.0000
                                Pseudo R2      =    0.0080

Log likelihood = -13269.855
```

SHOPLYP	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
DIVORCED	.5827652	.0384961	15.14	0.000	.5073142	.6582162
_cons	-3.058387	.0211073	-144.90	0.000	-3.099757	-3.017018

2. Hva er sannsynligheten for å naske blant de som har gifte foreldre?
3. Hva er oddsratioen for at kvinner nasker (relativt til menn), når foreldres samlivsstatus holdes konstant, i følge Modell 2?

### Stata-output 3: Modell 2

```
.
. logit SHOPLYP DIVORCED FEMALE
Iteration 0:  log likelihood = -13376.394
Iteration 1:  log likelihood = -13245.701
Iteration 2:  log likelihood = -13240.039
Iteration 3:  log likelihood = -13240.03
Iteration 4:  log likelihood = -13240.03

Logistic regression              Number of obs   =    65,871
                                LR chi2(2)      =    272.73
                                Prob > chi2     =    0.0000
                                Pseudo R2      =    0.0102

Log likelihood = -13240.03
```

SHOPLYP	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
DIVORCED	.5924427	.0385418	15.37	0.000	.5169022	.6679832
FEMALE	-.2733949	.035502	-7.70	0.000	-.3429775	-.2038124
_cons	-2.929305	.0262915	-111.42	0.000	-2.980835	-2.877774

4. Har en modell med kontroll for kjønn signifikant bedre prediksjonskraft enn en modell uten slik kontroll? Gjør en LR-test vha. log likelihood-verdier

fra Modell 1 og 2.

5. Bruk resultatene i Modell 2 og Modell 3 til å vise at FEMALE er en nødvendig kontrollvariabel i en modell der SHOPLY er avhengig variabel og DIVORCE er forklaringsvariabel. Trekk gjerne inn formelen for Omitted Variable Bias, og knytt denne til et kausaldiagram.

#### Stata-output 4: Modell 3

```
. logit DIVORCED FEMALE
Iteration 0:  log likelihood = -33384.241
Iteration 1:  log likelihood = -33360.656
Iteration 2:  log likelihood = -33360.647
Iteration 3:  log likelihood = -33360.647 (backed up)
Logistic regression               Number of obs   =    65,871
                                LR chi2(1)       =     47.19
                                Prob > chi2        =     0.0000
Log likelihood = -33360.647      Pseudo R2       =     0.0007
```

DIVORCED	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
FEMALE	.1328466	.0193552	6.86	0.000	.0949111	.170782
_cons	-1.426392	.0140853	-101.27	0.000	-1.453999	-1.398786

Forskeren går videre til å analysere inntekten fra nasking *blant de som nasker*. Samplet blir altså mindre, i og med at vi kaster ut de som ikke har denne typen inntekter. Stata-output 5 viser deskriptiv statistikk for det begrensede samplet.

#### Stata-output 5: Deskriptiv statistikk for subsample som har nasket.

```
. su SHOPLYI DIVORCED FEMALE WEEKDINNERS if SHOPLY != 0
```

Variable	Obs	Mean	Std. Dev.	Min	Max
SHOPLYI	3,396	5.880742	16.0982	1	300
DIVORCED	3,396	.3080094	.4617385	0	1
FEMALE	3,396	.4490577	.4974714	0	1
WEEKDINNERS	3,396	4.985277	2.473736	0	7

6. Modell 4 (Stata-output 6) er estimert ved hjelp av en annen regresjonsteknikk enn modell 1-3. Hvilken regresjonsteknikk benyttes nå? Foreslå en begrunnelse for dette skiftet av regresjonsteknikk.

7. Hva er gjennomsnittlig inntekt fra nasking blant naskere som har skilte foreldre (iflg. Modell 4)?

8. Hva er forskjellen i gjennomsnittlig inntekt i nasking mellom naskere som har skilte foreldre og naskere som har gifte foreldre (iflg. Modell 4)?

**Stata-output 6: Modell 4**

```
. reg SHOPLYI DIVORCED if SHOPLYP != 0
```

Source	SS	df	MS	Number of obs	=	3,396
Model	1433.83146	1	1433.83146	F(1, 3394)	=	5.54
Residual	878386.869	3,394	258.805795	Prob > F	=	0.0186
Total	879820.701	3,395	259.1519	R-squared	=	0.0016
				Adj R-squared	=	0.0013
				Root MSE	=	16.087

  

SHOPLYI	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
DIVORCED	1.40745	.5979585	2.35	0.019	.2350552 2.579846
_cons	5.447234	.3318586	16.41	0.000	4.796571 6.097897

Forskeren undersøker en hypotese om at det ikke er skilsmisse i seg selv som påvirker inntekten fra nasking, men heller at barn med skilte foreldre tilbringer mindre tid med foreldrene sine. Som en proxy for tid med foreldre bruker forskeren variabelen 'WEEKDINNERS', som måler hvor mange ganger i uka ungdommen spiser middag med en eller flere foreldre.

9. Diskuter kort styrker og svakheter ved WEEKDINNERS som et mål på tid brukt med foreldrene.

10. Sammenlikn resultatene fra Modell 4 og Modell 5. Er det støtte for at sammenhengen mellom skilsmisse og inntekt fra nasking skyldes at ungdommer med skilte foreldre tilbringer mindre tid med foreldrene sine?

**Stata-output 7: Modell 5**

```
. reg SHOPLYI DIVORCED FEMALE WEEKDINNERS if SHOPLY != 0
```

Source	SS	df	MS	Number of obs	=	3,396
Model	6339.3545	3	2113.11817	F(3, 3392)	=	8.21
Residual	873481.346	3,392	257.512189	Prob > F	=	0.0000
				R-squared	=	0.0072
				Adj R-squared	=	0.0063
Total	879820.701	3,395	259.1519	Root MSE	=	16.047

  

SHOPLYI	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
DIVORCED	1.527644	.5998523	2.55	0.011	.3515351	2.703752
FEMALE	-2.400746	.5553927	-4.32	0.000	-3.489684	-1.311808
WEEKDINNERS	-.0911636	.111907	-0.81	0.415	-.3105756	.1282485
_cons	6.942762	.7136012	9.73	0.000	5.54363	8.341894

Forskeren vil også undersøke om betydningen av foreldres skilsmisse for inntekt fra nasking er ulik for gutter og jenter. Forskeren konstruerer derfor et interaksjonsledd, og estimerer en ny regresjonsmodell der dette er inkludert.

11. Tolk koeffisienten til interaksjonsleddet. Er koeffisienten statistisk signifikant på 10%-nivå? Gjør en t-test.

12. Er det en statistisk signifikant sammenheng mellom å ha skilte foreldre og naske for gutter? Begrunn svaret.

### Stata-output 8: Modell 6

```
. gen INTER_FEM_DIV = FEM* DIVORCED
. reg SHOPLYI DIVORCED FEMALE WEEKDINNERS INTER_FEM_DIV if SHOPLYI != 0
```

Source	SS	df	MS	Number of obs	=	3,396
Model	7326.1772	4	1831.5443	F(4, 3391)	=	7.12
Residual	872494.523	3,391	257.297117	Prob > F	=	0.0000
				R-squared	=	0.0083
				Adj R-squared	=	0.0072
Total	879820.701	3,395	259.1519	Root MSE	=	16.04

  

SHOPLYI	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
DIVORCED	.4161935	.8255975	0.50	0.614	-1.202526 2.034913
FEMALE	-3.132211	.6691086	-4.68	0.000	-4.444108 -1.820314
WEEKDINNERS	-.0883839	.1118693	-0.79	0.430	-.307722 .1309541
INTER_FEM_DIV	2.343072	1.196419	1.96	0.050	-.0027047 4.688848
_cons	7.241628	.7294452	9.93	0.000	5.811431 8.671825

13. Skisser hvordan en modell med faste effekter kunne brukes til å komme nærmere kausal identifikasjon av hvorvidt skilsmisse påvirker sannsynligheten for nasking. Kan forskeren estimere en slik modell med dette datasettet?

14. Tilslutt estimerer forskeren Modell 7, som et alternativ til Modell 5. Forklar kort forskjellen på det to modellene. Tolk koeffisientene for WEEKDINNERS i Modell 7, og sammenlikn med tolkningen av koeffisienten for samme variabel i Modell 5. Foreslå forklaringer på at en t-test av regresjonskoeffisienten for WEEKDINNERS gir ulikt resultat i Modell 5 og Modell 7.

**Stata-output 9: Modell 7**

```
. gen LN_SHOPLYI = ln(SHOPLYI)
(62,475 missing values generated)
. reg LN_SHOPLYI DIVORCED FEMALE WEEKDINNERS if SHOPLY != 0
```

Source	SS	df	MS	Number of obs	=	3,396
Model	64.7644326	3	21.5881442	F(3, 3392)	=	20.65
Residual	3545.6499	3,392	1.04529773	Prob > F	=	0.0000
				R-squared	=	0.0179
				Adj R-squared	=	0.0171
Total	3610.41433	3,395	1.06345047	Root MSE	=	1.0224

  

LN_SHOPLYI	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
DIVORCED	.1089748	.0382178	2.85	0.004	.0340426 .183907
FEMALE	-.2562158	.0353852	-7.24	0.000	-.3255942 -.1868374
WEEKDINNERS	-.014491	.0071298	-2.03	0.042	-.0284702 -.0005119
_cons	1.10519	.0454649	24.31	0.000	1.016048 1.194331