

# SKOLEEKSAMEN I

## SOS1120 Kvantitativ metode

**13. desember 2013**

**4 timer**

Det er lov å bruke ikke-programmerbar kalkulator som hjelpemiddel

Sensur for eksamen faller 10.januar 2014 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 **8 sider** inkludert denne.

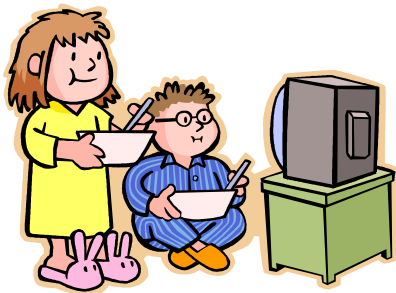
Kandidaten skal levere både originalen og kopien av besvarelsen. Kladd skal ikke leveres.

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

Husk å notere deg kandidatnummeret ditt.

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.

## Temaet på denne eksamen er tv-titting



Tv-titting kan betraktes som et av mange mulige mål på kulturelt konsum i dagens globale tidsalder. Men hvor mye ser vi på tv? Og er det forskjeller i tv-titting etter sosial bakgrunn og demografiske forhold (som familiestørrelse, alder, utdanning og kjønn)? Kort sagt: Hvem benytter seg av «friheten» til å kikke på tv?

Ved hjelp av den amerikanske spørreundersøkelsen General Social Survey (GSS) kommer vi til å undersøke tv-tittingsvaner i året 2012. Utvalget er representativt for den amerikanske populasjonen og består av 1974 respondenter.

### Variabelbeskrivelse (gjelder alle oppgaver i denne eksamen):

Variabelnavn	Betydning	Koding
tvhours	Antall timer med tv-titting per dag	Målt i faktiske timer (0 betyr «0 timer med tv-titting»)
tvtimer2kat	Dikotom variabel for antall timer med tv-titting per dag.	Delt i to kategorier (dikotom): 1 = «0-3 timer» 2 = «over 3 timer»
gender	Respondentens kjønn	Mann = 0, kvinne = 1
educ	Respondentens utdanning målt i antall år utover ingen skolegang	Målt i antall år. (0 betyr «0 år med skole»)
parenteduc	Foreldrenes høyeste utdanning, målt i antall år utover ingen skolegang	Målt i antall år. (0 betyr «0 år med skole»)
farsutd2kat	Fars utdanning (fathers highest degree), <b>delt i to kategorier</b>	1 = «mindre enn universitet», 2 = «grad fra universitet»
age20	Respondentens alder. Variabelen er sentrert på 20 år	Alder er sentrert på 20. (i.e. 0 betyr «20 år»)
aldersgrp	Respondentens alder delt inn i fem kategorier	1 = «20-25 år» 2 = «26-35 år» 3 = «36-50 år» 4 = «51-70 år» 5 = «71-100 år»
educXgender	Samspillsledd mellom utdanning og kjønn	Kjønn (gender) multiplisert med utdanning (educ)

### Oppgave 1-1 (1 poeng)

Anta at en forsker ønsker å ha et generelt mål på kulturelt konsum. Tv-titting måler kun et av flere aspekter av kulturelt konsum. Nevn minst tre alternative mål som forskeren kunne ha inkludert i et mer generelt mål på kulturelt konsum dersom han/hun sto fritt til å lage spørsmål til en spørreundersøkelse.

**Oppgave 1-2 (3 poeng)**

**Tabell 1: Frekvenstabell av aldersgrupper (aldersgrp)**

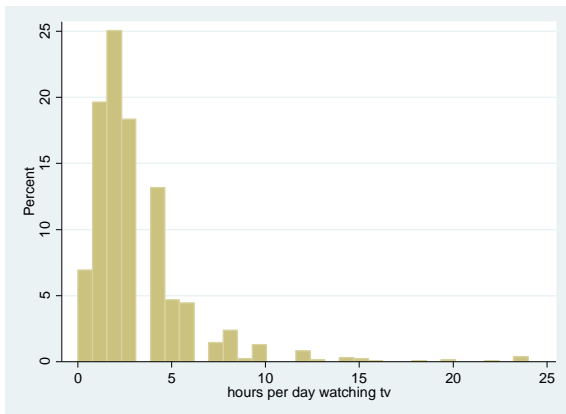
. ta aldersgrp

RECODE of age (age of respondent)	Freq.	Percent	Cum.
20-25 år	201	10.21	10.21
26-35 år	367	18.64	28.85
35-50 år	541	27.48	56.32
51-70 år	607	30.83	87.15
71-100 år	253	12.85	100.00
Total	1,969	100.00	

- a) Hvilket målenivå befinner variabelen «aldersgrp» seg på?
- b) Hvilken alderskategori er *median* i tabell 1?
- c) Tolke *modus* i tabell 1

**Oppgave 1-3 (1 poeng)**

Figur 1: Histogram av antall timer med tv-titting per dag (tvhours)



Kikk på figur 1. Figuren viser fordelingen av antall timer med tv-titting. Vurder om fordelingen av timer med tv-titting per dag er (angi gi kun en bokstav i svaret ditt):

- a) høyreskjev
- b) venstreskjev
- c) normalfordelt
- d) t-fordelt
- e) z-fordelt
- f) uniform

**Oppgave 1-4 (1 poeng)**

Tabell 2: Beskrivende statistikk for antall timer med tv-titting per dag. Mean= gjennomsnitt, p50= median, sd= standardavvik, se(mean) = standardfeil

. tabstat tvhours, stat(mean median sd sdmean min max N )

variable	mean	p50	sd	se(mean)	min	max	N
tvhours	3.088598	2	2.8651	.0795248	0	24	1298

Gi en kort begrunnelse for ditt svar i oppgave 1-3 basert på resultatene for gjennomsnitt og median i tabell 2.

**Oppgave 1-5 (2 poeng)**

Bruk tall fra tabell 2. Finn kritisk t-verdi i vedleggstabellen for et signifikansnivå på 99% (alfa= 0.01). Frihetsgrader er gitt ved  $df = N-1$ .

a. Regn ut 99% - konfidensintervallet basert på formelen (1 poeng):

$$KI_{99\%}: \text{mean} \pm t^*SE(\text{mean})$$

b. Gi også en kort tolkning av konfidensintervallet (1 poeng)

**Oppgave 1-6 (1 poeng)**

Ifølge norske analyser kikker nordmenn på tv ca 2 timer daglig (basert på European Social Survey data). Basert på konfidensintervallet, kan vi konkludere med at amerikanerne ser på mer tv enn 2 timer per dag? Angi: ja / nei

**Oppgave 2-1 (2 poeng)**

Kikk på tallrekken nedenfor. Anta at verdiene i denne tallrekken angir antall svitsjinger mellom ulike tv-kanaler i løpet av en time oppgitt av 7 ulike personer.

{ 0,1,2,3,3,5,6 }

Finn modus, median og gjennomsnitt i denne tallrekken. Vis utregning der det er relevant.

**Oppgave 2-2 (1 poeng)**

Anta at vi hadde spurt en person ekstra, og at denne personen fortalte oss at han/hun så svitsjet mellom 100 kanaler i timen. Angi hvilke av disse målene på sentraltendens som i liten grad hadde blitt påvirket av den nye verdien i tallrekken (flere svar er mulig) : median, modus, gjennomsnitt

**Oppgave 2-3 (1 poeng)**

Begrunn kort hvorfor dette/disse målene blir i liten grad påvirket av den nye verdien.

**Oppgave 3 (2 poeng)**

Anta at en forsker har foretatt et randomisert eksperiment. Hun har undersøkt tv-titting blant universitetsstudenter i Oslo. En testgruppe blir informert om negative konsekvenser av mye tv-titting. En kontrollgruppe blir ikke informert. Forskeren finner at de som har vært i testgruppen titter mindre på tv i etterkant av tiltaket enn de som var i kontrollgruppen. Forskeren konkluderer med at informasjonen om negative konsekvenser av mye tv-titting har en kausal effekt på tv-tittingsatferd.

Kan konklusjonen om årsakssammenheng generaliseres til resten av befolkningen? Begrunn svaret ditt kort.

**Oppgave 4**

Tabell 3 nedenfor viser sammenhengen mellom fars utdanning (farsutd2kat) og tv-timer (tvtimer2kat). Begge variablene er målt med to kategorier.

Tabell 3: fars utdanning og tv-timer

. ta farsutd2kat tvtimer2kat, row nof chi v

RECODE of padeg (fathers highest degree)	RECODE of tvhours (hours per day watching tv)		Total
	0-3 timer	over 3 ti	
Univ utd	84.39	15.61	100.00
Ikke univ utd	68.89	31.11	100.00
Total	72.07	27.93	100.00

Pearson chi2(1) = 19.4440 Pr = 0.000  
Cramér's V = 0.1395

- 4-1) Kikk på prosentueringsretningen i tabell 3. Hva er uavhengig og avhengig variabel i denne tabellen? (1 poeng)
- 4-2) Regn ut prosentdifferansen i denne tabellen. Vis utrekningen. (1 poeng)
- 4-3) Gi en kort tolkning av prosentdifferansen (1 poeng)
- 4-4) Tolk korrelasjonsmålet Cramers V (1 poeng)
- 4-5) Formuler en nullhypotese tilpasset en Kjikvadrattest (1 poeng)
- 4-6) Regn ut frihetsgrader for tabellen:  $df = (\text{antall kolonner}-1) * (\text{antall rader}-1)$ . (1 poeng)
- 4-7) Velg signifikansnivå 95% (alfa = 0.05). Finn kritisk kji-kvadratverdi i vedlagt tabell bakerst. Kan vi forkaste nullhypotesen med en klassisk kjikvadrattest? Begrunn kort. (1 poeng)

### Oppgave 5

Nedenfor ser du en t-test for gjennomsnittsforskjeller mellom to grupper.

. tttest tvhours, by(gender) uneq

Two-sample t test with unequal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
male	591	3.128596	.1289519	3.134885	2.875335	3.381856
female	707	3.055163	.0985412	2.620157	2.861694	3.248631
combined	1298	3.088598	.0795248	2.8651	2.932587	3.244609
diff		.0734329	.1622928		-.2449897	.3918556

diff = mean(male) - mean(female) t = 0.4525

- a. Hvor stor er gjennomsnittsforskjellen i tv-titting mellom menn og kvinner? (1 poeng)
- b. Formuler en nullhypotese for en to-halet t-test for menn og kvinner. (1 poeng)
- c. Utfør en klassisk tohalet t-test når alfa = 0.05 og frihetsgrader er oppgitt til  $df = 591-1 = 590$ . (1 poeng)

### Oppgave 6

Vi ønsker her å undersøke sammenhengen mellom sosial bakgrunn og tv-titting (tvhours). Vi inkluderer foreldrenes utdanning (parenteduc) som et mål på sosial bakgrunn, samt to kontrollvariabler: kjønn (gender) og alder (age20).

#### Modell 1: *Multipel regresjon hvor antall timer tv-titting avhenger av kjønn (gender), foreldrenes høyeste utdanning (parenteduc) og alder(age20)*

```
. reg tvhours parenteduc age20 gender, beta
```

Source	SS	df	MS		
Model	238.451632	3	79.4838773	Number of obs =	1215
Residual	9498.4817	1211	7.84350264	F( 3, 1211) =	10.13
Total	9736.93333	1214	8.02053817	Prob > F =	0.0000
				R-squared =	0.0245
				Adj R-squared =	0.0221
				Root MSE =	2.8006

tvhours	Coef.	Std. Err.	t	P> t	Beta
parenteduc	-.0462591	.0207073	-2.23	0.026	-.0659886
age20	.0200945	.0048468	4.15	0.000	.1226164
gender	-.1980305	.1618097	-1.22	0.221	-.0347911
_cons	3.141818	.338437	9.28	0.000	.

- 6-1) Tolk konstantleddet i modell 1 (1 poeng)
- 6-2) Tolk den ustandardiserte regresjonskoeffisienten til *foreldreutdanning* i modell 1. (1 poeng)
- 6-3) Tolk den ustandardiserte regresjonskoeffisienten til *kjønn* i modell 1(1 poeng)
- 6-4) Fortolk den (ujusterte)  $R^2$  (1 poeng)
- 6-5) Avgjør om nullhypotesen « $H_0$ : modellen forklarer ingenting i tv-titting ( $R^2 = 0$ )» kan forkastes eller ikke. Bruk moderne F-test. Velg et signifikansnivå. (1 poeng)
- 6-6) Angi variabel som har *minst* betydning på utfallsvariabelen i modell 1 (1 poeng)
- 6-7) Predikér antall timer tv-titting per dag for en kvinne som er 21 år gammel, og som har foreldre med 10 år som høyeste utdanning. Vis utrekning (2 poeng)

### Oppgave 7

Vi kjører en ny regresjonsmodell (modell 2) hvor vi bygger ut modell 1 med en ekstra kontrollvariabel: respondentenes egen utdanning (educ).

**Modell 2: *Multipel regresjon som inkluderer i tillegg en kontrollvariabel for respondentenes egen utdanning (educ)***

. reg tvhours parenteduc age20 gender educ, beta

Source	SS	df	MS		
Model	537.100978	4	134.275244	Number of obs =	1215
Residual	9199.83236	1210	7.60316724	F( 4, 1210) =	17.66
Total	9736.93333	1214	8.02053817	Prob > F =	0.0000
				R-squared =	0.0552
				Adj R-squared =	0.0520
				Root MSE =	2.7574

tvhours	Coef.	Std. Err.	t	P> t	Beta
parenteduc	.0262597	.0234423	1.12	0.263	.0374595
age20	.0249563	.0048346	5.16	0.000	.1522836
gender	-.2063437	.1593169	-1.30	0.196	-.0362516
educ	-.1866572	.0297825	-6.27	0.000	-.2013918
_cons	4.682271	.4140566	11.31	0.000	.

7-1) Bruk relevante mål for å vurdere om modell 1 eller 2 forklarer mest variasjon i den avhengige variabelen (1 poeng)

7-2) Hva er den mest sannsynlige grunnen til at betydningen av foreldrenes utdanning synker og blir statistisk insignifikant i modell 2? (1 poeng) Angi bokstav for ett av følgende alternativer:

- a) Foreldreutdanning virker indirekte på tv-titting via egen utdanning
- b) Det er en større betydning av egen utdanning på tv-titting blant de med høyt utdannede foreldre
- c) Det er utelatte variabler som påvirker både høy foreldreutdanning og egen utdanning
- d) Foreldreutdanning har en betydning på tv-titting som er uavhengig av sammenhengen mellom egen utdanning og tv-titting

7-3) Kan vi på grunnlag av denne regresjonsmodellen slutte til at mer utdanning er en årsak til mer tv-titting? (1 poeng) Angi bokstav for ét av følgende svaralternativer:

- a) Ja, fordi det er en statistisk signifikant sammenheng
- b) Nei, fordi det kan være en kurvilinear sammenheng mellom utdanning og tv-titting
- c) Nei, fordi tv-titting også kan avgjøre hva slags utdanning du har tatt tidligere
- d) Ja, fordi beta også viser at utdanning har den sterkeste effekten på tv-titting
- e) Nei, fordi utdanning kan være bestemt av kulturelle preferanser som også er korrelert med tv-titting.

**Oppgave 8 Samspillsledd**

Anta at du har kjørt en ny regresjonsanalyse på samme data hvor antall tv-timer er avhengig av kjønn og utdanning (educ). Det er også inkludert et statistisk signifikant *samspillsledd* (educXgender) i modellen. Resultatet er oppstilt slik:

$$tv\widehat{titting} = 4 - 0,05 * educ - 1 * gender - 0,05 * educXgender$$

- 8-1) Hvor *mye mer* reduseres betydningen av utdanning på tv-titting for kvinner enn for menn? (2 poeng)
- 8-2) Utled to nye regresjonslikninger ut fra regresjonslikningen over: én likning for kvinner og en likning for menn. (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

**Kjikkvadratfordelingen [Chi-square distribution] (forkortet tabell)**

df	Signifikansnivåer for kjikkvadrattest (alfa)						
	0.5	0.25	0.1	0.05	0.025	0.01	0.001
1	0.45	1.32	2.71	3.84	5.02	6.63	10.83
5	4.35	6.63	9.24	11.07	12.83	15.09	20.52
10	9.34	12.55	15.99	18.31	20.48	23.21	29.59
15	14.34	18.25	22.31	25.00	27.49	30.58	37.70
20	19.34	23.83	28.41	31.41	34.17	37.57	45.31
30	29.34	34.80	40.26	43.77	46.98	50.89	59.70
40	39.34	45.62	51.81	55.76	59.34	63.69	73.40
60	59.33	66.98	74.40	79.08	83.30	88.38	99.61
80	79.33	88.13	96.58	101.88	106.63	112.33	124.84
>100	99.33	109.14	118.50	124.34	129.56	135.81	149.45