



SKOLEEKSAMEN MET4

Høst, 2019

Dato: 21.11.2019

Tidsrom: 09:00 – 12.00

Antall timer: 3

Foreleser/emneansvarlig kan kontaktes av eksamensvakt på telefon: 55959902

TILLATTE HJELPEMIDLER:

Alle trykte/egenskrevne, kalkulator

Ordbok: én tospråklig ordbok tillatt

Antall sider, inkludert forside: 9

Instruksjoner: Bruk tester med 5 % signifikans. Rund av tall til to desimaler med mindre noe annet er presisert.

Sosiale normer varierer betydelig fra land til land. Den europeiske samfunnsundersøkelsen (European Social Survey – ESS) er en tverrnasjonal spørreundersøkelse som måler holdninger, oppfatninger og atferdsmønstre i varierte populasjoner i mer enn 30 land. I denne eksamenen vil vi undersøke data fra den åttende ESS-runden, som ble gjennomført i 2016–2017. Hver rad rådata representerer én respondents svar.

For å registrere en verdi for variabelen **lykkelig** ber intervjueren respondentene om å vurdere hvor lykkelige de er på en skala fra 1 til 10, der 10 er svært lykkelig og 1 er svært ulykkelig. Tabell 1 beskriver variablene i analysen. I tabell 2 finner du sammendragstatistikk for variabelen **lykkelig** for forskjellige land. I del (B) av tabell 2 kan du se svarene for variabelen lykkelig ordnet i tabellform. For eksempel har 1 norsk, 7 italienske og 20 russiske respondenter vurdert sin lykke til nivå 1, som blir tolket som "minst lykkelig". Nivå 2 betegner et noe høyere lykkenivå enn nivå 1, og så videre. Nivå 10 velges av respondenter som er svært lykkelige.

La oss kalle dem som rapporterer nivå 10 for **kjempelykkelige**. La verdiene 7, 8 og 9 betegne **forsiktige optimister**.

- Hvilken test/hvilke tester ville du kunne utføre for å finne ut hvilket land som har den signifikant største andelen **kjempelykkelige** mennesker? Hvordan ville du transformert dataene fra del (B) av tabell 2? Gi et eksempel på datatransformasjon for Norge.
- Sammenlign Norge og Italia, og finn ut om de to landene har ulike andeler **forsiktig optimistiske** mennesker ved å transformere dataene fra del (B) av tabell 2.
- Utfør en test for å finne ut om Norge og Italia har forskjellig varians for variabelen **lykkelig**.
- Utfør en test for å finne ut om det gjennomsnittlige lykkenivået i Norge og Italia er forskjellig. Bruk frihetsgrader lik uendelig ($df = \infty$).

Vi er interessert i å finne ut om lykke kan følge en uniform fordeling. Det første steget er å sammenligne den observerte fordelingen av lykke i Russland med en uniform fordeling.

- Gjør en kjikvadrattest av tilpasning ("goodness-of-fit") der russere representerer den observerte populasjonen og de forventede frekvensene alle er lik 236.

Vi vil gjøre en kjikvadrattest av en krystabell for å finne ut om det er en sammenheng mellom lykkenivå og hvilket land folk kommer fra. I del (C) av tabell 2 lager vi en tabell over antallet respondenter som er **kjempelykkelige**, **forsiktige optimister** og andre. Legg merke til at vi i denne delen av tabell 2 har endret definisjonen av hvem som er **kjempelykkelige** og hvem som er **forsiktige optimister**.

- Sett opp tabellen med de forventede frekvenser for en kjikvadrattest av en krystabell. Rund av til hele tall. Hva blir resultatet av testen hvis vi regner ut at $\chi^2 = 782$?
- Skriv ned en regresjonsmodell som kan brukes til å teste forskjeller i gjennomsnittlig lykkenivå mellom de tre landene. Definer variablene i modellen. Hvordan bør vi tolke koeffisientene i denne modellen? Skisser et eksempel på de fem første observasjonene i datasettet som trengs for å estimere denne regresjonen (ta bare med nødvendige kolonner).

I tabell 3 viser vi et sammendrag over variablene som brukes i analysen. I tabell 4 viser vi korrelasjonsmatrisen, og i tabell 5, 6 og 8 viser vi noen regresjonsestimater.

- Når vi ser på regresjonene i tabell 5 og i korrelasjonene i tabell 4, hvilke variabler vil kunne gi multikollinearitet? Når vi ser på beskrivelsene av variabler i tabell 1, finnes det begrepsmessige problemer med multikollinearitet der flere variabler sannsynligvis fanger opp samme informasjon? (Maksimalt 100 ord for hvert drøftet variabelpar.)
- Kan heteroskedastisitet og autokorrelasjon utgjøre et problem i estimeringen av kolonne (8) i tabell 5? Passer modellen godt til dataene? Kommenter kort diagnoseplottet i figur 1. (Maksimalt 200 ord.)

- j) Skisser regresjonslinjen for **råalders** samlede påvirkning på lykkenivået ved hjelp av kolonne (8) i tabell 5. Legg merke til at regresjonen er basert på logaritmen til alder, mens spørsmålet bare forutsetter alder. Gi en kort intuitiv forklaring av figuren.
- k) Vi ønsker å vite om rettferdighetspreferanser påvirker lykkenivået på ulike måter for menn og kvinner. Skriv ned regresjonsmodellen vi kan bruke for å finne ut dette ved hjelp av variablene fra tabell 6.
- l) Tostegsprediksjon: Bruk modellene i kolonne (1) og (2) av tabell 6 til å predikere menns og kvinners rettferdighetspreferanser. Bruk så verdiene du finner til å predikere hva den samlede påvirkningen det å være mann eller kvinne vil ha på lykkenivået i kolonne (3) av tabell 6. Hvem er lykkeligst i gjennomsnitt, menn eller kvinner? (Maksimalt én setning.)

La oss anta at vi gjennom data fra undersøkelsen kan observere en tidsserie for lykkenivået gjennom livet for den gjennomsnittlige nordmann, russer og italiener. Begrepsmessig er vi interessert i hva slags prosess lykke er: Er det en autoregressiv prosess (AR) eller en glidende gjennomsnitt-prosess (MA)? Med andre ord, er lykkenivået svært historieavhengig, eller er hver periode en uavhengig trekning fra lykkefordelingen?

- m) Se på autokorrelasjonsplottene i figur 2. Hva slags prosess ser lykke ut til å være for hvert av landene?
- n) I tabell 7 viser vi resultatene fra flere ARIMA-modeller tilpasset lykkenivået i Frankrike. Hvilken modell passer best? Skriv ned ligningen i modellen. Er det indikasjoner på at noen av leddene i den beste modellen ikke er signifikante? Hvordan vil den nye ligningen bli?
- o) Gitt resultatene fra b), c), d), e), f), m) og n): Hva har vi funnet ut om lykkenivået i forskjellige land? Hvilken modell i tabell 8 forklarer variasjonen i lykke best: kolonne (1) eller kolonne (2)? Hvorfor? Maksimalt 300 ord. (NB! Du kan fremdeles svare på dette spørsmålet selv om du ikke har svart på alle spørsmålene i b), c), d), e), f), m) og n).)

Tabell 1. Variabler som brukes i analysen

Variabelnavn	Beskrivelse	Koding
lykkelig	Svar på spørsmålet: Hvor lykkelig er du?	1 = Svært ulykkelig, 10 = Svært lykkelig
sosial	Omgås du med folk minst flere ganger i uka?	1 = Ja, 0 = Nei
nære_bånd	Har du minst én person som står deg nær?	1 = Ja, 0 = Nei
sosialt_aktiv	Hvor ofte deltar du i sosiale aktiviteter sammenlignet med andre på din alder?	1 = Mer enn de fleste, 5 = Mindre enn de fleste
kriminalitet	Har du eller et annet medlem av din husstand blitt utsatt for innbrudd, ran eller overfall i løpet av de siste fem årene?	1 = Ja, 0 = Nei
trygghet	Hvor trygg føler du deg, eller ville du følt deg, hvis du går alene i ditt nabolag etter at det har blitt mørkt?	1 = Veldig trygg, 4 = Veldig utrygg
helse	Subjektiv generell helsetilstand	1 = Svært god, 5 = Svært dårlig
kvinne	Indikator for respondentens kjønn	1 = Kvinne, 0 = Mann
sivilstatus	Gift eller i registrert partnerskap	1 = Ja, 0 = Nei
inntektsforskjell	Store inntektsforskjeller er akseptabelt for å belønne talent og innsats	1 = Svært enig, 5 = Svært uenig
rettferdighet	For at et samfunn skal være rettferdig, bør forskjeller i levestandard være små	1 = Svært enig, 5 = Svært uenig
alder	Respondentens alder	Kontinuerlig variabel
lalder	Logaritmen til respondentens alder	Kontinuerlig variabel

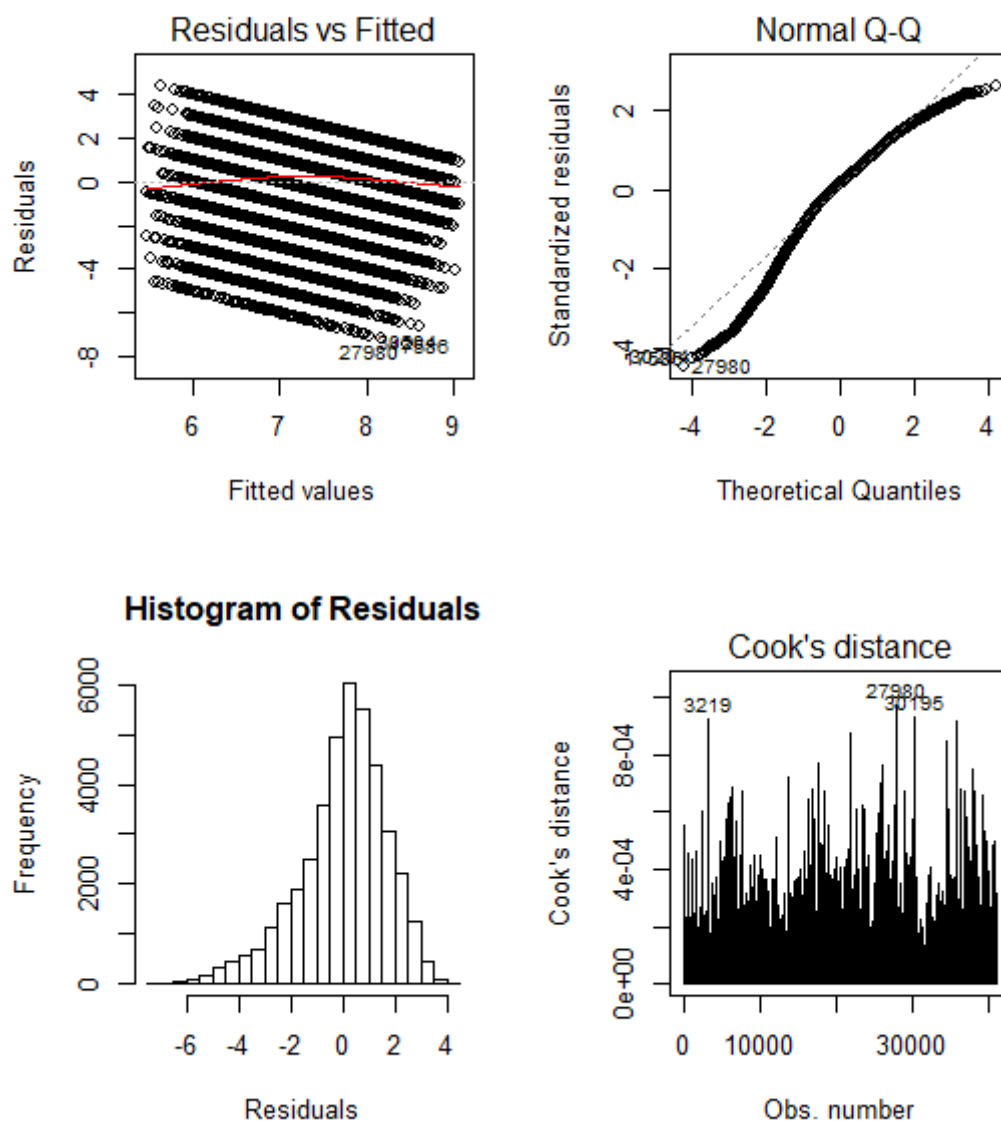
Tabell 2. Del (A). Sammendrag for variabelen <i>lykkelig</i>			
Land:	Norge	Italia	Russland
Gjennomsnitt	8,12	7,18	6,32
Standardavvik	1,47	1,66	2,05
Antall observasjoner	1541	2591	2364
Del (B). Antall svar i tabellform.			
	Norge	Italia	Russland
Svar = 1, minst lykkelig	1	7	20
2	9	25	53
3	13	68	161
4	26	69	194
5	41	198	454
6	61	356	295
7	229	655	482
8	521	732	377
9	378	299	146
Svar = 10, Svært lykkelig	262	182	182
Del (C). Krysstabell for spørsmål f)			
	Norge	Italia	Russland
Svært lykkelig	640	481	328
Forsiktige optimister	811	1743	1154
Andre	90	367	882

Tabell 3. Sammendrag				
Variabelnavn	Gjennomsnitt	Standardavvik	Min.	Maks.
lykkelig	7,51	1,77	1	10
sosial	0,41	0,49	0	1
nære_bånd	0,80	0,40	0	1
sosialt_aktiv	2,72	0,90	1	5
kriminalitet	0,15	0,36	0	1
trygghet	1,92	0,76	1	4
helse	2,17	0,91	1	5
kvinne	0,52	0,50	0	1
sivilstatus	0,02	0,15	0	1
inntektsforskjell	2,98	1,13	1	5
rettferdighet	2,40	0,95	1	5
alder	48,71	18,33	15	90
lalder	3,80	0,43	2,71	4,50

Tabell 4. Korrelasjonstabell													
	lykkelig	sosial	nære_bånd	sosialt_aktiv	kriminalitet	trygghet	helse	kvinne	sivilstatus	inntektsforskjell	rettferdighet	agea	lage
lykkelig	1,00												
sosial	0,16	1,00											
nære_bånd	0,15	0,16	1,00										
sosialt_aktiv	0,19	0,26	0,19	1,00									
kriminalitet	-0,02	0,05	0,03	0,01	1,00								
trygghet	-0,19	-0,09	-0,08	-0,08	0,09	1,00							
helse	-0,33	-0,14	-0,14	-0,19	0,03	0,21	1,00						
kvinne	0,00	-0,01	0,05	-0,02	-0,01	0,24	0,05	1,00					
sivilstatus	-0,01	-0,01	0,00	-0,01	0,00	0,00	0,00	0,00	1,00				
inntektsforskjell	-0,07	-0,01	-0,01	-0,01	0,00	0,02	0,08	0,06	0,00	1,00			
rettferdighet	0,04	0,02	0,02	0,02	0,00	-0,02	-0,06	-0,03	0,00	-0,29	1,00		
alder	-0,08	-0,17	-0,14	-0,06	-0,06	0,07	0,38	0,03	0,02	0,04	-0,07	1,00	
lalder	-0,08	-0,20	-0,14	-0,07	-0,05	0,06	0,37	0,03	0,03	0,04	-0,07	0,98	1,00

Tabell 5. Stegvis regresjon								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLER	lykkelig	lykkelig	lykkelig	lykkelig	lykkelig	lykkelig	lykkelig	lykkelig
inntektsforskjell	-0,10*** (0,01)	-0,09*** (0,01)	-0,09*** (0,01)	-0,09*** (0,01)	-0,09*** (0,01)	-0,09*** (0,01)	-0,09*** (0,01)	-0,08*** (0,01)
rettferdighet		0,05*** (0,01)	0,04*** (0,01)	0,04*** (0,01)	0,03*** (0,01)	0,03*** (0,01)	0,03*** (0,01)	0,03*** (0,01)
sosial			0,58*** (0,02)	0,50*** (0,02)	0,38*** (0,02)	0,39*** (0,02)	0,35*** (0,02)	0,33*** (0,02)
nære_bånd				0,57*** (0,02)	0,48*** (0,02)	0,48*** (0,02)	0,44*** (0,02)	0,42*** (0,02)
sosialt_aktiv					0,28*** (0,01)	0,28*** (0,01)	0,27*** (0,01)	0,27*** (0,01)
kriminalitet						-0,15*** (0,02)	-0,07*** (0,02)	-0,08*** (0,02)
trygghet							-0,36*** (0,01)	-0,36*** (0,01)
lalder								-0,14*** (0,02)
Konstant	7,82*** (0,02)	7,68*** (0,04)	7,45*** (0,04)	7,03*** (0,04)	6,39*** (0,05)	6,41*** (0,05)	7,18*** (0,05)	7,73*** (0,09)
Observasjoner	40 992	40 992	40 992	40 992	40 992	40 992	40 992	40 992
R ²	0,004	0,005	0,031	0,048	0,066	0,067	0,091	0,092
F-Observator	184,6	103,7	437,2	511,5	583,8	493,4	587,7	520,7
Standardfeil i parentes								
*** p<0,01, ** p<0,05, * p<0,1								

Figur 1. Diagnoseplott for regresjonsmodell (8) fra tabell 5

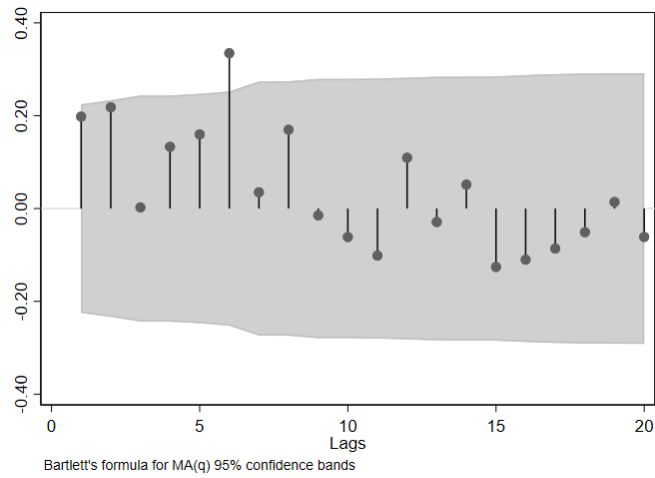


Tabell 6. Kjønnets påvirkning på rettferdighet og lykke

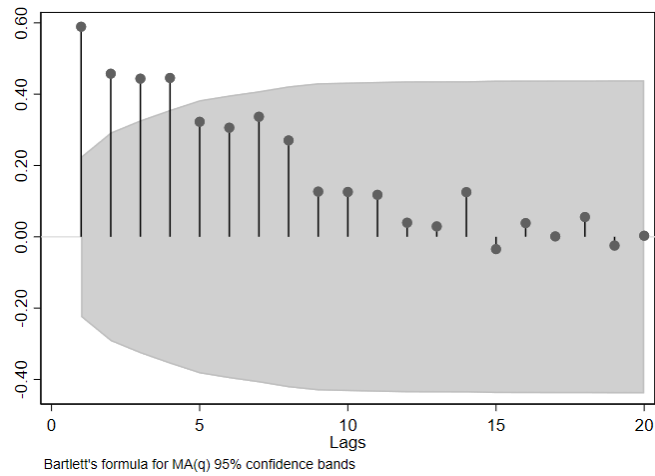
	(1)	(2)	(3)
VARIABLER	inntektsforskjell	rettferdighet	lykkelig
inntektsforskjell			-0,09*** (0,01)
rettferdighet			0,05*** (0,01)
kvinne	0,14*** (0,01)	-0,05*** (0,01)	0,03* (0,02)
Konstant	2,90*** (0,01)	2,42*** (0,01)	7,67*** (0,04)
Observasjoner	40 992	40 992	40 992
R²	0,004	0,001	0,005
F-observator	168,1	28,88	70,10
Standardfeil i parentes			
*** p<0,01, ** p<0,05, * p<0,1			

Figur 2. Autokorrelasjon av lykke gjennom livet

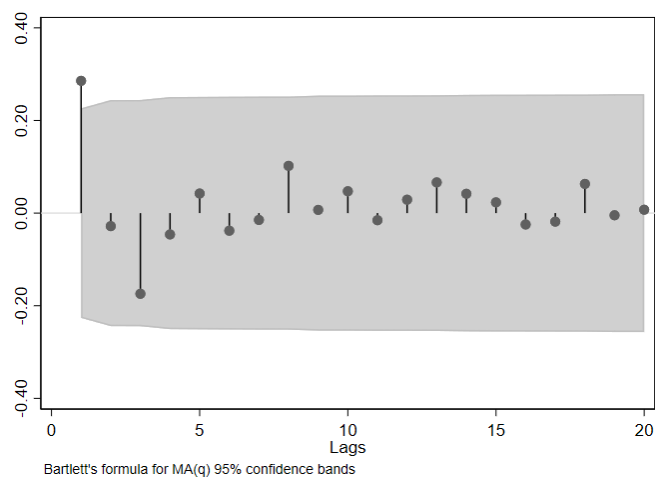
Del (A). Norge (under)



Del (B). Italia



Del (C). Russland



Tabell 7. Fitting ARIMA models to the process of happiness in France

```

> summary(Arima(France, order = c(0, 1, 2)))
Series: France
ARIMA(0, 1, 2)

Coefficients:
      ma1      ma2
    -0.6452 -0.0992
s.e.    0.1290  0.1232

sigma^2 estimated as 0.1026: log likelihood=-20.41
AIC=46.81 AICc=47.15 BIC=53.77

Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE
ACF1
Training set -0.05614993 0.3139894 0.2458546 -0.8991682 3.422542 0.8082646 -0.
04991224
> summary(Arima(France, order = c(2, 1, 0)))
Series: France
ARIMA(2, 1, 0)

Coefficients:
      ar1      ar2
    -0.4928 -0.3546
s.e.    0.1080  0.1080

sigma^2 estimated as 0.1134: log likelihood=-23.97
AIC=53.94 AICc=54.28 BIC=60.9

Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE
ACF1
Training set -0.0282777 0.3299994 0.2511585 -0.5203383 3.489935 0.8257016 -0.1
161642
> summary(Arima(France, order = c(2, 1, 2)))
Series: France
ARIMA(2, 1, 2)

Coefficients:
      ar1      ar2      ma1      ma2
    0.7151 -0.2450 -1.4009  0.5650
s.e.    0.4691  0.1475  0.4700  0.3616

sigma^2 estimated as 0.1021: log likelihood=-19.27
AIC=48.55 AICc=49.42 BIC=60.13

Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE
ACF1
Training set -0.04774024 0.3088865 0.2455316 -0.7791432 3.410757 0.8072026 -0.
01937606

```


Tabell 8. Modelltilnærming til lykke		
	(1)	(2)
VARIABLER	lykkelig	lykkelig
sosial	0,31***	0,21***
	(0,02)	(0,02)
nære_bånd	0,35***	0,18***
	(0,02)	(0,02)
sosialt_aktiv	0,19***	0,21***
	(0,01)	(0,01)
kriminalitet	-0,03	-0,11***
	(0,02)	(0,02)
trygghet	-0,27***	-0,20***
	(0,01)	(0,01)
inntektsforskjell	-0,07***	-0,06***
	(0,01)	(0,01)
rettferdighet	0,02**	0,03***
	(0,01)	(0,01)
helse	-0,56***	-0,53***
	(0,01)	(0,01)
kvinne	0,16***	0,17***
	(0,02)	(0,02)
sivilstatus	-0,07	-0,10*
	(0,05)	(0,05)
alder	0,26***	0,18***
	(0,02)	(0,02)
Konstant	7,41***	7,59***
	(0,09)	(0,09)
Observasjoner	40 992	40 992
R ²	0,159	0,197
Faste landeffekter	nei	ja
F-observator	705,7	547,5
Standardfeil i parentes		
*** p<0,01, ** p<0,05, * p<0,1		