

Eksamen STV1020 våren 2022 – *sensorveiledning i blått*

Del 1 (30 av 100 poeng)

Flervalgsoppgaver, rettes automatisk

Del 2 (70 av 100 poeng)

Du skal svare på alle oppgavene. Tallene i parentes viser antallet poeng du kan få for oppgaven. Hjelpemidler er tillatt, men det er ikke lov å samarbeide med andre om svarene, dette er en individuell eksamen som vil bli sjekket for plagiat både mot vitenskapelige tekster og hverandre.

Flere studier ser på effektene av å få barn på ulike livsutfall, som lykke, karriere og inntekt. Du ønsker å studere hvordan politisk deltagelse (i valg, i partier og andre organisasjoner) påvirkes av å få barn, hos kvinner og menn. Du undersøker et datasett med informasjon om et utvalg av den norske befolkningen, og tester en hypotese om at å få barn reduserer politisk deltagelse.

1. Hva er populasjonen, analyseenheten, den avhengige og den uavhengige variabelen for denne hypotesen? (4 poeng)

Populasjonen: Den norske befolkningen (1 poeng)

Analyseenheten: Individer i den norske befolkningen (et utvalg) (1 poeng)

Avhengig variabel: Politisk deltagelse (1 poeng)

Uavhengig variabel: Om individer har barn eller ikke (1 poeng)

Du definerer barn enkelt, som hvorvidt individene i undersøkelsen har barn, eller ikke (*barn*) og deltagelse i politikk er definert som svaret på følgende spørsmål: «Hvor ofte vil du si at du deltar i politikk (som kommunestyremøter, valg, politiske debatter o.l.)?» Svar: 1=veldig sjeldent, 2=ganske sjeldent, 3=verken sjeldent eller ofte, 4=ganske ofte, 5=veldig ofte.

2. Hva er målenivået til disse variablene? Begrunn svaret ditt ved å vise til egenskapene til oppgitt målenivå (4 poeng)

Barn er kategorisk målenivå . Det har bare to gjensidige utelukkende verdier (barn eller ikke barn) og kan ikke si noe om styrke (2 poeng). Politisk deltagelse er ordinal: Verdiene kan rangeres, men man kan ikke si noe om avstand eller styrke. (2 poeng).

3. Hva kan være problematisk med måten man har operasjonalisert «politisk deltagelse» på? (2 poeng)

Her spør man om et subjektivt mål. Det er ikke gitt at det fanger opp faktisk politisk deltagelse. Det kan sies å ha lav operasjonell validitet/målevaliditet. De får uansett poeng hvis de nevner dette, også poeng ved andre treffende kritikker av målet.

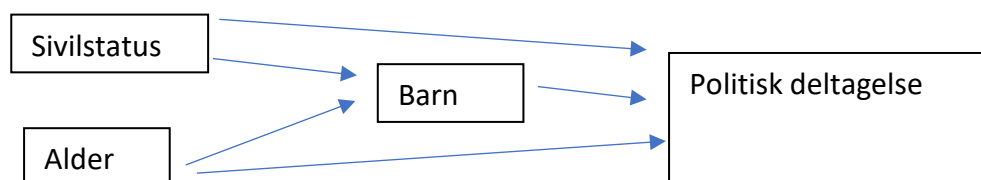
4. Hvilken bivariat analyse (dvs. korrelasjonsmål og hypotesetester) ville du brukt for å teste om det er sammenheng mellom de to variablene? Begrunn svaret kort. (2 poeng)

Tabellanalyse med kji kvadrattest fordi begge variablene vil behandles som kategoriske i en bivariat analyse.

5. Kan du foreslå en annen operasjonell definisjon av politisk deltagelse? (2 poeng)

Den operasjonelle definisjonen må forklare hvordan man konkret skal måle begrepet. Deltagelse må defineres på en måte som gjør det mulig å måle dette empirisk. Eksempler: Antall timer man rapporterer å ha deltatt i uka, stemmegivning i valg, medlemskap i organisasjoner, medlemskap i partier osv. (2 poeng).

6. En medstudent foreslår at alder kan påvirke både politisk deltagelse og ønsket om å få barn, men også at det samme gjelder sivilstatus (at de som er gift oftere vil ha barn, og delta mer i politikken). Tegn en kausalmodell av sammenhengen mellom barn og politisk deltagelse (som du ønsket å teste over) men hvor din medstudents innvendig også er med i modellen. En kausalmodell med barn, politisk deltagelse, alder og sivilstatus (4 poeng).



NY SENSORVEILEDNING: Her skal studentene få full pott, siden fasit stod i eksamen. Dvs. GI alle 4 poeng.

Opprinnelig sensorveiledning: Modell med piler mellom bakenforliggende variable, barn og politisk deltagelse (1 poeng), og med direkte piler mellom bakenforliggende og deltagelse (3 poeng) (4 totalt).

7. Du velger å endre operasjonaliseringen av «barn» til å måle antall barn (0, 1, 2, 3, osv.). Du estimerer så en bivariat regresjonsmodell. Politisk deltagelse er definert på en skala fra 0-100 hvor 100 er veldig høy deltagelse og 0 er ingen politisk deltagelse (her er vi ikke opptatt av operasjonaliseringen). Den bivariate regresjonsanalysen gir deg følgende resultat:

Tabell 1 - Bivariat sammenheng mellom barn og politisk deltagelse

Variabelnavn	Modell 1
Barn (β_1)	-2,5 (0,043)
Konstantledd (α)	20,1(0,002)
R^2	0,05
N	1000

P-verdier i parentes. Analyseenheten er 1000 tilfeldig utvalgte norske borgere over 18 år.

Hva forteller modellen deg? Tolk konstantledd og stigningskoeffisienten til Barn (4 poeng).

Konstantleddet forteller oss at gjennomsnittlig deltagelse for de uten barn er 20,1 (2 poeng). For hvert ekstra barn synker deltagelsen med 2,5. (2 poeng).

8. Er sammenhengen mellom barn og politisk deltagelse statistisk signifikant på 5% nivå?

Utrekning ikke strengt nødvendig (selv om det vil hjelpe), men begrunn svaret kort. (2 poeng)

Ja. P-verdien er under .05.

9. Hvis det er slik medstudenten din foreslår (altså at alder og sivilstatus både påvirker beslutningen om å få barn og politisk deltagelse) hva kan være ulempene ved å undersøke sammenhengen mellom barn og politisk deltagelse gjennom en bivariat OLS regresjon? (3 poeng)

Ulempen er at man ikke får kontrollert for de to bakenforliggende variablene, sivilstatus og alder. Dette fordi analysen er bivariat. Ønsker man å kontrollere for disse må man gjøre en multivariat regresjon.

10. Du velger å inkludere alder i modellen fra oppgave 6, som en kontrollvariabel, i en multipel regresjon. En økning i alder tilsvarer ett år, men siden 18 er laveste alder i undersøkelsen koder du om alder slik at nullpunktet er 18 år (verdien 0 tilsvarer 18 år, verdien 1 tilsvarer 19 år osv.). Du inkluderer også sivilstatus, hvor 1=gift, 0=ugift. Du får følgende resultater:

Tabell 2 - Multivariat sammenheng mellom barn og politisk deltagelse

Variabelnavn	Modell 2
Barn (β_1)	-2,2 (0,042)
Alder (β_2)	0,5 (0,001)
Sivilstatus (gift=1, ugift=0) (β_3)	3(0,13)
Konstantledd (α)	10,1(0,0001)
R^2	0,10
N	1000

P-verdier i parentes. Analyseenheten er 1000 tilfeldig utvalgte norske borgere over 18 år.

Tolk følgende: Konstantleddet, stigningskoeffisienten til alder, stigningskoeffisienten til sivilstatus og til slutt stigningskoeffisienten til barn. Gi en substansiell tolkning av hver av dem, og kommenter hvilke sammenhenger som er sterke (signifikante på 5%, med T-verdier over 2) (8 poeng).

Tolkning (2 poeng for hver ok tolkning, den trenger ikke å være så omfattende som dette), 1 poeng for «så som så» tolkninger). 2 poeng hvis signifikans kommenteres.

Konstantleddet uttrykker gjennomsnittlig deltagelse for ugifte personer med null barn som er 18 år. Det har 10,1 deltagelse. Sammenhengen er signifikant på 5% nivå (p-verdien er lavere enn 0.05).

Stigningskoeffisienten til barn er negativ, det vil si at deltagelse i gjennomsnitt synker med 2.2 for hvert ekstra barn personen har. Sammenhengen er signifikant på 5% nivå (p-verdien er lavere enn 0.05).

Stigningskoeffisienten til alder er positiv. Det vil si at hvert års økning i alder gir en økning i politisk deltagelse på 0,5 i snitt. Sammenhengen er signifikant på 5% nivå.

Stigningskoeffisienten til sivilstatus indikerer at gifte i snitt har 3-poengs høyere politisk deltagelse, men sammenhengen er ikke signifikant ((p-verdien er over 0.05).

11. Hva uttrykker R2 i analysen? Tolk R2 i tabell 2. Hva indikerer forskjellen mellom R2 i tabell 1 og R2 i tabell 2? (5 poeng)

R2 indikerer forklart varians i politisk deltagelse. Modellen i tabell 2 har en R2 på 0,10 som vil si at den forklarer 10% av variasjonen i avhengig variabel. Forskjellen mellom de to modellene tilsier at modellen i tabell 2 forklarer mer enn modellen i tabell 1 (5 poeng).

12. Er hypotesen du innledet med, om en negativ sammenheng mellom alder og deltagelse, styrket eller svekket av denne analysen? (3 poeng)

Hypotesen er styrket ettersom det er signifikant og substansiell negativ sammenheng mellom barn og deltagelse (3 poeng).NB: Her er det noe forvirring pga. feil i eksamen. Det skulle stå «negativ sammenheng mellom BARN og deltagelse». Studenter får riktig uavhengig av om de inkluderer ALDER eller BARN som variabel. Noe forvirring er å forvente her.

13. En kollega innvender at sammenhengen mellom barn og deltagelse sannsynligvis er ulik for kvinner og menn. Du har en dummyvariabel i datasettet på kjønn (kvinne=1, mann=0)¹. Hva kan du gjøre for å teste hypotesen til kollegaen din? Forklar hvordan du vil gjøre dette i modellen, og hvordan det kan endre tolkningen av koeffisienten for «barn». (6 poeng)

Man kan teste en interaksjonseffekt. Dvs. interagere variabelen «barn» med variabelen «kjønn». Mange vil nok svare at man bør kontrollere for kjønn, men dette er ikke riktig. (3 poeng). Man introduserer da et interaksjonsledd barn × kjønn, og dette vil endre tolkningen slik (3 poeng):

Koeffisienten «barn» vil uttrykke gjennomsnittlig endring i deltagelse for menn for hver økning i antall barn.

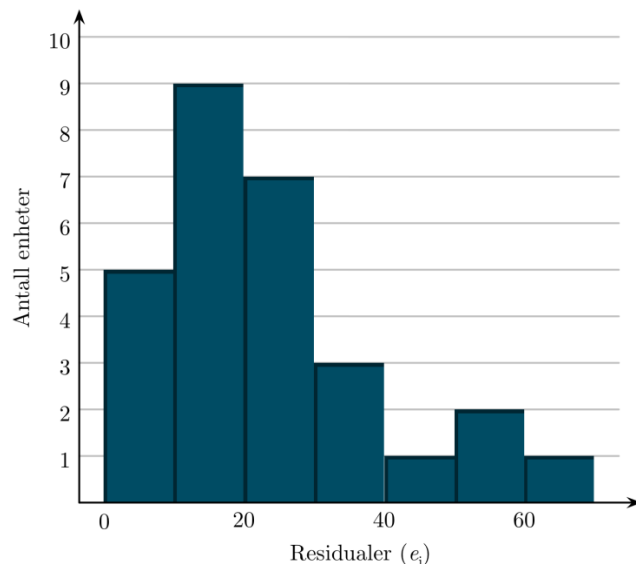
Koeffisienten «barn × kjønn» (interaksjonsleddet) vil uttrykke gjennomsnittlig endring i deltagelse for kvinner for hver økning i antall barn .

14. Du ønsker å se om geografisk tilhørighet påvirker politisk deltagelse. Du har en variabel med følgende fem kategorier i datasettet: «Osloregionen/hovedstaden», «stor by (over 50 000 innbyggere)», «liten by (under 50 000 og over 10 000 innbyggere)», «tettsted», «bygd». Hvordan vil du gå frem for å undersøke sammenhengen mellom denne variabelen og politisk deltagelse i en regresjonsmodell? 6 poeng

¹ Av tilfeldige grunner er det ingen transpersoner i utvalget, kategoriene «mann» og «kvinne» blir derfor uttømmende.

Det naturlige her er at man lager dummyvariable. 1 for hver kategori, men så må man velge en referansekategori som de andre skal sammenlignes med. F.eks kan «bygd» være referansekategori. Da blir alle koeffisientene sammenligning av politisk deltagelse i den gitte dummyvariabelen og «bygd» kategorien. Studenter som er inne på dette får full pott. 6 poeng. Studenter som er inne på at det må lages dummyvariable eller at hver variabel må inn som en kategorisk variabel, eller at variabelen må inn som en kategorisk variabel, uten å spesifisere videre får 3 poeng.

15.



Figur 1: Fordeling av residualer fra en regresjonsanalyse hvor OLS er benyttet.

Figur 1 viser fordelingen av residualene fra en regresjonsanalyse estimert med minste kvadraters metode (OLS). Kan du stole på signifikanstester av regresjonskoeffisientene fra denne regresjonsanalysen? Begrunn svaret. (3p)

Denne oppgaven omhandler forutsetninger for OLS, og mer spesifikt forutsetningen om at residualene skal være tilnærmet normalfordelte, spesielt når utvalget er lite som her ($N=28$). Dersom residualene er skjevfordelte kan standardfeilen bli over- eller underestimert (1p). I så fall vil konklusjonene fra eventuelle hypotesetester også kunne bli feil (1p). Siden residualene i figur 1 er skjevfordelte, er ikke denne forutsetningen for OLS oppfylt, og vi kan derfor ikke stole på signifikanstester av regresjonskoeffisientene fra denne regresjonsanalysen (1p).

15. Her har vi to regresjonsmodeller. En «sann» modell, som tilsvarer sammenhengen slik den egentlig er og en «feil» modell («vår modell») som vi har estimert:

Sann modell: $Deltagelse = \alpha + \beta_1 Barn_1 + \beta_2 Alder_2 + \beta_3 Sivilstatus_3 + \mu$

Vår modell: $Deltagelse = \alpha + \beta_1 Barn_1 + \beta_2 Alder_2 + \mu$

I vår modell har vi glemt å inkludere sivilstatus. Hva vil skje med residualene i vår modell? (5 poeng)

Effekten av sivilstatus vil befinne seg i restleddet, residualene vil få systematisk skjevhet. Vi vil få skjeve estimater av Y. Utelatt variabelskjevhet er ett begrep for dette. Dette bryter forutsetningen om ingen skjevhet i residualene.

16. Du har fremsatt og analysert en hypotese om at barn påvirker politisk deltagelse. Hva kan være en troverdig mekanisme for å forklare denne sammenhengen? (2 poeng).

En mekanisme knytter sammen de to variablene i en årsaksforklaring. I tilfellet barn og deltagelse er det mange mulige. En mer eller mindre rimelig en som svarer til definisjonen «mekanisme» gir full pott.

17. Ett av Kellstedt og Whittens kriterier for at vi har å gjøre med en kausal sammenheng er at det ikke forekommer motsatt kausalitet («reverse causality»). Hva betyr dette? Og er det troverdig i tilfellet barn og politisk deltagelse? (3 poeng)

Dette innebærer at politisk deltagelse påvirker antall barn (altså at det går andre veien). Det kan ikke utelukkes. Men så lenge studenten har en forståelse av dette og gjør seg noen tanker den ene eller andre veien om det er rimelig, så gir det uttelling.