

Sensorveiledning for eksamen i SOS1002 høsten 2019

Eksamen består av tre oppgaver. Hver besvarelse teller en tredjedel av den samlede karakteren. For å stå på eksamen må en oppnå minst karakter E på hver av de tre oppgavene.

Oppgave 1

- a) Beskriv de viktigste forskjellene mellom deduktive og induktive forskningsdesign. Gi gjerne noen eksempler.

Besvarelsen bør omhandle både kvantitative og kvalitative forskningsdesign. I besvarelsen bør studentene drøfte ulike måter å kombinere teoretiske modeller med empiriske data. I begrepsavklaringene til Ringdal (2018) defineres deduksjon som logiske slutninger fra en generell regel eller lovmessighet til enkelttilfeller, mens induksjon defineres som logiske slutninger fra enkelttilfeller til en generell regel eller lovmessighet. Her er det viktig vurdere om studentene har forstått at de deduktive design går ut på å teste ut teorier og modeller med empiriske data (ovenfra og ned), mens de induktive designene har som mål å utvikle generell teoretisk kunnskap gjennom systematiske analyser av empiri (nedenfra og opp). Rindal bruker på side 37 og 40 den logiske positivismen (empirismen) som eksempel på induktive design, og Poppers kritiske rasjonalisme som eksempel på deduktive design. På side 45 skriver han videre hvordan om hvordan induktiv og deduktiv tankegang kombineres i den hypotetisk-deduktive metoden. Studentene bør i tillegg også kunne trekke inn kvalitative forskningsdesign i denne drøftingen, og da særlig hvordan SDI-metoden som er beskrevet i Tjora bygger både på induksjon og deduksjon.

- b) Drøft kriterier som brukes for å vurdere forskningskvalitet. Nevn de mest sentrale kriteriene for kvalitetsvurderinger i kvalitative og kvantitative metoder.

I kvantitative metoder er det særlig begrepene validitet, reliabilitet og generaliserbarhet som er de viktigste knaggene i denne drøftingen. Ringdal beskriver validitet som et spørsmål om vi faktisk måler det vi vil måle, og reliabilitet som om gjentatte målinger med samme måleinstrument vil gi samme resultat. I kvalitativ forskning erstattes ofte begrepet reliabilitet med troverdighet, mens validitet ofte skiftes ut med bekreftbarhet. I kvantitative metoder forbindes begrepet generaliserbarhet ofte med statistisk generaliserbarhet fra et sannsynlighetsutvalg til en konkret eller teoretisk populasjon, mens Tjora lanserer flere alternative begrep for generalisering i sin bok.

- c) Av og til forekommer det anklager om uredelighet og fusk bak publiserte forskningsresultater. Drøft mulige eksempler på uredelig forskning.

Ringdal skriver om forskningsetikk i kapittel 3, Tjora i kapittel 8, og Fangen i kapittel 11 og 13. Her forventer vi å få gjengitt en del av disse drøftingene.

Oppgave 2

- a) Vedlegg 1 gjengir en loggfil fra statistikkprogrammet Stata, og viser en analyse av data fra det norske utvalget av European Social Survey fra 2018. Beskriv oppbyggingen av modell 1 og modell 2 i vedlegg 1, og forklar hva disse to modellene viser (se vedlegg 1 s. 5-8).

Målet med denne oppgaven er å teste om studentene klarer å se hvordan to lineære regresjonsmodeller er bygd opp i statistikkprogrammet Stata. Her selekterer vi først ut det norske utvalget av ESS9. Deretter lager vi en refleksiv indeks som måler politisk tillit ut fra de tre variablene `trstplc`, `trstplt` og `prstprt`. Reliabilitetstesten viser at disse tre variablene har en alfa på 0,805, og ut fra Ringdal bør de vite at alfaverdier over 0,7 viser at indeksen har intern konsistens-reliabilitet, og at vi her har et godt grunnlag for å kunne slå sammen de tre variablene til en refleksiv indeks for politisk tillit.

Videre viser utskriftet hvordan vi behandler de uavhengige variablene. Her blir variabelen `gndr` kodet om til dummyen `kvinne`, der kvinner får verdien 1 mens menn får verdien 0. Videre fjerner vi verdien 999 fra aldersvariabelen ved å kode den om til missingverdien «.». Grunnen til denne omkodingen blir ikke forklart i oppsettet, men de som studerer kodene på side 7 vil kunne se at alder er målt i antall år, og at 999 da blir en ulogisk kode som må kodes om til missing. Til slutt viser vi hvordan politisk interesse er målt med en ordinalnivåvariabel med 4 trinn. Her blir ikke de underliggende tallkodene presentert, og de fleste må sannsynligvis bruke litt tid på å skjønne hvordan variabelen er kodet. De som har mye erfaring med Stata bør vite at `tab`-kommandoen presenterer verdiene fra laveste verdi (øverst) til høyeste verdi (nederst) i tabellen. I tillegg er det mulig å se på dummyleddene for politisk interesse i modell 2, og da vil de mest erfarne kunne se at politisk tillit går ned for hver økende verdi på politisk interesse. Ut fra denne informasjonen blir det mest naturlig å konkludere med at «Very interested» er kodet 1, «Quite interested» er kodet 2, «Hardly interested» er kodet 3 og «Not at all interested» er kodet 4. Men her må vi godta også alle andre koder så lenge de er logisk konsistente med variabelens koeffisienter i modell 1 og 2. I tillegg bruker vi også `eduys` som et mål på antall år informantene har gått på skole, men denne variabelen brukes slik den opprinnelig er kodet i `ESS9e01.dta`. De som er usikre på hva variabelnavnene `women`, `age` og `eduys` kan lese variablenes fullstendige navn under `summarize`-kommandoen på side 7. Her har vi i tillegg lagt inn informasjon om variabelen `nwspol`, men denne blir ikke brukt videre utover i analysen. De som ikke har erfaring med å lese utskrift fra `summarize`-kommandoen med subkommandoen `d` (detail) bør likevel ut fra sin generelle kunnskap om statistiske begreper klare å tolke informasjonen i utskiftet.

På side 8 presenterer vi de to lineære regresjonsmodellene. I modell 1 er effektene av de fire uavhengige variablene estimert ut fra den presenterte kodingen. I modell 1 er det kun variabelen politisk interesse som har statistisk signifikant effekt på 5%-nivå på politisk tillit. Her vil sannsynligvis en del stusse over at det er negativ sammenheng mellom politisk interesse og politisk tillit, men dette har sammenheng med den «ulogiske» kodingen av politisk interesse som gjør at økende verdi betyr minkende politisk interesse. Det er derfor en sterk sammenheng mellom økende politisk interesse og økende politisk tillit. De studentene som har fulgt forelesningene har vært bort i tilsvarende måleproblemer på forelesningene. R^2 viser at modellen forklarer litt over 2 prosent av variansen i politisk tillit.

I modell 2 har utvidet modellen ved å estimere en kurvelineær alderseffekt og ved å dummykode variabelen som måler politisk interesse. I denne modellen får vi en negativ signifikant koeffisient for alder og en positiv signifikant koeffisient for alder i andre. De beste studentene bør da kunne se at modellen viser at alder danner en u-formet kurve i forhold til politisk tillit, og de som har jobbet grundig med dette vil kunne regne ut at bunnpunktet på denne kurven ligger rundt 53 år ($-0,1285672 / (2 * 0,0012163) = 52,85176$). I modell 2 blir også utdanningsvariabelen signifikant på 5%-nivået, og viser at tilliten øker med 0,086 tinn for hvert år med utdanning. Den dummykodede variabelen som måler politisk interesse viser også at de som ikke er så veldig politisk interesserte (hardly interested) i gjennomsnitt ligger omtrent ett trinn lavere på den avhengige variabelen enn kontrollgruppa som er veldig politisk interessert (very interested), og at denne forskjellen er statistisk signifikant på 5%-nivået. Videre ser vi at de som ikke er politisk interesserte (not at all interested) i

gjennomsnitt ligger over 3 trinn lavere på den avhengige variabelen som måler den politisk tillit, sammenlignet med de som er svært interesserte. Denne forskjellen er statistisk signifikant både på 5%-nivået og 1%-nivået. Modell 2 forklarer rundt 3,5 prosent av variansen i indeksen for politisk tillit. Etter som modell to estimerer tre flere parametre enn modell 1, så vil sikkert noen velge å se på justert R^2 når de skal sammenligne de to modellenes forklaringskraft, og da vil de se at også justert R^2 øker fra 2 prosent til 3 prosent forklart. Og det indikerer at modell 2 er vesentlig bedre enn modell 1.

- b) Beskriv oppbyggingen av modell 3 og modell 4, og forklar hva disse to modellene viser (se vedlegg 1 s. 9-10).

På side 9 deler vi indeksen for politisk tillit inn i to grupper ut fra medianverdien på rundt 18, og koder de som har lav politisk tillit til verdien 0 og de som har høy politisk tillit til verdien 1. Dermed kan vi bruke denne som avhengig variabel når vi estimerer to logistiske regresjonsmodeller med de samme uavhengige variablene som i modell 1 og modell 2.

Ut fra Ringdal kan en logistisk regresjonsmodell etter tre skalaer:

- *Tolkning av de logistiske regresjonskoeffisientene*
 - Ser på koeffisientens fortegn, og ser om det er statistisk signifikans
- *Tolkning av koeffisientene i oddsskalaen*
 - Her får vi bedre forståelse av styrken på sammenhengen.
- *Omregning til sannsynligheter*
 - Resultatene blir enkle å formidle, men det krever en god del beregningsarbeid for å få fram sannsynlighetene.
 - Dette ser vi på i neste forelesning

Ut fra den første metoden tolker vi modell 3 til at variabelen *eduysr* har en signifikant positiv effekt på å tilhøre gruppen med høy politisk tillit, og det viser at hvert år med utdanning øker sannsynligheten for å havne i gruppen med høy politisk tillit. Videre ser vi for hvert trinn oppover i politisk interesse, så minskes sannsynligheten for å havne i gruppen med høy politisk tillit. Etter som kodingen av politisk interesse her gjør at interessen blir lavere for hvert trinn, så betyr dette at jo lavere politisk interesse du har jo lavere blir også sannsynligheten for å tilhøre gruppe med høy politisk interesse. I modell 4 ser vi at vi også her får en signifikant kurvelineær alderseffekt, og at de to gruppene med lavest politisk interesse har signifikant lavere sannsynlighet for å tilhøre gruppe med høy politisk tillit når vi sammenligner dem med gruppe som er svært politisk interessert.

De som ønsker å tolke modellen baser på oddsskalaen må se på de estimerte oddsratene på side 10. Her ser vi i modell 3 at oddsen for å ha høy politisk tillit øker med 2,8 prosent [$100(1,028-1)=2,8$] for hvert økte trinn på utdanningsvariabelen. Og for hvert økte trinn på variabelen for politisk interesse så senkes oddsen for å ha høy politisk tillit med 30 prosent [$100(0,6999-1)=30$]. I modell 4 blir det vanskelig å bruke oddsratene til å beskrive alderseffekten, så her ser vi helst at studentene hopper over denne variabelen. For utdanningsvariabelen ser vi at oddsen for å ha høy politisk tillit øker med 3,7 prosent [$100(1,037-1)=3,7$] for hvert økte trinn på utdanningsvariabelen. Og at de med liten politisk interesse har 48 prosent [$100(0,58-1)=48$] lavere odds for å ha høy politisk tillit enn de som er svært politisk interesserte, og at de som ikke er politisk interessert har 29 prosent [$100(0,313-1)=29$] lavere odds for å ha høy politisk tillit.

De som ønsker å tolke en av modellene med sannsynligheter gjør det enklest for seg selv om de velger den enkleste modellen som er modell 3. Her må de selv velge hvilke verdier de vil sette opp i sine prediksjoner, og dette vil variere fra person til person. Hvis jeg f.eks. ønsker å sammenligne sannsynligheten for å ha høy politisk tillit for en mann på 50 år med 10 års utdanning og stor politisk

interesse med en kvinne med de samme kjennetegnene så setter jeg opp følgende beregninger ut fra tabell 3 side 9:

$$L_{\text{mann}} = 0,7053 + 0,0921*0 - 0,0037*50 + 0,0281*10 - 0,3568*0 = 0,8013$$

Denne logiten kan så regnes om til en sannsynlighet med formelen $1/(1-\text{eksp}(-L)) = 1/(1 + \text{exp}(-0,8013)) = 0,69$

Det vil si at en mann med disse kjennetegnene har 69 prosent sannsynlighet for å ha høy politisk tillit.

$$L_{\text{kvinne}} = 0,7053 + 0,0921*1 - 0,0037*50 + 0,0281*10 - 0,3568*0 = 0,8934$$

Denne logiten kan så regnes om til en sannsynlighet med formelen $1/(1-\text{eksp}(-L)) = 1/(1 + \text{exp}(-0,8934)) = 0,71$

Det vil si at en mann med disse kjennetegnene har 71 prosent sannsynlighet for å ha høy politisk tillit.

- c) Fremmøteprosenten ved Stortingsvalget 2017 var ifølge nrk.no 78,2 prosent. I det norske utvalget av ESS9 fra 2018 var det 90,3 prosent av de stemmeberettigede som hevder at de stemte ved siste Stortingsvalg. Drøft mulige årsaker til at hjemmesitterne blir underrepresentert i utvalget fra ESS.

Her ønsker vi å undersøke hva studentene har fått med seg av seleksjonsmekanismer knyttet til surveyundersøkelser og eventuelle intervju effekter som kan oppstå ved personlig intervju. De flinkeste studentene bør ha fått med seg at den norske delen av ESS er basert på personlige intervju, og at det av problemene som kan oppstå når vi bruker intervjuere er at informantene kan svare ut fra det de oppfatter som riktige forventninger og at sosialt aksepterte svar ofte blir overrapportert ved personlige intervju. Men sannsynligvis har seleksjonseffektene enda større betydning for den høye valgdeltakelsen, og den mest opplagte forklaringen er at de som er villige til å bli med på en slik spørreundersøkelse også har høyere sannsynlighet for å delta i valg. I tillegg til disse forklaringene vil kanskje noen også kunne se at ESS er et sannsynlighetsutvalg for personer som 15 år eller eldre, mens stemmerettsalderen i Norge er 18 år. De rundt 90 prosent som hevder de har stemt er derfor 90 prosent av de som er 18 år eller eldre i ESS-utvalget.

Opgave 3

- a) Ved de siste fire Stortingsvalgene er det altså litt under 80 prosent av de stemmeberettigede som har stemt. Du ønsker å finne ut hvorfor mer enn 20 prosent av de stemmeberettigede ikke stemmer. Hvordan ville du ha utformet et kvalitativt forskningsprosjekt som kan gi oss mer informasjon om hvorfor så mange ikke stemmer ved Stortingsvalgene? Formuler en problemstilling for undersøkelsen, begrunn valg av metode(r), og beskriv videre trinn i metodeopplegget frem til data er klargjort for analyse.

Her forventer vi at studentene klarer å rendyrke kvalitative design, og at de ikke begynner å blande inn bruken av strukturerte spørreskjema og lignende i denne drøftingen. Og her bør dere legger mer vekt på logiske begrunnelser bak det valgte designet enn på hvilket design som velges.

- b) Drøft hvilke etiske utfordringer du kan bli stilt overfor hvis du skal gjennomføre dette forskningsprosjektet.

Her bør studentene klare å bruke pensumstoffet både fra Ringdal, Tjora og Fangen i drøftingen.