

Sensurveiledning SOS1002 Høst 2018

Merknad til veiledningen. Det ble avdekket et par småfeil i oppgavesettet. Faglærer Hilde B gikk trøsterunden og informerte om feil i Tabell 1 (Beta koeffisienten for Alder i blokk 2 skulle være -0.017). Det var imidlertid ikke mulig å rekke over alle. Noen vil ha oppdaget feilen selv (ved å bruke appendix), andre ikke. Det må tas hensyn til dette i sensuren. Hilde ble også gjort oppmerksom på at det ikke var samsvar mellom Beta koeffisientene for frukt/grønnsaker i tabell 1 og i Stata Output (appendix). I oppgave 2 d og e, samt i og j, kan det derfor være ulike svar på oppgaven etter hvilke tall som er brukt i prediksjonen (fra tabell eller appendix).

Ang aldersvariabel: Her kom det spm. om hvilken verdi man skal bruke for 25 år og 74 år. I tidligere års veiledninger har det vært brukt reelle tall. Her kan årets foreleser / gruppeleder gitt andre instruksjoner (25=0, 74=49). Det er derfor OK om de bruker 0 i stedet for 25 etc. i prediksjonene.

Oppgave 2 a og b opplevdes svært like. I utskriften (bokmål) kom b på ny side. Mange hadde derfor svart svært utdypende på a før de så b, og var usikker på hva mer som var forventet.

Ved kobling av kvantitativ og kvalitativ oppgavedel har det blitt med en feil. Opprinnelig var det laget oppgave 3 og 4. I endelig versjon ble disse til 4 og 5. I oppgave 5 henvises det tilbake til oppgave 4a. Dette skulle vært 5a. forhåpentligvis var dette intuitivt for de aller fleste.

Bokmål

Eksamen er delt i to deler, en kvantitativ og en kvalitativ del. Begge deler må besvares og de teller 50% hver. Begge deler må bestås for å få eksamen godkjent (Tips: Begynn med det du kan og disponer tiden godt. Ikke «heng deg opp» i detaljer).

Lykke til!

Kvantitativ del

Alle tre oppgaver skal besvares. Oppgave 1 teller 25%, oppgave 2 teller 50%, og oppgave 3 teller 25% av den samlede (kvantitative) karakteren.

Oppgave 1

Gjør kort rede for 3 av de 4 begrepene. Bruk inntil ½ side på hvert begrep.

- a) Operasjonalisering

Her bør studentene kunne vise til overgangen fra teoretiske til empiriske definisjoner av begreper. I Ringdal heter det «å knytte empiriske indikatorer til teoretiske begreper». Dette skal det være nok å nevne, men om studentene kan vise til utfordringer ved denne overgangen (for eksempel når det gjelder validitet) vil det telle som et pluss.

b) Kausalitet

Her bør studentene vite at vi er ute etter å kunne forstå årsakssammenhenger og at dette er problematisk i samfunnsvitenskapen. En mulig inngang er å vise til Hempels forklaringsmodell (som det refereres til i Ringdal) som har tre hovedelementer: at det er snakk om en generell lovmessighet (med allmenn gyldighet i tid og rom), randbetingelser og en utløsende årsak. Studentene bør også ta opp det faktum at årsaksmodeller er omdiskuterte i samfunnsvitenskapene (i motsetning til i naturvitenskapene). I samfunnsvitenskapen benytter vi ofte betegnelser som assosiasjoner og korrelasjoner i stedet for kausale forhold. I stedet for deterministiske sammenhenger: «hvis X, så alltid Y», så opererer samfunnsvitenskapen gjerne med sannsynlighetsformuleringer: «hvis X, så ofte Y».

c) Samspill/interaksjon

Her må studenten frem til at effekten av én X-variabel er betinget av en annen. Eksempler er ikke nødvendige, men et pluss: for eksempel at effekten på utdanning på timelønn er sterkere for menn enn for kvinner. Det er også et pluss om studenten nevner hvordan et samspillsledd kan testes (ved å gange sammen to variabler og se signifikansen). Det er også et pluss om studenten nevner at det finnes alternative måter å studere samspill på: ved å splitte analysen i to (for eksempel: én analyse for menn, og én for kvinner).

d) Kurvelineær sammenheng

Her må studenten komme frem til at effekten av én X-variabel på den avhengige variabelen ikke følger en rett line, men at den er bøyelig/kurvelineær. Dette er tilstrekkelig, men det er pluss å nevne eksempler, samt nevne hvordan dette testes statistisk (ved bruk av andregradsledd). De sterkeste studentene vil også kunne nevne at det å forstå sammenhengen bare kan oppnås ved bruk av effektplott.

Oppgave 2

I analysene under skal vi se på to sett av regresjoner (én lineær og én logistisk), som hver har to blokker. Analysene tar utgangspunkt i hele den Europeiske befolkningen og er basert på European Social Survey (ESS) fra 2014. ESS er en multinasjonal spørreundersøkelse som analyserer samfunnsvitenskapelige spørsmål.

Regresjonsanalysene under skal kunne gi svar på om det er kjønns- og utdanningsforskjeller i helse Europa.

Selvopplevd helse er avhengig variabel i alle analysene.

I de *lineære regresjonsanalysene* går verdiene på den avhengige variabelen fra 0 (veldig dårlig helse), 1 (dårlig helse), 2 (gjennomsnittlig helse), 3 (god helse), til 4 (veldig god helse)

I de *logistiske regresjonsanalysene* har vi kodet om den samme variabelen til en dikotom/todelt variabel, med verdiene 1 (veldig dårlig helse, dårlig helse og gjennomsnittlig helse) og 0 (god helse og veldig god helse).

Forklaringsvariablene (x-variablene) er de samme både for de lineære og logistiske regresjonsanalysene.

Blokk 1 (for alle analyser):

Alder: kontinuerlig variabel fra 25 til 75 år.

Kjønn: *kvinne* (kodet 1), *mann* (kodet 0).

Utdannelsesnivå: Dummyvariabler: *høy*-, *middels*- og *lav utdanning*, hvor høy utdanning er referansekategori

I Blokk 2 legges følgende variabler til analysene.

Inntak av frukt og grønnsaker: Dummyvariabler: *Høyt inntak* (referanse), *moderat inntak*, og *lavt inntak*.

Alkoholkonsum: Dummyvariabler: *Høyt forbruk*, *moderat forbruk*, og *lavt forbruk* (referanse).

Fysisk aktivitet: Dummyvariabler: *Fysisk aktiv* (kodet 0) og *fysisk inaktiv* (kodet 1).

Røyking: Dummy-variabler: *Ikke røyker* (referanse), *tidligere røyker*, og *røyker*.

Ta utgangspunkt i Tabell 1 (samt eventuelt appendiks):

- a) Viser analysen at det er kjønns- og utdanningsforskjeller i Europa i rapportering av selv-opplevd helse? Forklar hvorfor/hvorfor ikke.

Her kan det svares kort: ja, det er forskjeller både mellom høy-middels og mellom høy-lav. Dette kan enkelt ses av signifikansen til disse koeffisientene, både i blokk 1 og blokk 2. Om studentene bare viser til blokk 1 eller bare blokk 2, så er det OK.

- b) Kan vi ut fra tabellen forklare disse forskjellene? Forklar i så fall hvordan og eventuelt hvor mye.

Ja, blokk 2 viser at koeffisientene (både for forskjellene mellom høy-middels og mellom høy-lav) reduseres fra blokk 1. Det er tilstrekkelig å nevne hvor mye den reduseres (i desimaler eller prosent).

- c) Hvordan kan vi tolke regresjons-konstanten?

Konstanten angir gjennomsnittsverdi for predikert Y når alle verdier av uavhengige variabler er satt til 0. Dette er tilstrekkelig å nevne, men sterke studenter vil også nevne at dette betegner en 0 år gammel mann med lav utdanning etc.

- d) *Prediker* nivået av selvopplevd helse for en lavt utdannet kvinne på 74 år, som har lavt inntak av frukt/grønt, som er fysisk inaktiv, som drikker lite alkohol og som røyker.

Minimumskravet er å sette opp riktig ligning: Konstant + alder*74 + Kjønn*1 + Lav utdanning*1 + Lavt inntak av frukt*1 + Fysisk inaktiv*1 + Lite alkohol*1 + Røyker daglig*1. Det er ikke viktig at svaret er rett pga feilen for «alderskoeffisienten»

- e) *Prediker* nivået av selvopplevd helse for en høyt utdannet mann på 25 år, som har høyt inntak av frukt/grønt, som er fysisk aktiv, som har høyt alkoholforbruk og som ikke røyker.

Minimumskravet er å sette opp riktig ligning: Konstant + alder*25 + Kjønn*0 + Høy utdannelse*1 + Høy inntak av frukt*1 + Fysisk aktiv*1 + Høyt alkoholinntak*1 + Ikke-røyker*1. Det er ikke viktig at svaret er rett pga feilen for «alderskoeffisienten»

- f) Hvor mye av variasjonen i selv-rapportert helse er vi i stand til å forklare ut fra analysene i Tabell 1?

R2 viser i appendiks viser 16%. De må også vise om de referer til R2 eller adjusted R2 (de kan ikke bare skrive 16%). Dette er opp til dem selv, men om de velger å rapportere adjusted R2 og forklarer hvorfor er dette et pluss.

Ta utgangspunkt i Tabell 2 (samt eventuelt appendiks).

- g) Gir også denne analysen grunnlag for å hevde at vi har kjønns- og utdanningsforskjeller i selv-rapportert helse i Europa? Forklar hvorfor/hvorfor ikke.

Ja. Argumentasjonen er akkurat den samme som over. Resultatene er de samme.

- h) Gir den logistiske regresjonsanalysen grunnlag til å trekke noen andre konklusjoner sammenlignet med den lineære regresjonsanalysen?

Det er tilstrekkelig å nevne at den substansielle konklusjonen er den samme. Om de svarer ja, på bakgrunn av at tolkningen av odds ratioer er ulik tolkningen av OLS-konstanten, så er det OK. Men da må det nevnes at konklusjonen er den samme.

- i) *Prediker* sannsynligheten for at en lavt utdannet kvinne på 74 år, som har lavt inntak av frukt/grønt, som er fysisk inaktiv, som drikker lite alkohol og som røyker, har dårlig selvopplevd helse. Du kan bruke følgende formel:

$$\hat{P} = \frac{1}{1 + e^{-\hat{L}}}$$

Samme prediksjon som i forrige oppgave. Her er minimumskravet at det er LOGITENE og ikke ODDS RATIOENE som er satt inn i formelen.

- j) Bruk samme formel for å *predikere* sannsynligheten for at en for en høyt utdannet mann på 25 år, som har høyt inntak av frukt/grønt, som er fysisk aktiv, som har høyt alkoholforbruk og som ikke røyker har dårlig selvopplevd helse.

Samme prediksjon som i forrige oppgave. Her er minimumskravet at det er LOGITENE og ikke ODDS RATIOENE som er satt inn i formelen.

Oppgave 3

- a) Diskuter hvilke av de to modellene (den lineære eller den logistiske) du synes egner seg best. Det er ikke sikkert svaret er opplagt, og du kan gjerne diskutere for og imot.

Her er vi særlig ute etter to ting:

1) at logistisk regresjon krever todelt avhengig variabel, og dette reduserer mengden informasjon vi har (sammenlignet med en kontinuert avhengig variabel). Dette heller i retning av å bruke lineær regresjon.

2) at den avhengige variabelen i den lineære regresjonen bare har 5 verdier. Dette er ikke ideelt og gir begrensninger for analysen, noe som heller i retning av logistisk regresjon.

Svar utover dette er et pluss, og gjennomtenkte svar på premieres. For eksempel kan man foreslå å kjøre begge analyser for å se om man får de samme resultatene.

- b) Tenk deg at du hadde tilgang til disse dataene, samt egnet analyseverktøy (som for eksempel STATA eller SPSS). Hvordan ville du gått frem for å forbedre modellene?

Her er vi først og fremst ute etter 1) å teste samspillsledd, 2) å teste for kurvlineære effekter (på alder), 3) legge til flere relevante variabler. Flere forslag er et pluss.

EKSTRA-OPPGAVE

Her kan du få plusspoeng. INGENTING fratrekkes hvis den ikke besvares.

Forskning skal danne grunnlag for politikk. Hvordan ville du kommunisert funnene over til politikerne våre, og hvilke råd ville du gitt?

Dette er en oppgave som er ment å fange opp faglig modenhet. En god diskusjon med støtte i analysen skal premieres.

Hovedtema i oppgaven er sosial ulikhet i helse og dets determinanter. Politiske forslag kan derfor gå i retning av å foreslå tiltak (spise sunnere, trene etc) for å redusere sosial ulikhet helse. Merk ellers at det ligger et kontraintuitivt funn i modellen: at de som drikker mest er sunnest (se derfor etter om studenten foreslår å drikke mindre, noe som ikke støttes av analysen).

Man kan også foreslå tiltak direkte for å bedre helsen (altså ikke bare tiltak for å redusere sosial ulikhet i helse). Som for eksempel å sørge for tiltak til at flere tar høyere utdanning/ hindre drop-out.

Det er også bra om studenten uttrykker seg forsiktig (siden samfunnsvitenskapen ser på assosiasjoner og ikke kausale forhold), i motsetning til deterministisk. Som forskere må vi alltid ta høyde for usikkerheten i data og metode.

Det er også bra om studenten gir forslag basert på et europeisk perspektiv, siden det brukes data basert på europeiske gjennomsnitt.

Kvalitativ del

Oppgave 4 (teller 20% av den samlede kvalitative karakteren)

Gjør kort rede for tre av de fem begrepene. Bruk maksimum en ½ side på hvert begrep

a) **Deltagende observasjon**

I Tjora defineres deltagende observasjon i begrepslisten – som «observasjons som er åpen og deltagende». Tjora åpner for en breiere definisjon i boka. I forelesning har vi gått gjennom både åpen og skjult deltagende observasjon. Eksempler kan styrke besvarelsen, evt kontrastering til skjult deltagende observasjon. Refleksjon over etiske problemstillinger vil styrke besvarelsen.

b) **Induktiv**

I Tjora 2017 gis følgende definisjon: Induktiv betegner forskning som er eksplorerende og/eller empiridrevet. Vi bruker også induktiv om det å slutte fra et enkelt tilfelle (eller et begrenset antall enkelttilfeller) til en generell regel. Eksempel, kontrastering til deduktiv og/eller utviding til abduktiv, kan styrke besvarelsen.

c) **Transparens**

Defineres i Tjora 2017 som gjennomsiktighet og brukes i metodologisk sammenheng om hvorvidt og hvordan detaljene i en studie beskrives for leseren. Stor grad av transparens betraktes som positivt for høy forskningsmessig kvalitet.

I forelesningene har det blitt lagt en del vekt på dette som element i kvalitets- og etikkvurderinger i forskning. Kobling til gyldighet og pålitelighet, og/eller generalisering, vil styrke besvarelsen.

d) **Pålitelighet (reliabilitet)**

Handler ifølge Tjora 2017, om hvorvidt vi oppfatter en klar sammenheng mellom empiri, analyse og resultater i en undersøkelse, og at dette ikke er styrt av personlige, politiske eller andre faktorer som det ikke er redegjort for. Her vil det være sentralt å trekke inn refleksjon over egen tolkning – refleksivitet – i besvarelsen. Eksempler kan være en styrke, samt kommentar mot Gyldighet (validitet).

e) **Forskerssubjektivitet**

Brukes i Tjora (2017) om den personlige påvirkningen som forskeren vil ha på forskningen.

I pensum listes grad av forskerssubjektivitet (og inter-subjektivitet) opp i forbindelse med design, datagenerering, analyse og tolkning i ulike kvalitative metoder, mens objektivitet knyttes til analyse i kvantitative metoder.

Refleksjon om «grad» av subjektivitet, og/eller eksempler vil styrke besvarelsen.

Oppgave 5 (teller 80% av den samlede kvalitative karakteren)

- a. Studentenes helse- og trivselsundersøkelse 2018 viser at det har vært en økning i andel studenter som bruker rusmidler og at mange opplever at dette går utover studieresultatene.
Du jobber i en forskningsgruppe som skal gjennomføre en kvalitativ undersøkelse for å få økt innsikt og kunnskap om hvordan studenter ved NTNU opplever ruskulturen i studiemiljøet sitt.
Lag en problemstilling og beskriv et kvalitativt metodeopplegg for denne studien.
Begrunn valget ditt.

I denne oppgaven står studenten fritt til å utvikle eget metodeopplegg for studien.

Det er avgjørende at det utvikles en forskbar problemstilling og at metoden muliggjør operasjonalisering av problemstillingen.

Oppgaven muliggjør at kandidaten tar i bruk et mangfold av begreper fra kurset. Gode besvarelser er innom alle delene av metoden og reflekterer over dilemma/avveininger som kan oppstå og ha betydning for forskningsprosessen, og hvordan disse evt kan løses/begrenses. Gode besvarelser vil også reflektere over begrensninger i metoden, og evt hvilke metoder (andre kvalitative eller kvantitative) som kan belyse nye sider av problemfeltet.

På grunn av plass og tidshensyn er det imidlertid viktigst at problemstilling og metode(r) er godt beskrevet og begrunnet.

- b. I Tjora (2017) blir en skjematisk stegvis-deduktiv induktiv modell (SDI) for kvalitativ forskning presentert. Forskningsgruppen skal benytte SDI modellen i analysen av datamaterialet dere genererte i undersøkelsen om ruskultur i studiemiljøet (4a). Gjør rede for de ulike stegene i modellen og hvordan den kan anvendes i analysen av datamaterialet deres.

Kandidaten bør kunne gjøre rede for at dette er en skjematisk modell for kvalitativ forskning hvor grunnprinsippet er en induktiv utvikling fra empiri til konsepter eller teorier med deduktive trinnvise tilbakekoblinger. Kandidaten bør kunne vise/beskrive at modellen starter med datagenerering og at de ulike stegene handler om reduksjon av data gjennom analyse. Modellen er grundig gjennomgått på forelesning og (tekstnær) koding, kodegruppering og utvikling av typologier/konsepter/modeller, og teori, bør være kjente begreper for kandidaten. Det forventes at det knyttes eksempler fra opplegget i a til presentasjonen.

En sterk besvarelse viser forståelse for at den induktive utviklingen starter i empirien mens den stegvise deduktive (iterasjoner) handler om kvalitetssikring. Herunder følger det at målsettingen med modellen er teori(konsept)utvikling og kvalitetssikring (trinnvis deduktivt), og at i sum brukes begrepet

**abduktiv/abduksjon. Den ekstra sterke besvarelsen reflekterer rundt teori-
utvikling i SDI-modellen.**

- c. Forskningsgruppen blir bedt om å vurdere hvilke etiske problemstillinger som kan oppstå i gjennomføringen av undersøkelsen. Gjør rede for mulige etiske problemstillinger i datainnsamlingsprosessen og under håndtering, bearbeiding og presentasjon av data, og hvilken relevans de har for deres undersøkelse (4a).

Pensum og forelesninger vektlegger etikk i alle deler av forskningsprosessen. Besvarelsen bør være innom etiske problemstillinger som kan oppstå ved rekruttering av informanter (f.eks eksponering av informant), ved gjennomføring av datainnsamling (f. eks hensyn til informant), ved håndtering av data (anonymisering og trygg oppbevaring) og i presentasjon av data (Anonymitet & transparens)

En sterk besvarelse vil også inkludere en refleksjon over forskerens integritet (kvalitet og uavhengighet).

Det forventes at besvarelsen knyttes til opplegget i a.

Nynorsk

Eksamen er delt i to delar, ein kvantitativ og ein kvalitativ del. Begge delane må svarast på og dei tel 50 % kvar. Begge delar må vere bestått for å få eksamen godkjend (Tips: Start med det du kan og disponer tida godt. Ikkje «heng deg opp» i detaljar).

Lukke til!

Kvantitativ del

Alle tre oppgåver skal svarast på. Oppgåve 1 tel 25%, oppgåve 2 tel 50 % og oppgåve 3 tel 25% av den samla (kvantitative) karakteren.

Oppgåve 1

Gjer kort greie for 3 av dei 4 omgrepa. Bruk inntil ½ side på kvart omgrep.

- a) Operasjonalisering
- b) Kausalitet
- c) Samspel/interaksjon
- d) Kurvelineær samanheng

Oppgåve 2

I analysane under skal vi sjå på to sett av regresjonar (ein lineær og ein logistisk), som kvar har to blokkar. Analysane tek utgangspunkt i heile den europeiske befolkninga og baserer seg på European Social Survey (ESS) frå 2014. ESS er ei multinasjonal spørjeundersøking som analyserer samfunnsvitskapelege spørsmål.

Regresjonsanalysane under skal kunne gje svar på om det er kjønns- og utdanningskilnadar i helse i Europa.

Sjølvrapportert helse er avhengig variabel i alle analysane.

I dei *lineære regresjonsanalysene* går verdiane på den avhengige variabelen frå 0 (veldig dårleg helse), 1 (dårleg helse), 2 (gjennomsnittleg helse), 3 (god helse), til 4 (veldig god helse)

I dei *logistiske regresjonsanalysene* har vi koda om den same variabelen til ein dikotom/todelt variabel, med verdiane 1 (veldig dårleg helse, dårleg helse og gjennomsnittleg helse) og 0 (god helse og veldig god helse).

Forklaringsvariablane (x-variablane) er dei same både for dei lineære og logistiske regresjonsanalysane.

Blokk 1 (for alle analysar):

Alder: kontinuerleg variabel frå 25 til 75 år.

Kjønn: *kvinne* (koda 1), *mann* (koda 0).

Utdanningsnivå: Dummyvariablar: *høg*-, *middels*- og *låg utdanning*, kor høg utdanning er referanse.

I Blokk 2 vert følgjande variablar lagt til analysane.

Inntak av frukt og grønnsaker: Dummyvariablar: *Høgt inntak* (referanse), *moderat inntak*, og *lågt inntak*.

Alkoholkonsum: Dummyvariablar: *Høgt forbruk*, *moderat forbruk*, og *lågt forbruk* (referanse).

Fysisk aktivitet: Dummyvariablar: *Fysisk aktiv* (koda 0) og *fysisk inaktiv* (koda 1).

Røking: Dummyvariablar: *Ikkje-røykjar* (referanse), *tidlegare røykjar*, og *røykjar*.

Ta utgangspunkt i Tabell 1 (samt eventuelt appendiks):

- a) Viser analysen at det er kjønns- og utdanningsskilnadar i Europa i rapportering av sjølv-rapportert helse? Forklar kvifor/kvifor ikkje.
- b) Kan vi ut frå tabellen forklare desse skilnadane? Forklar i så fall kvifor og eventuelt kor mykje.
- c) Korleis kan vi tolke regresjons-konstanten?
- d) *Prediker* nivået av sjølvrapportert helse for ei lågt utdanna kvinne på 74 år, som har lågt inntak av frukt/grønt, som er fysisk inaktiv, som drikk lite alkohol og som røykjer.
- e) *Prediker* nivået av sjølvrapportert helse for ein høgt utdanna mann på 25 år, som har høgt inntak av frukt/grønt, som er fysisk aktiv, som har høgt alkoholforbruk og som ikkje røykjer.
- f) Kor mykje av variasjonen i sjølvrapportert helse er vi i stand til å forklare ut i frå analysane i Tabell 1?

Ta utgangspunkt i Tabell 2 (samt eventuelt appendiks).

- g) Gjer også denne analysen grunnlag for å hevde at vi har kjønns- og utdanningsskilnadar i sjølvrapportert helse i Europa? Forklar kvifor/kvifor ikkje.
- h) Gjer den logistiske regresjonsanalysen grunnlag for å trekkje nokon andre konklusjoner samanlikna med den lineære regresjonsanalysen?
- i) *Prediker* sannsynet for at ei lågt utdanna kvinne på 74 år, som har lågt inntak av frukt/grønt, som er fysisk inaktiv, som drikk lite alkohol og som røykjer, har dårleg sjølvrapportert helse. Du kan nytte følgjande formel:

$$\hat{P} = \frac{1}{1 + e^{-\hat{L}}}$$

- j) Bruk same formel til å *predikere* sannsynet for at ein høgt utdanna mann på 25 år, som har høgt inntak av frukt/grønt, som er fysisk aktiv, som har høgt alkoholforbruk og som ikkje røykjer, har dårleg sjølvopplevd helse.

Oppgåve 3

- a) Diskuter kva for ein av dei to modellene (den lineære eller den logistiske) du tykkjer høver best. Det er ikkje sikkert svaret er opplagt, og du kan gjerne diskutere for og imot.
- b) Tenk deg at du hadde tilgang til desse dataa, samt eigna analyseverktøy (som til dømes STATA eller SPSS). Korleis ville du gått fram for å betre modellane?

EKSTRA-OPPGÅVE

Her kan du få plusspoeng. INGENTING trekkjast frå dersom du ikkje svarar på denne.

Forskning skal danne grunnlag for politikk. Korleis hadde du kommunisert funna til politikarane våre, og kva for råd hadde du gjeve?

Kvalitativ del

Oppgåve 4 (tel 20% av den samla kvalitative karakteren)

Gjer kort greie for tre av dei fem omgrepa. Bruk maksimum ein ½ side på kvart omgrep

- Deltakande observasjon
- Induktiv
- Transparens
- Pålitelegheit (reliabilitet)
- Forskarsubjektivitet

Oppgåve 5 (tel 80% av den samla kvalitative karakteren)

- a. Studentane si helse- og trivselsundersøking 2018 syner at det har vore ein auking i andel studentar som bruker rusmiddel og at mange opplever at dette går utover studieresultantane.
Du jobbar i ei forskingsgruppe som skal gjennomføre ein kvalitativ undersøking for å få auka innsikt og kunnskap om korleis studentar ved NTNU opplever ruskulturen i studiemiljøet sitt.
Lag ein problemstilling og beskriv eit kvalitativt metodeopplegg for denne studien. Grunnlegg valet ditt.
- b. I Tjora (2017) vert ein skjematisk stegvis-deduktiv induktiv modell (SDI) for kvalitativ forskning presentert. Forskingsgruppa skal nytte SDI modellen i analysen av datamaterialet de genererte i undersøkinga om ruskultur i studiemiljøet (4a). Gjør greie for dei ulike stega i modellen og korleis den kan nyttast i analysen av datamaterialet dykkar.
- c. Forskingsgruppa blir bede om å vurdere kva slags etiske problemstillingar som kan oppstå i gjennomføringa av undersøkinga. Gjer greie for moglege etiske problemstillingar i datainnsamlingsprosessen og under handtering, handsaming og presentasjon av data, og kva slags relevans dei har for dykkar undersøking (4a).

English

The exam has two parts, a quantitative and a qualitative one, which count 50% each. Both parts have to be answered and passed to pass the exam. (Tip: Answer those questions first that you know immediately. Do not get caught up in too many details.)

God luck!

Quantitative section

Please answer all three exercises. Exercise 1 counts 25%, exercise 2 counts 50% and exercise 3 counts 25% of the quantitative grade.

Exercise 1

Give a short description of 3 out of the 4 concepts. Use up to ½ a page on each concept.

- a) Operationalization
- b) Causality
- c) Interaction effect
- d) Curvilinearity

Exercise 2

The tables below present two sets of regression analyses (one linear and one logistic regression analysis), both of which have two blocks. The analyses are based on the entire European sample from the European Social Survey (2014). The ESS is a cross-national survey, which aims at giving answers to research questions within the social sciences.

The regression analyses below should be able to address whether there are differences in self-perceived health between men and women, and between educational groups.

Self-perceived health is the dependent variable in all analyses.

In the *linear regression analyses*, the values of the dependent variable are the following: 0= very poor health, 1= poor health, 2=fair health, 3= good health, and 4=very good health.

In the logistic regression analyses, the dependent variable is recoded to a dichotomous variable with the following values: 1= very poor, poor and fair health, 0=good and very good health.

The explanatory variables (x-variables) are identical in all analyses (i.e. both for the linear and logistic regression analyses).

Block 1 (for all analyses):

Age: Continuous variable ranging from 25 to 75 years of age.

Sex: *Women* (coded 1), *men* (coded 0).

Educational level: Dummy variables: *high-*, *medium-* and *low education*, where high education is the reference category.

The following variables are added in Block 2 (in all analyses):

Fruit and vegetable consumption: Dummy variables: *High intake* (reference), *moderate intake*, and *low intake*.

Alcohol consumption: Dummy variables: *High consumption*, *moderate consumption*, and *low consumption* (reference).

Physical activity: Dummy variables: *physical activity* (coded 0), and *physical inactivity* (coded 1).

Smoking: Dummy variables: *Non-smoker* (reference), *former smoker*, and *smoker*.

Study Table 1 (and the appendix if need be):

- a) Does the analysis demonstrate any differences in health according to sex and education in Europe? Please explain why/why not.
- b) Can Table 1 explain any observed differences in health? Explain how and the extent to which this is the case.
- c) Please interpret the regression coefficient.
- d) Please *predict* the level of self-perceived health of a low educated woman aged 74, who has a low intake of fruit/vegetables, who is physically inactive, who has a low intake of alcohol, and who is a smoker.
- e) Please *predict* the level of self-perceived health of a highly educated man aged 25, who has a high intake of fruit/vegetables, who is physically active, who drinks a lot and smokes.
- f) How much of the variation in self-perceived health is explained in Table 1?

Study Table 2 (and the appendix if need be):

- g) Does this analysis also support the view that there are differences in self-perceived health between men and women, and between educational groups in Europe? Explain why/why not.
- h) Does the logistic regression analysis give reason to draw any different conclusions as compared to Table 1?
- i) Please *predict the probability* that a low educated woman aged 74 low educated woman aged 74, who has a low intake of fruit/vegetables, who is physically inactive, who has a low intake of alcohol, and who is a smoker, has poor health. You may use the following formula:

$$\hat{P} = \frac{1}{1 + e^{-\hat{L}}}$$

- j) Use the same formula to *predict the probability* that a high educated man aged 25, who has a high intake of fruit/vegetables, who is physically active, and who drinks a lot and smokes, has poor health.

Exercise 3

- a) Discuss which of the two models (the linear regression or the logistic regression respectively) you find most suitable. It may not be that the answer is given, implying that you can find arguments in favor of, and against both models.
- b) Imagine that you had access to the ESS data and suitable analytical tools (such as STATA eller SPSS). How would you go about to improve the regression models?

EXTRA EXERCISE

This is an opportunity to gather some extra points if you provide a good answer. There will be NO negative points (no reduction in the grade) if you do not answer this exercise.

Research should inform policy. How would you communicate the findings above to our politicians, and which recommendations would you suggest?

Qualitative part

Exercise 4 (counts 20% of the qualitative grade)

Briefly explain three of the five concepts. Use maximum a ½ page on each concept

- a) Participatory observation
- b) Inductive
- c) Transparency
- d) Reliability (qualitative)
- e) Researcher subjectivity

Exercise 5 (counts 80% of the qualitative grade)

- a) The students' health and well-being survey 2018 shows that there has been an increase in the proportion of students who use drugs, and that many experience that it influences on study results.
You work in a research group that will conduct a qualitative study to gain insight and knowledge about how students at NTNU experience the drug culture in their study environment.
Develop a research question and describe a qualitative methodology for this study.
Explain your choice.
- b) In Tjora (2017) a schematic stepwise-deductive-inductive model (SDI) for qualitative research is presented. The research group will employ the SDI model in the analysis of the data you generated in the study of the drug culture in the study environment (4a). Explain the different steps in the model and how it can be used in the analysis of your data.
- c) The research group is being asked to consider which ethical issues that may arise in conducting the study. Explain possible ethical issues in the data generation process, during the handling, processing and presentation of data, and the relevance of these issues for your study (4a).

Tabell 1: Lineær regresjonsanalyse av selvopplevd helse (STATA output i appendiks)

Variabler	Verdier	Blokk 1			Blokk 2		
		B	St.f.	Sig.	B	St.f.	Sig.
Konstant		3.90	0.00	0.00	3.97	0.03	0.00
Alder	(kontinuerlig)	-0.02	0.01	0.00	-0.17	0.00	0.00
Kjønn	Kvinne	-0.08	0.01	0.00	-0.06	0.01	0.00
	Mann	0			0		
Utdanning	Høy	0			0		
	Middels	-0.20	0.12	0.00	-0.13	0.12	0.00
	Lav	-0.40	0.15	0.00	-0.28	0.15	0.00
Frukt/grønt	Høyt inntak				0		
	Moderat inntak				-0.16	0.13	0.23
	Lavt inntak				-0.28	0.14	0.00
Fysisk aktivitet	Aktiv				0		
	Inaktiv				-0.25	0.01	0.00
Alkoholkonsum	Lavt				0		
	Moderat				0.13	0.13	0.00
	Høyt				0.23	0.12	0.00
Røyking	Ikke-røyker				0		
	Tidligere røyker				-0.05	0.12	0.00
	Røyker				-0.17	0.12	0.00

B=lineær regresjonskoeffisient, st.f.= standardfeil, sig=signifikans.

Tabell 2: Logistisk regresjonsanalyse av selvopplevd helse (STATA output i appendiks)

Variabler	Verdier	Blokk 1				Blokk 2			
		OR	Logit	St.f.	Sig.	OR	Logit	St.f.	Sig.
Konstant		0.03	-3.55	0.00	0.00	0.02	-3.82	0.00	0.00
Alder	(kontinuerlig)	1.04	0.00	0.00	0.00	1.04	0.04	0.00	0.00
Kjønn	Kvinne Mann	1.27 1	0.24	0.03	0.00	1.20	0.18	0.04	0.00
Utdanning	Høy Middels Lav	1 1.68 2.61	0.52 0.96	0.06 0.11	0.00 0.00	1.47 2.07	0.38 0.73	0.06 0.09	0.00 0.00
Frukt/grønt	Høyt inntak Moderat inntak Lavt inntak					1 1.01 1.34		0.01 0.04 0.06	0.77 0.00
Fysisk aktivitet	Aktiv Inaktiv					1 1.75		0.56 0.05	0.00
Alkoholkonsum	Lavt Moderat Høyt					1 0.78 0.58		-0.25 0.29 -0.54	0.00 0.00
Røyking	Ikke-røyker Tidligere røyker Røyker					1 1.11 1.45		0.11 0.04 0.05	0.00 0.00

OR=odds ratio, st.f.= standardfeil av OR, sig=signifikans.

Table 1: Linear regression analysis (STATA output in appendix)

Variables	Values	Block 1			Block 2		
		B	s.e.	Sig.	B	S.e.	Sig.
Constant		3.90	0.00	0.00	3.97	0.03	0.00
Age	(continuous)	-0.02	0.01	0.00	-0.17	0.00	0.00
Sex	Women	-0.08	0.01	0.00	-0.06	0.01	0.00
	Men	0			0		
Education	High	0			0		
	Medium	-0.20	0.12	0.00	-0.13	0.12	0.00
	Low	-0.40	0.15	0.00	-0.28	0.15	0.00
Fruit/veg	High intake				0		
	Moderate intake				-0.16	0.13	0.23
	Low intakea				-0.28	0.14	0.00
Physical activity	Active				0		
	Inactive				-0.25	0.01	0.00
Alcohol consumption	Low				0		
	Moderate				0.13	0.13	0.00
	High				0.23	0.12	0.00
Smoking	Non-smoker				0		
	Former smoker				-0.05	0.12	0.00
	Smoker				-0.17	0.12	0.00

B=linear regression coefficient, s.e.=standard error, sig=significance

Table 2: Logistic regression analysis (STATA output in appendix)

Variables	Values	Block 1				Block 2			
		OR	Logit	S.e.	Sig.	OR	Logit	S.e.	Sig.
Constant		0.03	-3.55	0.00	0.00	0.02	-3.82	0.00	0.00
Age	(continuous)	1.04	0.04	0.00	0.00	1.04	0.04	0.00	0.00
Sex	Women	1.27	0.24	0.03	0.00	1.20	0.18	0.04	0.00
	Men	1							
Education	High	1							
	Medium	1.68	0.52	0.06	0.00	1.47	0.38	0.06	0.00
	Low	2.61	0.96	0.11	0.00	2.07	0.73	0.09	0.00
Fruit/veg	High intake					1			
	Moderate intake					1.01	0.01	0.04	0.77
	Low intakea					1.34	0.30	0.06	0.00
Physical activity	Active					1			
	Inactive					1.75	0.56	0.05	0.00
Alcohol consumption	Low					1			
	Moderate					0.78	-0.25	0.29	0.00
	High					0.58	-0.54	0.20	0.00
Smoking	Non-smoker					1			
	Former smoker					1.11	0.11	0.04	0.00
	Smoker					1.45	0.37	0.05	0.00

OR=odds ratio, s.e.= standard error of OR, sig=signifikans.

APPENDIX: STATA OUTPUT

Block 1: linear regression analysis

Source	SS	df	MS	Number of obs	=	27,533
Model	2425.28562	4	606.321405	F(4, 27528)	=	891.26
Residual	18727.2376	27,528	.680297792	Prob > F	=	0.0000
				R-squared	=	0.1147
				Adj R-squared	=	0.1145
Total	21152.5232	27,532	.768288654	Root MSE	=	.8248

SRH	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
agea	-.0166789	.0003715	-44.90	0.000	-.017407	-.0159508
female	-.0811908	.0099598	-8.15	0.000	-.1007125	-.0616691
edu012						
1	-.1966763	.0120809	-16.28	0.000	-.2203554	-.1729971
2	-.4041244	.0147431	-27.41	0.000	-.4330217	-.3752271
_cons	3.897667	.0205256	189.89	0.000	3.857436	3.937898

Block 2: linear regression analysis

Source	SS	df	MS	Number of obs	=	26,472
Model	3252.1556	11	295.650509	F(11, 26460)	=	459.55
Residual	17023.146	26,460	.643353967	Prob > F	=	0.0000
				R-squared	=	0.1604
				Adj R-squared	=	0.1601
Total	20275.3016	26,471	.765943922	Root MSE	=	.80209

SRH	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
agea	-.0173903	.0003779	-46.02	0.000	-.0181311	-.0166496
female	-.0605779	.0104345	-5.81	0.000	-.0810302	-.0401257
edu012						
1	-.1294428	.0121857	-10.62	0.000	-.1533274	-.1055582
2	-.2846377	.0150201	-18.95	0.000	-.3140779	-.2551974
fruitveg						
1	-.0162224	.0134024	-1.21	0.226	-.0424918	.010047
2	-.137008	.0142982	-9.58	0.000	-.1650333	-.1089827
1.pa_guide	-.2534667	.0100429	-25.24	0.000	-.2731514	-.233782
alcohol_consumption						
1	.1269648	.013036	9.74	0.000	.1014137	.152516
2	.2339603	.012176	19.21	0.000	.2100947	.2578258
smokebhv						
1	-.0534955	.0124604	-4.29	0.000	-.0779186	-.0290723
2	-.1742277	.0124786	-13.96	0.000	-.1986863	-.1497691
_cons	3.969581	.0252771	157.04	0.000	3.920036	4.019125

