

Info

Eksamen KRS4105 høst 2023

Om eksamen

Eksamensbesvarelsen skal leveres i Inspera senest fredag 20. oktober kl. 9.00. Du leverer oppgaven ved å laste opp ett PDF-dokument. Vær oppmerksom på at kun én fil skal leveres inn.

Alle hjelpemidler er tillatt under eksamen, inkl. lærebok, notater, forelesningsslides og andre ressurser. Både oppgaveløsningen og selve besvarelsen skal være et individuelt arbeid, og alle former for samarbeid er derfor forbudt. Plagiatkontroll vil bli gjennomført for å sikre at alle leverer en selvstendig besvarelse.

Besvarelsen skal være maksimalt på 4000 ord, som tilsvarer om lag 10 sider med de fleste moderne skrifttyper, font 12 og linjeavstand 1½. Forside, innholdsfortegnelse, kode, tabeller, figurer og litteraturliste kommer i tillegg. Besvarelser på over 4000 ord vil ikke bli sensurert.

- Når du presenterer resultater med mange desimaler runder du av sånn som du selv synes er rimelig.
- Når oppgaveteksten angir at du skal «tolke» resultater innebærer dette at du skal gi en intuitiv tolkning av betydningen av resultatet, og det holder altså ikke å henvise kun til en tallverdi.
- Du står fritt til å velge signifikansnivå (alpha).
- I oppgavene hvor du må vise koden du har benyttet kan du ha en del med tittel "Kode" hvor du limer inn koden du har benyttet som tekst (helst), eller bilde. Hvis du velger å lime den inn som bilde, må det være synlig på bildet hvilke kommandoer du har benyttet.
- Tabellene skal helst limes inn som bilder.
- Noen av oppgavene har veiledende lengde oppgitt i oppgaveteksten. Dette er kun veiledende og dine svar kan besvare spørsmålet også med færre ord. Hvis du har behov kan du også skrive noe lenger enn det som er angitt.
- Hvis du opplever problemer med R på eget PC kan du benytte UIOs R workbench løsning (<https://www.uio.no/tjenester/it/utdanning/rstudioworkb/index.html>)

Du kan få maksimalt 64 poeng på eksamen. Hver oppgave følges av en parentes som angir det maksimale antallet poeng du kan få på hver oppgave. For å bestå må du ha minst 16 poeng.

Du skal benytte to datasett for å besvare disse oppgavene. Datasettene finnes i Canvas under Filer - Eksamen 2023. Begge datasettet er i Stata (*.dta) format.

Oppgave 1 (total 34 poeng)

I denne oppgaven skal du undersøke sammenhengen mellom å kjenne en politibetjent og tillit til at politiet (som yrkesgruppe) gjør en god jobb.

Du skal benytte datasettet `tillit_politi.dta` som er et Stata-datasett for å besvare delspørsmålene under. Dataene er fra en spørreundersøkelse som ble gjennomført høsten 2022. Undersøkelsen skal være representativ for den norske befolkningen i forhold til alder, kjønn og geografi. Likevel er andelen som har høyere utdanning høyere blant respondentene enn den er i befolkningen.

Datasettet inneholder følgende variabler:

- **RespondentID** - ID-nummer for respondenten
- **tillit_politi** - Hvilken tillit har du til at politibetjenter gjør en god jobb? Svaralternativer: fra 1 - "Svært lav tillit" til 7 - "Svært høy tillit"
- **alder** - Hva er din alder? Svaralternativer: fritekst, numerisk
- **kvinne** - Registrer ditt kjønn. Svaralternativer: 0 = Menn, 1 = Kvinner
- **people_fair** - Tror du at folk flest vil prøve å utnytte deg hvis de fikk sjansen, eller vil de prøve å behandle deg rimelig og rettferdig? Svaralternativer: 1 - "Folk flest vil prøve å utnytte meg" til 10 - "Folk flest vil prøve å være rimelige og rettferdige"
- **fodel1** - Er du født i Norge? Svaralternativer: 1 = Ja, 2 = Nei
- **utd6delt** - Hva er din høyeste fullførte utdanning? Svaralternativer: 1 - "Grunnskoleutdanning - 10-årig, grunnskole, 7-årig folkeskole eller lignende", 2 - "Videregående allmennfaglig utdanning", 3 - "Videregående yrkesfaglig utdanning", 4 - "Fagskole – Yrkesrettede utdanninger (1/2 – 2 år) som bygger på videregående", 5 - "Universitets-/høgskoleutdanning med inntil 4 års varighet", 6 - "Universitets-/høgskoleutdanning med mer enn 4 års varighet"
- **kjennerPoliti** - Kjenner du noen som jobber som politibetjent? Svaralternativer: 0 - "Nei", 1 - "Ja, som familiemedlem eller venn"

Spørsmål 1a) (1 poeng)

Last inn datasettet `tillit_politi.dta` i R/RStudio. Du skal vise koden du har brukt for å laste inn datasettet. Fra koden må navnet du gir til datasettet i R fremkomme. Husk å fjerne delen av stien som inneholder ditt brukernavn!

Spørsmål 1b) (1 poeng)

Hvor mange rader (observasjoner) og kolonner (variabler) har du i datasettet?

Spørsmål 1c) (3 poeng)

Presenter gjennomsnittet og standardavviket for følgende variabler: `alder`, `people_fair`, og `tillit_politi`.

Vis koden som du har brukt for å lage tabellen og kommenter kort resultatet. Du skal oppgi hva som er gjennomsnittsverdien og standardavviket for hver variabel og kommentere hvor mye missing (ikke kjent verdi) det er på hver variabel.

Spørsmål 1d) (5 poeng)

Presenter en deskriptiv tabell som inneholder andelene for variablene kvinne, fodel1, utd6felt, kjennerPoliti.

Vis koden du har benyttet for å lage tabellen og kommenter kort resultatet. Du skal si noe kort om fordelingen på hver variabel, samt kommentere hvor mange observasjoner som har missing-verdier på disse variablene.

Spørsmål 1e) (3 poeng)

Kod om variabelen utd6felt slik at kun de som oppgir at de har fullført høyere utdanning (Universitets-/høgskoleutdanning) skal få verdien 1, mens de øvrige svaralternativene skal få verdien 0. Gi den nye variabelen du lager navnet utd2kat. Vis koden du bruker for denne omkodningen, samt fordelingen på denne variabelen.

Spørsmål 1f) (3 poeng)

Kjør en bivariat regresjon med tillit_politi som avhengig variabel og alder som uavhengig variabel. Du skal vise koden du har benyttet, presentere resultatene i en tabell, deretter skal du tolke samtlige koeffisienter, vurdere om de er substansielt meningsfulle og statistisk signifikante.

Spørsmål 1g) (4 poeng)

Du snakker med en kollega som påstår at sammenhengen mellom alder og tillit til at politibetjenter gjør en god jobb er ikke-lineær. Hen påstår at tilliten til at politibetjenter er uformet i forhold til alder. Undersøk denne påstanden ved hjelp av en regresjonsmodell. Vis koden du har brukt for å kjøre regresjonsmodellen og kommenter resultatet kort. Din kommentar skal svare på: Er det slik at sammenhengen mellom alder og tillit til politiet best beskrives av en rett linje basert på den regresjonen du har estimert eller er sammenhengen ikke-lineær?

Spørsmål 1h) (totalt 6 poeng)

Kjør en modell hvor tillit til at politibetjenter gjør en god jobb (tillit_politi) er avhengig variabel. Du skal kontrollere for om respondentene kjenner en politibetjent, alder, kjønn og om de har høyere utdanning (variabelen fra oppgave 1e).

Du må vise koden du har benyttet, presentere resultatene i en tabell, deretter skal du tolke samtlige koeffisienter og kommentere om de er statistisk signifikante.

Spørsmål 1i) (totalt 2 poeng)

Du får tilbakemelding fra en kollega om at din modell ikke tar høyde for folks syn på andre mennesker (people_fair). Du skal bygge videre på modellen fra oppgave 1h og legger til variabelen people_fair. Tolk kun koeffisienten for people_fair og beregn 95%-konfidensintervallet for denne variabelen. Du skal legge inn variabelen som kontinuerlig.

Spørsmål 1j) (totalt 4 poeng)

Du skal nå undersøke følgende spørsmål: Er sammenhengen mellom at respondentene tenker at folk prøver å utnytte dem (people_fair) og tillit til at politiet gjør en god jobb (tillit_politi) avhengig av om de selv kjenner en politibetjent?

Bygg videre på modellen fra oppgave 1i) med et samspillsledd mellom people_fair og kjennerPoliti. Tolk samspills-koeffisienten(e). I tillegg må du presentere koden du har brukt for å estimere modellen og resultatet fra modellen i en tabell.

Spørsmål 1k) (totalt 1 poeng)

Vurder om du også skal kontrollere for om respondentene er født i Norge eller ikke i modellen du har estimert under 1h). Hva ville vært hovedfordelene og ulempene med å inkludere denne variabelen? (Forventet lengde: fra 20 til 150 ord)

Spørsmål 1l) (totalt 1 poeng)

Hva er ulempene med å benytte et spørsmål som “Hvilken tillit har du til at politibetjenter gjør en god jobb?” med svaralternativer: 1 - “Svært lav tillit” til 7 - “Svært høy tillit” som kontinuerlig avhengig variabel i en regresjonsmodell? (Forventet lengde: fra 20 til 175 ord)

Oppgave 2 (30 poeng)

Inntil nokså nylig kunne amerikanske arbeidsgivere spørre arbeidssøkere om de har blitt dømt for en forbrytelse. Dette var ofte i form av en avkryssingsboks for “ja” og “nei”. “Ban the Box” (forkortet BTB) er en amerikansk kampanje ledet av talspersoner for tidligere straffedømte, med mål om å fjerne avkryssingsboks som spør om søkerne har et kriminelt rulleblad fra jobbsøknader. Fra og med 1.03.2015 i New Jersey og fra og med 27.10.2015 i New York ble det forbudt også for arbeidsgivere i privat sektor å stille spørsmål om kriminelt rulleblad i jobbsøknader eller under intervjuer.

For å besvare oppgaven trenger du ikke å ha kjennskap til denne politikken eller til det amerikanske systemet.

I denne oppgaven skal du benytte følgende data. Dataene består av svarene for nesten 15 000 fiktive jobbsøknader på vegne av unge menn til ulike stillinger, før og etter de effektive datoene for innføring av BTB-forbudet for arbeidsgivere i privatsektor i New Jersey (1. mars 2015) og New York City (27. oktober 2015). Søknadene ble sendt i par der søkerne var like når det gjaldt om de identifiserte seg som svart eller hvit, og det ble også **tilfeldig variert om søkeren hadde en straffedom**.

Forfatterne av denne studien studerer også om det har blitt mer diskriminering etter BTB-forbudet på bakgrunn av hudfarge. Forfatterne har en hypotese om at etter BTB-forbudet blir arbeidssøkere som identifiserer seg som svarte i større grad diskriminert, fordi arbeidsgivere i større grad tenker at de er lovbrøyttere.

Datasettet har navnet oppgave2.dta og inneholder følgende variabler

- **response** - Om arbeidsgivere har tatt kontakt for å invitere til intervju. Koder: 0 - “arbeidsgiver har ikke tatt kontakt”; 1 - “arbeidsgiver har tatt kontakt og inviterte til intervju” (*engelsk: “Application rec’d Positive Response”*)
- **white** - Om fiktiv arbeidssøker identifiserte som hvit eller svart i jobbsøknaden. Koder: 0 - “arbeidssøker identifiserer seg selv om svart”; 1 - “arbeidssøker identifiserer seg selv som hvit”
- **crimbox** - Om arbeidssøker må krysse “ja” eller “nei” for om hen har blitt straffedømt. Koder: 0 - “ikke avkrysningsboks for domfellelse”; 1 - “avkrysningsboks for domfellelse” - (*engelsk: Application has Box*).
- **crime** - Om fiktiv arbeidssøker hadde krysset av for at hen var dømt for en forbrytelse. Koder: 0 - “arbeidssøker krysset nei, dvs. ikke dømt for forbrytelse”; 1 - “arbeidssøker krysset ja, dvs. arbeidssøker er dømt for forbrytelse” (*engelsk: Applicant has Criminal Record*)
- **empgap** - Om fiktiv arbeidssøker hadde hull i CVen (dvs. lengere periode med arbeidsledighet). Koder: 0 - “ikke hull i CV-en”; 1 - “fiktiv arbeidssøker har hull i CV-en” (*engelsk: Applicant has Employment Gap*)
- **ged** - Om fiktiv arbeidssøker hadde fullført videregående. Koder: 0 - “ikke fullført videregående”; 1 - “fullført videregående” (*engelsk: Applicant has GED*)
- **pre** - Om dataene er samlet inn før BTB-forbudet trådte i kraft. Koder: 0 - “etter forbudet mot å spørre om domfelles (BTB) trådte i kraft” (*engelsk: =0 if post-BTB period*); 1 - “perioden før forbudet mot BTB trådte i kraft” (*engelsk: =1 if pre-BTB period*)
 - Hvis pre har kode 1 betyr det at dataene er samlet inn mens det var lov å spørre om arbeidssøkere har blitt dømt for en forbrytelse.

Når dere laster inn datasettet kommer disse navnene til å være på engelsk. Det er kun høye verdier som har labels i datasettet.

Oppgave 2a) (4 poeng)

Designet som er beskrevet her et felteksperiment. Hvis du kun skal undersøke perioden **før det ble ulovlig å spørre om personen var straffedømt**, kunne du, med en slikt design si noe kausalt, i så fall om hva? Kan dette svaret generaliseres til alle arbeidsgivere i USA? (100 til maks 400 ord).

Oppgave 2b) (6 poeng)

Man kan sammenligne perioden før BTB-forbudet ble innført og perioden etter ved å bruke en forskjeller-i-forskjeller (difference-in-difference) design. Hva er en forskjeller-i-forskjeller-design og hvilke antakelser må være på plass/undersøkes for at en slik design skal muliggjøre identifikasjonen av kausal effekter? (150 til cirka 550 ord)

Oppgave 2c) (6 poeng)

Beskriv med ord hvordan du ville estimert ved hjelp av en regresjon om forskjellene i å bli kontaktet av arbeidsgivere endres mellom de som identifiserer som svarte, eller hvite, før og etter reformen? Beskriv hva som er avhengig variabel, uavhengige variabler og hvordan du ville

bygget opp modellen. For denne oppgaven skal du ta utgangspunkt i variablene inkludert i datasettet med navnet oppgave2.dta (200 til cirka 400 ord)

Oppgave 2d) (2 poeng)

Last inn datasettet med navnet oppgave2.dta. Hvor mange observasjoner (rader) og variabler (koloner) har du i datasettet?

Du skal vise koden du har brukt for å laste inn datasettet (Husk å fjerne brukernavn fra stien).

Oppgave 2e) (2 poeng)

I denne oppgaven skal du kun undersøke forskjeller før forbudet trådte i kraft. Hvor stor andel av dem som identifiserte seg som hhv. svarte og hvite ble kontaktet og invitert til intervju?

Tips: Du skal kun ha med data fra perioden før forbudet trådte i kraft, dvs. `pre==1`

Oppgave 2f) (2 poeng)

I denne oppgaven skal du kun undersøke forskjeller etter forbudet trådte i kraft. Er det en statistisk signifikant forskjell mellom dem som identifiserte seg som hvite eller svarte i om de ble kontaktet og invitert til intervju?

Du skal også ha med koden du brukte til å undersøke dette i svaret.

Tips: Du skal kun ha med data fra perioden etter forbudet trådte i kraft, dvs. `pre==0`

Oppgave 2g) (8 poeng)

Er det flere, færre, eller ingen endring i hvor stor andel av dem som identifiserte seg som svarte som ble kontaktet og invitert til intervju etter at endringen trådte i kraft (`pre == 0`) relativt til situasjonen før (`pre == 1`)?

Forklar hvordan du undersøker dette empirisk og hva du har kommet fram til. Hvordan tolker du dette resultatet? Kan vi si noe om årsakssammenhenger basert på din analyse? Hva er mulige svakheter av din analysemetode og hvordan kunne designet styrkes, hvis du hadde tilgang til mer data?

I tillegg skal du ha med koden du har brukt for å undersøke dette og tabellen(e) som presenterer resultatet.