

i Forside

SOSGEO1120 - Kvantitativ metode

- Skoleeksamen
- 24. mai 2019 kl 14.30 (4 timer)

Om eksamen

Eksamen består av tre bolker. Les informasjonsdokumentet før hver bolk nøye.

Bolk 1: Oppgaver med STATA

Bolk 2: Flervalgsoppgaver

Bolk 3: Håndregning

Alle bolker må besvares for å bestå eksamen.

Du kan bla deg frem og tilbake ved hjelp av piltastene, eller ved hjelp av innholdsfortegnelsen nederst.

Eksamensoppgaven er på norsk. Du kan besvare eksamenen på norsk, svensk, dansk eller engelsk.

Lykke til!

i STATA

Denne bolken innehold tre hovedoppgåver. Du skal bruke STATA til å svare på oppgåvene.

Denne bolken består av tre hovedoppgaver. Du skal benytte STATA for å løse oppgavene.

I noen oppgaver er det et åpent tekstfelt der du kan oppgi kodene du har brukt.

Altså: skriv kode i STATA slik du er vant til. Når du har besvart oppgavene kopierer du koden fra STATA over i tekstfeltet. Du får ikke poeng for STATA-koden, men koden kan være korrigerende hvis f.eks. feil svar skyldes en slurvefeil etc.

2.1 V19 STATA kommuner

Informasjon om oppgaven

Datasettet vil lagres i mappen for nedlastinger.

Vedlagt i denne oppgaven er et datasett i stata-format her: [V19_kommuner](#)

I STATA kan du spesifisere filbane slik:

```
cd C:/Users/brukernavn/Downloads/
```

der du bytter ut '**brukernavn**' med ditt uio-brukernavn.

Se også vedlagte instruksjoner om å laste ned/åpne datasett.

Deretter kan du bruke use-kommandoen til å åpne datasettet. (Se evt. vedlagte veiledning).

Av tekniske grunner vil Inspira endre navnet på datafilen når den lastes ned, ved å legge til en del siffer på slutten av filnavnet. Du må derfor passe på å gi datafilen riktig navn når du leser det inn. (Du kan evt. endre filnavnet i nedlastingsmappen før du åpner filen).

Kommandoen for å åpne datasettet kan da se noe slikt ut:

```
use V19_kommuner_43132846_1556266463244.dta
```

(Hvis *Inspira* skulle komme til å endre filnavnet slik at det ikke slutter på *.dta* skal use-kommandoen fungere likevel hvis filnavnet er likt som den filen som ligger i download-mappen).

For sikkerhets skyld er også datasettet i csv-format her: [V19 kommuner csvformat](#) og kan åpnes med insheet-kommandoen slik:

insheet using V19_kommuner_43132846_1556266463244.csv

Disse dataene er aggregerte tall for norske kommuner i 2016 og inneholder følgende variable:

vold - antall voldshendelser anmeldt til politiet per 1000 innbyggere

prop_unge_menn - andel menn i alderen 18-34 år i forhold til totalt antall innbyggere

prop_unge_kvinner - andel kvinner i alderen 18-34 år i forhold til totalt antall innbyggere

stor_by - en dummy-variabel der verdien 1 indikerer at kommunen er en av følgende: Oslo, Bergen, Trondheim, Kristiansand eller Stavanger, og verdien 0 er øvrige kommuner.

Du skal her bruke hensiktsmessige teknikker og kommandoer i STATA til å besvare spørsmålene nedenfor.

a) Hvilke av følgende utsagn vil du si beskriver best fordelingen av voldskriminalitet i norske kommuner?

Velg ett alternativ

- Fordelingen er høyreskjev (hale til høyre)
- Fordelingen er venstreskjev (hale til venstre)
- Tilnærmet normalfordelt
- Symmetrisk fordeling

b) Undersøk i kva grad det er forskjell i gjennomsnittleg valdsrate i dei store byane samanlikna med dei andre kommunane.

Hva er differansen i gjennomsnittene? Skriv svaret uten fortegn her

c) Gjennomfør en t-test for å se om voldsraten er lik i byer vs. øvrige kommuner. Vil du da bruke en ensidig test eller en tosidig test?

Velg ett alternativ

- Tosidig test. H_0 er at byer og øvrige kommuner har lik voldsrate, mens H_a er at byene har forskjellig voldsrate enn øvrige kommuner.
- Ensidig test. H_0 er at byer og øvrige kommuner har lik voldsrate, mens H_a er at byene har høyere voldsrate enn øvrige kommuner.
- Tosidig test. H_0 er at byer og øvrige kommuner har forskjellig voldsrate, mens H_a er at byene har lik voldsrate som øvrige kommuner.
- Ensidig test. H_0 er at byer og øvrige kommuner har lik voldsrate, mens H_a er at byene har høyere voldsrate enn øvrige kommuner.

d) Hva er p-verdien i foregående deloppgave? Skriv svaret her uten fortegn og med minst 2 desimaler:

e) Formuler kort en konklusjon på t-testen her: (OBS! det er veldig liten plass, så stikkordsform er ok).

f) Hva er median andel unge menn i norske kommuner? (Bruk minst 3 desimaler)

g) Noen vil hevde at det er en sammenheng mellom vold og andelen relativt unge menn i kommunen. Formuler nullhypotese og alternativ hypotese for å teste denne påstanden:

h) Estimer en enkel lineær regresjonsmodell med vold som avhengig variabel og andel unge menn er forklaringsvariabel. Skriv inn stigningstallet (eng: "slope") i regresjonen her:

i) Hvordan vil du tolke usikkerheten ved estimatet i oppgave h)?

Velg ett alternativ

- Konfidensintervallet er ganske bredt og sammenhengen er så usikker at det er vanskelig å konkludere.
- Det er en høy p-verdi og det styrker nullhypotesen om ingen sammenheng
- Estimaten er statistisk signifikant på signifikansnivå 0.01. Vi konkluderer med at det er en sammenheng mellom andel unge menn og vold.
- Estimaten er statistisk signifikant på signifikansnivå 0.01. Vi konkluderer med at høy andel unge menn henger sammen med færre voldshendelser.

j) Estimer en lineær regresjonsmodell med vold som avhengig variabel der du bruker både andel unge menn, andel unge kvinner og variabelen `stor_by` som forklaringsvariable.

I denne modellen, er andel unge kvinner i kommunene assosiert med voldsnivå?

Velg ett alternativ

- Nei. Estimaten tilsier at høyere andel unge kvinner henger sammen med mindre vold, men er ikke statistisk signifikant på signifikansnivå 0.05.
- Ja. Estimaten for kvinner er ikke statistisk signifikant, men estimaten for menn reduseres når andel kvinner øker.
- Ja. Høyere andel unge kvinner er assosiert med en reduksjon i voldshendelser, signifikant på signifikansnivå 0.05
- Nei. Endringen i r^2 er for liten til at det har noe å si.

k) Ta utgangspunkt i estimatene i oppgave j). Hva er riktig tolkning av estimaten for menn? (Hint: vær ekstra nøye med hvilken skala hver variabel er på).

Velg ett alternativ

- Et prosentpoengs økning i unge menn tilsvarer 80% økning i voldshendelser
- Et prosentpoengs økning i unge menn tilsvarer 80 flere voldshendelser per 1000 innbyggere
- Et prosentpoengs økning i unge menn tilsvarer 80 flere voldshendelser
- Et prosentpoengs økning i unge menn tilsvarer en økning på 0.80 voldshendelser per 1000 innbyggere

Maks poeng: 11

2.2 V19 pussig regresjon

Bruk datasettet i forrige oppgave.

Estimer først en regresjonsmodell med vold som utfallsvariabel og andel unge kvinner som forklaringsvariabel. Estimer så en regresjonsmodell der du har både andel unge kvinner og andel unge menn som forklaringsvariable.

Vi er her kun interessert i regresjonskoeffisienten for unge kvinner, og den er vesentlig forskjellig i de to regresjonsmodellene. Si hva hvert av disse to estimatene betyr, og gi en forklaring på hvorfor estimatene er så forskjellige.

Skriv ditt svar her...

Maks poeng: 3

i **Digitale håndtegninger**

- Du kan svare med digital håndregning på oppgave 4.1, 4.2 og 4.5.
- Du kan kun bruke skisseark du får utdelt.
- Du kan bruke flere ark per oppgave.
- Se instruksjon for utfylling av skisseark, lenke nederst på skjermen (åpnes i eget vindu).
- Du får IKKE ekstra tid til å fylle ut informasjonsboksene på skisseark (engangskoder, kand.nr. o.l.).
- Du skal IKKE bruke digital håndregning på andre oppgaver enn oppgave 4.1, 4.2 og 4.5.

4.1 V19 t-test og frokostblanding

I denne oppgaven kan du svare med digital håndtegning. Bruk eget skisseark (utdelt). Se instruksjon for utfylling av skisseark.

En pakke av din favorittfrokostblanding skal inneholde 500 gram, men du er bekymret for at pakkene inneholder mindre enn det som er oppgitt. At det er litt variasjon må man godta, og siden du spiser ganske mye frokostblanding vil tilfeldig variasjon jevne seg ut over tid. Hvis du får systematisk for lite er det derimot et problem.

Du kjøper derfor 7 pakker frokostblanding og veier innholdet meget nøye og får følgende resultat oppgitt i gram:

502, 498, 479, 492, 488, 494, 494

a) Hva er standardavviket for disse dataene? Legg ved utregning på eget ark.

Skriv svaret her:

b) Gitt problemstillingen, hvilken av følgende er korrekt hypoteser for en hypotesetest?

4.2 V19 Regresjon og alkoholkonsum

I denne oppgaven kan du svare med digital håndtegning. Bruk eget skisseark (utdelt). Se instruksjon for utfylling av skisseark.

I en britisk studie av alkoholkonsum (Britton et al 2015, BMC Medicine, 13:47) ble det estimert hvordan alkoholkonsum varierte med alder og kjønn ($n = 1485$). Variabelen for kjønn er en dummy-variabel med verdiene 0 = mann og 1 = kvinne. Utfallsvariabelen er antall enheter alkohol i uka. En regresjonsmodell beskriver sammenhengen på denne formen:

$$y = \alpha + \beta_1 \times \text{alder} + \beta_2 \times \text{kjønn}$$

der parametrene er estimert til følgende: $\alpha = 15$, $\beta_1 = -0.187$ og $\beta_2 = -11.5$

a) Hva er forventet antall enheten alkohol i uka for 40 år gamle menn? Oppgi svaret med to desimaler her:

b) Standardfeilen til β_1 er estimert til 0.051. Hva er et 99 prosents konfidensintervall for dette estimatet? Bruk tre desimaler i utregning og svaret. Oppgi nedre grense for konfidensintervallet her: og øvre grense for

konfidensintervallet her:

Maks poeng: 3

4.3 V19 Standardavvik og standardfeil

Forklar med egne ord hva som er forskjellen på standardavvik og standardfeil.

Skriv ditt svar her...

Maks poeng: 2

4.4 V19 kjikvadrat

Vi kan teste om det er en sammenheng mellom to kategoriske variable ved bruk av kji-kvadrat test. Forklart kort med egne ord hva som er prinsippet for denne testen.

Du trenger ikke skrive noen formler eller utlede om selve kji-kvadratfordelingen.

Skriv ditt svar her...

Maks poeng: 3

4.5 V19 Studenter og snus

I denne oppgaven kan du svare med digital håndtegning. Bruk eget skisseark (utdelt). Se instruksjon for utfylling av skisseark.

I en undersøkelse blant et tilfeldig utvalg studenter ble det spurt om de brukte snus. Antall spurte var 250 og av disse var det 29.2 prosent som svarte at de brukte snus.

a) Hva er standardfeilen til dette estimatet. Gjør utregning på eget ark og oppgi svaret som andel med tre desimaler her:

b) Regn ut et 95% konfidensintervall for andelen som bruker snus i populasjonen.

Nedre grense:

Øvre grense:

c) I en diskusjon mellom to studenter hevder Ola at det neppe er mer enn 20% av studentene som bruker snus, og utvalget i undersøkelsen er såpass lite at det høyere estimatet skyldes tilfeldigheter. Per er uenig. Hvem av dem tror du har rett?

Gi en begrunnelse for svaret ditt sammen med utregning på eget ark, men kryss av konklusjonen her:

(For å få poeng på denne deloppgaven må det legges ved en begrunnelse).

Velg et alternativ

Per

Ola

Maks poeng: 4