

2. gangs prøve i R for STV 1020 V18, fredag 25. mai 12.15-14.00

Kodebok

Datasettet «beer» består av 336 observasjoner og 10 variabler. Observasjonene er av amerikanske stater i tidsperioden 1982-1988. Det er ikke missing-data.

Variabler:

state	state ID code
year	year
mrall	traffic fatality rate (deaths per 10000)
beertax	tax on case of beer
mlda	minimum legal drinking age
jaild	mandatory jail sentence ?
comserd	mandatory community service ?
vmiles	average miles per driver
unrate	unemployment rate
perinc	per capita personal income

Instruksjoner:

- Prøven skal besvares med et fungerende R-script (.R-fil, ikke sett noen punktum i filnavnet når du lagrer!) som leveres i innleveringsmappen «2.gangsprøve» på Fronter. Innleveringsmappen finner dere i arkiv-mappen i Fronter-rommet for seminargruppen.

- Scriptet skal inneholde nødvendig kode for å besvare oppgavene samt kommentarer markert med # som forklarer fremgangsmåten dere har valgt. Der oppgavene ber dere oppgi bestemte verdier eller tolkninger skal disse også oppgis som korte kommentarer i scriptet (du trenger ikke skrive mer enn en linje eller to).

- Sørg for at koden er oversiktlig. For å skille oppgavene fra hverandre i scriptet, anbefales overskrifter av typen:

```
### Oppgave 1 ####  
# Oppgave 1:
```

- Flere av oppgavene kan løses på forskjellige måter, du står fritt til å velge fremgangsmåte selv. Det er lov å google og bruke alle hjelpemidler (som oversikten over funksjoner og feilsøkningsdokumentet på github). Det eneste som ikke er lov, er kommunikasjon med medstudenter. Dersom dere skriver kode som er riktig, men ikke klarer å løse en oppgave fullstendig kan dere likevel få god uttelling. Dere kan hoppe over oppgaver dersom dere står fast, oppgavene er ikke sortert etter vanskelighetsgrad (heller ikke siste oppgave).

Lykke til!

Oppgaver:

- 1) Importer datasettet beer.csv eller beer.Rdata fra data-mappen på github (<https://github.com/langoergen/stv1020R/tree/master/data>) som et objekt i R-Studio. Du skal bruke dette datasettet i alle oppgavene.
- 2) Hvilken klasse har variablene i datasettet? Vis hvordan du finner ut av dette med kode.
- 3) Lag et spredningsplot (scatter-plot) med skatt på øl på x-aksen, og dødsrate i trafikken per 10000. innbygger på y-aksen. Tegn deretter en lineær regresjonslinje oppå plottet.
- 4) Lag et nytt datasett basert på beer, bestående av variablene year, mrall, beertax, vmiles, unrate og perinc. Lag en korrelasjonsmatrise med utgangspunkt i det nye datasettet. Gjør deretter en signifikanstest av sammenhengen mellom beertax og mrall. Hva indikerer korrelasjonen mellom disse to variablene?
- 5) Opprett et nytt datasett med alle observasjoner fra år 1982 i det opprinnelige datasettet, og et datasett med alle observasjoner fra år 1988 i det opprinnelige datasettet. Hva er gjennomsnittlig skatt på øl og gjennomsnittlig dødsrate per 10000 innbygger i de to datasettene?
- 6) Lag en tabell mellom variablene for obligatorisk fengselsstraff og obligatorisk samfunnstjeneste. Opprett deretter en ny variabel, punishment, slik at de observasjonene som hverken har obligatorisk fengselsstraff eller obligatorisk samfunnstjeneste får verdien 0, mens andre observasjoner får verdien 1.
- 7) Kjør en lineær regresjon med mrall som avhengig variabel og beertax, vmiles, unrate og perinc som uavhengige variabler. Lagre modellen som et objekt. Indikerer modellen at skatt på øl reduserer dødsfall i trafikken? Tolk effekten substansielt.
- 8) Opprett en ny variabel i datasettet ditt, state_fac, ved å omkode variabelen state til en factor. Lag deretter et boxplot med state_fac på x-aksen og mrall på y-aksen. Vil du si det er store variasjoner i dødsrate mellom statene (du kan trykke på Zoom over plottet for å se tydelig)?
- 9) Kjør en lineær regresjon med mrall som avhengig variabel og beertax, vmiles, unrate og perinc og state_fac som uavhengige variabler. Lagre modellen som et objekt. Indikerer modellen at skatt på øl reduserer dødsfall i trafikken? Tolk effekten substansielt. (P.S.: ved å legge til variabelen state_fac spesifiserer vi en modell med det som kalles fixed effects).