

Oppgaver seminar 5 – STV 4020A R H19

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:

Flere av oppgavene kan løses på forskjellige måter, du står fritt til å velge fremgangsmåte selv. Det er lov å google og bruke alle Dere kan hoppe over oppgaver dersom dere står fast, oppgavene er ikke sortert etter vanskelighetsgrad (heller ikke siste oppgave).

Oppgaver:

- 1) Importer datasettet beer.csv eller beer.Rdata fra data-mappen på github (<https://github.com/langoergen/stv4020aR/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.
Siden dette er et paneldatasett – sjekk om datasettet er balansert også.
- 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) 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.
- 7) 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)? Lag deretter det samme plottet som i oppgave 3, men legg til argumentet `facet_wrap(~state_fac)`. Hva leser du fra dette plottet?
- 8) 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).