

Instruksjoner

- Prøven skal besvares med et fungerende R-script som lastes opp i innleveringsmappen på Fronter.
- 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 kommentarer i scriptet.
- Sørg for at koden er oversiktlig. For å skille oppgavene fra hverandre, anbefales overskrifter av typen:
`### Oppgave 1 #####` i scriptet
`# Oppgave 1:`
eller lignende.
- Lykke til!

Variabelforklaringer:

rep_id Unik id for hver Stortingsrepresentant

sesjon Indikator for Stortingssesjon

parti_id Representantens partitilhørighet

ant_taler Antall taler representanten holdt i sesjonen

gj_ant_ord Gjennomsnittlig antall ord i talene til representanten

alder Representantens alder

kjonn Representantens kjønn

talers_rolle Indikator for om taler var medlem av Stortinget eller regjeringen (minister/-statsminister)

liste_nummer_valg Representantens plassering på partiets valglister ved forrige valg

Oppgaver

1. Last inn datasettet **talk-of-norway**. Enhetene i datasettet Stortingsrepresentanter (per sesjon). Du kan enten laste inn data direkte fra linken:
`http://folk.uio.no/martigso/jippii/talk-of-norway.csv`
eller laste ned fra samme link og inn i R fra lokalt på din data.
2. Lag et box-plot med talers rolle (**talers_rolle**) på x-aksen og antall taler (**ant_taler**) på y-aksen. Kommenter kort hva plotet viser.
3. Vis hvordan du finner median- og gjennomsnittalder (**alder**). Lag en ny variabel i datasettet som sentrerer alder til den mest egnede av disse (gjennomsnitt eller median), og kontroller at variabelen ble riktig opprettet.
4. Lag et spredningsplot med listeplassering fra forrige valg (**liste_nummer_valg**) på x-aksen og antall taler (**ant_taler**) på y-aksen. Legg også til en bivariat regresjonslinje i plotet, og kommenter sammenhengen.
5. Estimer en OLS-modell med antall taler (**ant_taler**) som avhengig variabel og gjennomsnitt antall ord (**gj_ant_ord**), listeplassering ved forrige valg (**liste_nummer_valg**), kjønn (**kjonn**) og alder (sentrert) som uavhengige variabler. Tolk effekten av en valgfri uavhengig variabel substansielt.
6. Estimer samme OLS-modell som over, men nå også med talers rolle (**talers_rolle**) som uavhengig variabel. Tolk konstantleddet substansielt.
7. Lag to nye variabel i datasettet som viser a) restleddene (residualene) og b) forventet (predikert) verdi til hver enhet fra den siste modellen (oppgave 6).
8. Lag et spredningsplot med forventede verdier på x-aksen og restledd på y-aksen (begge fra oppgave 7), og tegn en bivariat regresjonslinje mellom disse. Er forutsetningen om ikke-korrelerede restledd og forventede verdier oppfylt?