

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, skal disse også oppgis som kommentarer i scriptet.
- Sørg for at koden er oversiktlig. For å skille oppgavene fra hverandre, anbefaler vi overskrifter av typen: `### Oppgave 1 #####` i scriptet.
- Lykke til!

Oppgaver

1. Last inn datasettet **ministers.rdata**. Enhetene i datasettet er ministre.
2. Finn gjennomsnitt, median, skjevhet og kurtose for variabelen **duration**. Oppgi skjevheten i en kommentar.
3. Opprett variabelen **edu2** hvor alle med “Higher education” på **education** får verdien 1, og alle andre får verdien 0.
4. Gi verdien NA på **duration** til alle som har verdien “Lyng I” på variabelen **cabinet_name**.
5. Produser en korrelasjonsmatrise av variablene **duration**, **resigcalls** og **edu2**. Oppgi korrelasjonen mellom **duration** og **resigcalls**.
6. Plot et spredningsdiagram med **age_cen** langs X-aksen, og **duration** langs Y-aksen. Tegn en regresjonslinje i figuren med **duration** som avhengig variabel og **age_cen** som uavhengig variabel.
7. Estimer en OLS analyse med **duration** som avhengig variabel, og **resigcalls**, **edu2** og **age_cen** som uavhengige variabler.
8. Bruk analysen fra forrige oppgave og opprett variabelen **pred_dur** som inneholder den predikerte verdien til enhetene i datasettet.
9. Tegn et histogram av **pred_dur**
10. Gi verdien NA til enheten med høyest verdi på **pred_dur** og tegn histogrammet på nytt.

Variabelforklaringer:

duration Antall dager ministeren satt i regjering

resigcalls Antall ganger media har forlangt at ministeren går av

age_cen Alder i år, sentrert

education Utdannelse