Denne prøven skal leveres som et fungerende R-script. Scriptet skal inneholde nødvendig kode for å besvare oppgavene samt kommentarer markert med #. Der oppgavene ber dere oppgi bestemte verdier eller tolkninger skal disse oppgis som kommentarer i scriptet. For hvert spørsmål får du et poeng for fungerende R-kode og et poeng for riktig svar. Du må få 12 poeng for å bestå prøven. Besvarelsen leveres til seminarlederen din i Canvas innen kl. 15:00.

[INFO OM LASTE NED DATASETT]

Datasettet inneholder følgende variabler:

* country\_name: land
* prosent\_vaksinerte: prosentandel vaksinerte mot covid-19 i befolkningen.
* bnp\_per\_cap: BNP per innbygger målt i 1000 dollar
* regime: landets styresett. Variabelen tar følgende verdier:
  + 0 for land kategorisert som lukkede autokratier
  + 1 for land kategorisert som elektorale autokratier
  + 2 for land kategorisert som elektorale demokratier
  + 3 for land kategorisert som liberale demokratier

Oppgaver

1. Last inn datasettet vaksinedata.csv og se at det kjører. Rapporter størrelsen på datasettet (antall variabler og observasjoner).
2. Finn klassen til alle variablene i datasettet. Oppgi klassene i en kommentar.
3. Finn gjennomsnitt, median og standardavvik for variabelen prosent\_vaksinerte. Oppgi verdiene i en kommentar.
4. Hvilken av variablene prosent\_vaksinerte og bnp\_per\_cap har flest missingverdier? Vis hvordan du kom frem til svaret med kode og oppgi svaret i en kommentar.
5. Lag et spredningsplot med variabelen prosent\_vaksinerte på y-aksen og variabel bnp\_per\_cap på x-aksen. Gi plottet en tittel. Hva viser figuren?
6. Kjør en lineær regresjon med prosent\_vaksinerte som avhengig variabel og bnp\_per\_cap som uavhengig variabel. Dette er din modell1. Kommenter kort hva koeffisienten og p-verdien til prosent\_vaksinerte forteller oss.
7. Lag en ny variabel – regime\_2 – som tar verdien «autokrati» når observasjonen har en verdi som er lavere enn 2 på variabelen regime, og verdien «demokrati» ellers.

Hvordan kontrollerer du at omkodingen ble riktig?

1. Kjør en lineær regresjon med prosent\_vaksinerte som avhengig variabel. Som uavhengig variabel bruker du bnp\_per\_cap og regime\_2. Dette er din modell2. Kommenter kort hva koeffisienten og p-verdien til regime\_2 forteller oss.
2. Presenter resultatene fra modellene dine i en tabell. Rapporter hva som skjer med forholdet mellom prosent\_vaksinerte og bnp\_per\_innbygger når du inkluderer regime2 som en kontrollvariabel.
3. Plott regresjonslinjen til bnp\_per\_innbygger fra modell 2 med 95 prosent konfidensintervaller. Hva viser plottet?

Alternativt spm. 10: Plot et stolpediagram for variabelen regime. Hvilken verdi på variabelen har lavest frekvens?

Fjorårets prøve til referanse:

**Instruksjoner**

* Prøven skal besvares med et fungerende R-script som lastes opp i innleveringsmappen på Canvas.
* 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.
* Husk at riktig kode er det viktigste; pass på å ikke bruk for lang tid på tolkninger.
* 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!

Datasettet dere skal jobbe med er et utvalg av variablene fra runde 8 av European Social Survey for Norge. Respondentene er individer. Datasettet inneholder missingverdier.

Variabelforklaringer:

* trust\_politicians: Tilitt til politikere (0-10 der 0 = Lav, 10 = Høy)
* vote: Hvorvidt respondenten stemte i forrige stortingvalg (1 = stemte, 2 = stemte ikke)
* satisified\_democracy: Tilfredshet med demokratiet (0-10 der 0 = Lav, 10 = Høy)
* gender: Kjønn (1 = Mann, 2 = Dame)
* age: Alder (numerisk)
* freedom: Hvorvidt respondenten mener det er "viktig å ta egne valg" (1-6, der 1 = Veldig viktig, 6 = Ikke veldig viktig)

**Oppgaver**

1. Last inn datasettet ESS. Du kan laste ned Rdata fra [https://github.com/egen97/4020A\_RSeminar/blob/master/ESS\_Data.Rdata (Lenker til en ekstern side.)](https://github.com/egen97/4020A_RSeminar/blob/master/ESS_Data.Rdata) eller lese inn csv-data direkte fra [https://raw.githubusercontent.com/egen97/4020A\_RSeminar/master/ESS\_Data.csv (Lenker til en ekstern side.)](https://raw.githubusercontent.com/egen97/4020A_RSeminar/master/ESS_Data.csv).
2. Hvilken av variablene vote og trust\_politicians har flest missingverdier?
3. Finn følgende for variabelen age:

* Median
* Gjennomsnitt
* Standardavvik
* Laveste verdi
* Høyeste verdi

Oppgi verdiene i en kommentar.

1. Plot et stolpediagram for variabelen satisified\_democracy. Kommenter hvilken verdi på variabelen som har høyest frekvens.
2. Lag en ny variabel – satisified\_democracy\_2 – som tar verdiene:

* 1 når satisified\_democracy har verdien 5 eller høyere
* 0 når satisified\_democracy har verdien 4 eller lavere

Sjekk at det ble riktig med en tabell.

1. Lag et spredningsplot (scatter plot) med variabelen trust\_politicians på y-aksen og age på x-aksen.
2. Lag en ny variabel gender\_2 som tar verdiene:

* «Mann» når gender er lik 1
* «Kvinne» når gender er lik 2

1. Estimer en OLS-modell med satisified\_democracy som avhengig variabel og gender\_2, age og trust\_politicians som uavhengige variabler. Kommenter kort hva til koeffisienten til gender\_2 forteller oss.
2. Vis resultatene fra modellen i oppgave 8 i en tabell.
3. Plott fordelingen til variabelen freedom.