**R-prøve 2020:**

Prøven gjennomføres med datasettet «Wages» som du kan laste ned her: <https://github.com/liserodland/STV1020/blob/main/data/wages.Rdata> .   
Datasettet tar for seg temaer rundt lønn og utdanning.  
Datasettet inneholder 7425 observasjoner på 5 variabler. Enhetene er personer.    
Datasettet inneholder noen missing-verdier(NAs).

|  |  |
| --- | --- |
| **KODEBOK** |  |
| **Variabel navn** | **Forklaring** |
| Wages | Inntekt i pund(£), målt i pr 1000. |
| Education | År med gjennomført utdanning |
| Age | Alder |
| Sex | Kjønn |
| Language | Respondentens språk, tre kategorier: Engelsk, Fransk, Annet. |

1.

*- Last inn datasettet: «wages.Rdata» som ligger på Canvas.*

2.

*- Se på variabelen for inntekt og utdanning. Hent ut gjennomsnittet, minimumsverdi, maxverdi og standardavvik hos begge variablene. Hvor mange missing-verdier er det hos de to variablene?*

*3.*

*- Lag og presenter et histogram for inntektsvariabelen. Hva viser det?*

*4.*

*- Plott inntekt og utdanning mot hverandre i et spredningsplott/scatterplot. Gi en kort kommentar til plottet.*

*5.*

*- Opprett en krysstabell mellom kjønn og utdanning.*

*6.*

*- Opprett en numerisk, dikotom variabel for kjønn. Du kan selv velge om kvinne skal få verdien 0 eller 1, mens mann får den motsatte verdien ny.*

*- Opprett en dikotom utdannings variabel som har verdien 1 om originalvariabel har større verdier enn sitt gjennomsnitt og 0 om originalvariabelen har mindre verdier enn sitt gjennomsnitt.*

*- Kontroller om de to omkodingene har gått i orden med en krysstabell.*

*7.*

*- Lag et nytt datasett som inneholder variablene inntekt, utdanning, kjønn, og de to nye variablene du opprettet.*

*8.*

*- Estimer en enkel bivariat regresjonsmodell med inntekt som avhengig variabel, og utdanning som uavhengig variabel, tolk resultatene.*

*9.*

*- Legg til kjønn som en kontrollvariabel i denne modellen. Estimer modellen og tolk resultatene.*