

# R-seminar 6: Løsningsforslag

STV1020 Vår 2021

Uke 17

Datasettet du skal bruke er versjon 9 av spørreundersøkelsen European Social Survey (ESS) med tall fra Norge i 2018. Datasettet ligger på GitHub. Du skal bruke variablene "agea", "eduyrs", "gndr" og "imueclt".

- agea = Age of respondent
- eduyrs = Years of full-time education completed. About how many years of education have you completed, whether full-time or part-time? Please report these in full-time equivalents and include compulsory years of schooling.
- gndr = Gender, Male = 1, Female = 2
- imueclt = Country's cultural life undermined or enriched by immigrants, 0 = Cultural life undermined, 10 = Cultural life enriched

1. Last inn datasettet ESS9NO.dta fra GitHub. Lagre det i din prosjektmappe eller ditt working directory.

```
## Setter working directory
setwd("")

## Error in setwd(""): cannot change working directory

## Laster inn pakker
library(tidyverse)
library(stargazer)
library(haven)

## Laster inn datasett
df <- read_dta("ESS9NO.dta")
```

2. Her skal du lage et datasett det er oversiktelig å jobbe med. Du kan gjøre punkt a og b i samme kodesekvens ved å bruke %>% mellom hver funksjon, men du kan også løse hver oppgave for seg.
  - a. Opprett et nytt datasett der du kun inkluderer variablene "agea", "eduyrs", "gndr" og "imueclt".

- b. Omkod variabelen "gndr" (kjønn) slik at mann = 0 og kvinne = 1.

```
## Lager nytt datasett med relevante variabler, endrer navn og omkoder gndr
df2 <- df %>%
  select(agea, eduyrs, gndr, imueclt) %>%
  mutate(gndr_new = ifelse(gndr == 1, 0, 1))
```

3. Fjern enheter med missing i det nye datasettet ditt. Hvor mange enheter/observasjoner har datasettet nå?

```
## Fjerner NAs
df2 <- df2 %>%
  drop_na()

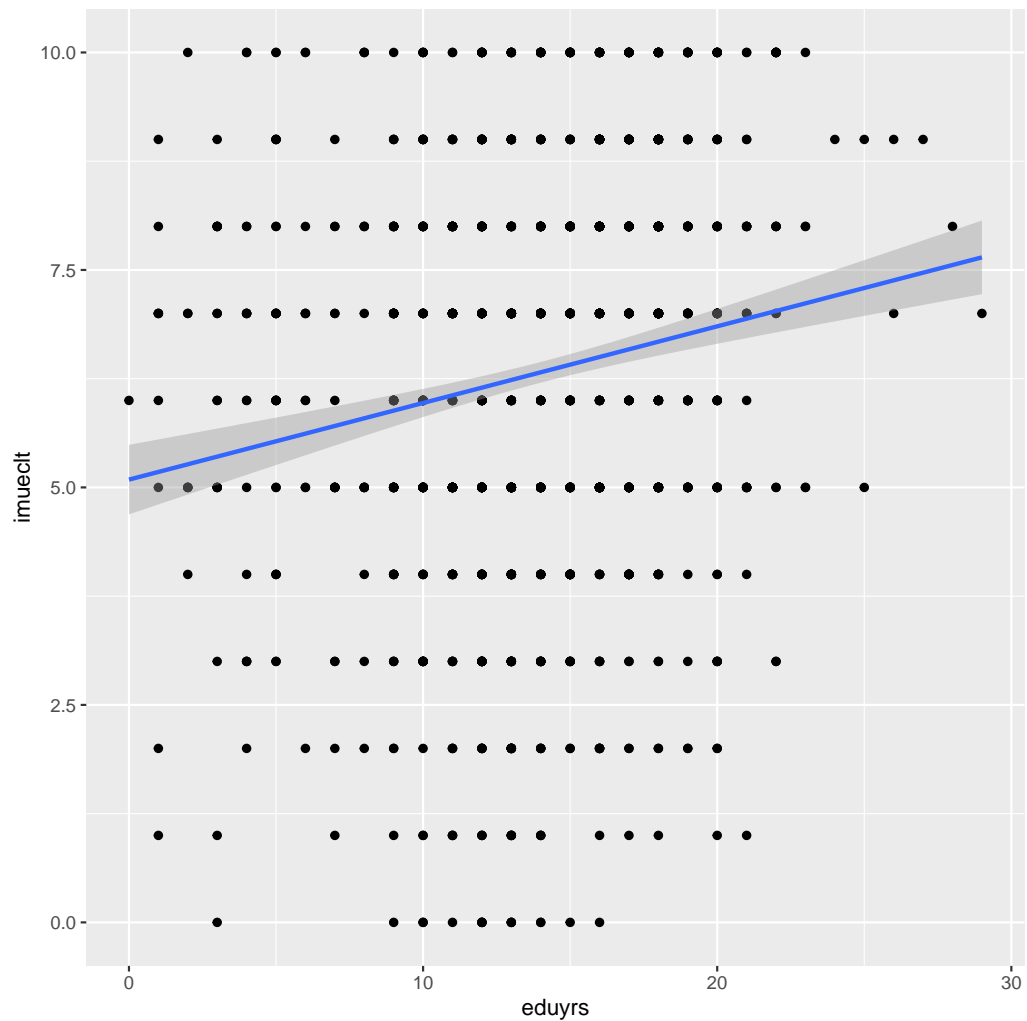
dim(df2)

## [1] 1354    5

# Har nå 1354 observasjoner
```

4. Lag et spredningsdiagram mellom "imueclt" (AV) og "eduyrs" (UV). Plott regresjonslingen til dette plottet.

```
## Spredningsplott for AV og UV med regresjonslinje
ggplot(df2, aes(x = eduyrs, y = imueclt)) +
  geom_point() +
  geom_smooth(method = "lm")
```



5. Estimer en lineær OLS regresjonsmodell med "imueclt" som AV og "eduyrs" som hoved-UV, og "agea" og "gndr" som kontrollvariabler.

```
## Kjører multippel regresjon
mod1 <- lm(imueclt ~ eduyrs +
            agea +
            gndr_new,
            data = df2)

summary(mod1)

##
## Call:
## lm(formula = imueclt ~ eduyrs + agea + gndr_new, data = df2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.6096 -1.4042  0.1723  1.6385  4.9827
##
## Coefficients:
```

```
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) 4.533042    0.264210  17.157 < 2e-16 ***
## eduyrs      0.082456    0.013710   6.014 2.32e-09 ***
## agea        0.006433    0.003310   1.943  0.0522 .
## gndr_new    0.739678    0.120620   6.132 1.14e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.199 on 1350 degrees of freedom
## Multiple R-squared:  0.05754, Adjusted R-squared:  0.05545
## F-statistic: 27.47 on 3 and 1350 DF,  p-value: < 2.2e-16
```

6. Print resultatene i en fin tabell ved hjelp av stargazer()-funksjonen.

```
## Presenterer resultatene i stargazer
stargazer(mod1,
           type = "text")

##
## =====
##                               Dependent variable:
##                               -----
##                               imueclt
## -----
## eduyrs                        0.082***
##                               (0.014)
##
## agea                          0.006*
##                               (0.003)
##
## gndr_new                      0.740***
##                               (0.121)
##
## Constant                     4.533***
##                               (0.264)
##
## -----
## Observations                  1,354
## R2                            0.058
## Adjusted R2                   0.055
## Residual Std. Error          2.199 (df = 1350)
## F Statistic                   27.475*** (df = 3; 1350)
## =====
## Note:                         *p<0.1; **p<0.05; ***p<0.01
```

7. Kommenter kort koeffisientene til modellen.

```
## Sammenhengen mellom utdanning og om lands kulturelle liv undergraves
# eller berikes av innvandrere er positiv og statistisk signifikant.
# En skalaenhets økning i utdanning (et år) medfører 0.082 skalaenheters
# økning i syn på hvorvidt lands kulturelle liv berikes av innvandrere eller ei,
```

# når de andre variablene holdes konstant. Kort sagt: Mer utdanning = man mener  
# at landets kulturelle liv berikes av innvandrere.