

```

# Laster nødvendige pakker
library(tidyverse)      # Generell pakke som inneholder mange ulike funksjoner
library(haven)           # lese .dta-filer
library(labelled)        # variabel- og verdilabels
library(summarytools)    # tabellfunksjoner
library(jtools)          # regresjonstabeller
library(sjPlot)          # plotting av samspillsmodeller
library(margins)         # margins() tilsvarende Stata

# Sette arbeidskatalog -----
# Mac-eksempel:
# setwd("/Volumes/arieble/Undervisning/SOS3003V2026/DATA")
# Windows-eksempel:
setwd("M:/Undervisning/SOS3003V2026/DATA")

# Import av data -----
ess <- read_dta("ESS11e04_1-subset.dta")

# Deskriptiv statistikk -----
summary(ess$freehms)
descr(ess$freehms)
table(ess$freehms)
freq(ess$freehms)
hist(ess$agea, breaks = 30)
# Bostedvariabel og fjerning av missing -----
freq(ess$domicil)
ess$domicil2 <- zap_missing(ess$domicil)
freq(ess$domicil2)

# Omkoding av kjønn til dummy -----
freq(ess$gndr)

ess$kvinne <- car::recode(ess$gndr, "1=0; 2=1; else=NA")
# Av og til vil du få en varsel om at as.numeric() burde brukes først. Slik:
ess$kvinne <- car::recode(as.numeric(ess$gndr), "1=0; 2=1; else=NA")
freq(ess$kvinne)

# Lage dikotom variabel fra kategorier -----
freq(ess$prtvtcno)

ess$frp <- car::recode(as.numeric(ess$prtvtcno), "1=0; 2=0; 3=0; 4=0; 5=0;
                           6=0; 7=0; 8=1; 9=0; 10=0; 11=0; else=NA")

freq(ess$frp)
ctable(ess$polintr, ess$frp, prop="c", useNA="no")

# Lineær regresjon -----
freq(ess$trstprl)

model1 <- lm(trstprl ~ eduyrs, data = ess)
summary(model1)
summ(model1)

model2 <- lm(trstprl ~ eduyrs + kvinne + agea, data = ess)
summ(model2)

# Dummysett: referanskategori (dplyr-faktor) -----
freq(ess$domicil2)
model3 <- lm(trstprl ~ eduyrs + kvinne + agea + domicil2, data = ess)
summ(model3)

ess$bosted <- as.factor(ess$domicil2)
model4 <- lm(trstprl ~ eduyrs + kvinne + agea + bosted, data = ess)
summ(model4)

```

```

# Stata: ib5.bosted → referansekategori = 5
ess$bosted_5 <- relevel(ess$bosted, ref = "5")
model5 <- lm(trstprl ~ eduyrs + kvinne + agea + bosted_5, data = ess)
summ(model5)

# Regresjon med andregradsledd -----
model6 <- lm(trstprl ~ eduyrs + kvinne + agea + I(agea^2) + bosted_5, data = ess)
summ(model6, digits=3)

# Margins-plot: kurvelinearitet -----
effect_plot(model6, pred = agea)

# Ny avhengig variabel og snuing -----
freq(ess$gincdif)

ess$redusere <- car::recode(as.numeric(ess$gincdif), "1=5; 2=4; 3=3; 4=2;
                                5=1; else=NA")
freq(ess$redusere)

# Samspillsgraf (grafisk fremstilling av samspill) -----
model7 <- lm(gincdif ~ kvinne + hinctnta, data = ess)
summ(model7, digits=3)
model8 <- lm(redusere ~ kvinne + hinctnta, data = ess)
summ(model8, digits=3)

model9 <- lm(redusere ~ kvinne + hinctnta + kvinne:hinctnta, data = ess)
summ(model9, digits=3)

# Alternativt:
model9 <- lm(redusere ~ kvinne * hinctnta, data = ess)
summ(model9, digits=3)

plot_model(model9, type = "int")

# Variabel for logistisk regresjon -----
ess <- ess %>% mutate(
  innvandringsvennlig = ifelse(imwbcnt == 88, NA, imwbcnt))

# Bivariat logistisk regresjon -----
log1 <- glm(frp ~ eduyrs, data = ess, family = binomial)
summ(log1, digit=3)

exp(coef(log1)) # OR tilsvarer ,or i Stata

# Margins for sannsynlighet
effect_plot(log1, pred = eduyrs)

# Multippel logistisk regresjon -----
log2 <- glm(frp ~ eduyrs + kvinne + agea, data = ess, family = binomial)
summ(log2, digit=3)

exp(coef(log2)) # OR tilsvarer ,or i Stata

# AME - average marginal effects:
# gjennomsnittlige endringen i sannsynligheten for å stemme FrP når
# en uavhengig variabel øker med én enhet (eller skifter kategori),
# mens alle andre variabler holdes konstante.
summary(margins(log2))

# Hvis du ønsker å fjerne alt som ligger i minnet til R -----
# rm(list =ls())      # Fjerner alt fra minnet til R
# Gå til start og last inn R-skriptet på nytt ved behov

```