Gruppeoppgave

Kandidatnummer??

Høst 2025

Laste inn data

```
library(haven) # Lese inn data sett
library(dplyr) # Databehandling
library(fixest) # Regresjon
library(knitr)
library(tidytext)
library(GenericML)
library(mlr3)
library(mlr3tuning)


df <- read_dta("data/Combined_allwaves_final.dta") # Leser data

df <- df |>
  mutate(miss_age_PAP = ifelse(is.na(age_med_BL), 1, 0)) |>
  mutate(age_med_BL_control= ifelse(is.na(age_med_BL), 0, 1))|>
  mutate(miss_household_size= ifelse(is.na(household_size), 1, 0))|>
  mutate(household_size_control= ifelse(is.na(household_size), 0, household_size)) |>
  mutate(miss_edu_category= ifelse(is.na(edu_nohs_BL), 1, 0))|>
  mutate(edu_nohs_BL_control= ifelse(is.na(edu_nohs_BL), 0, edu_nohs_BL))|>
  mutate(married_control= ifelse(is.na(married), 0, married))|>
  mutate(divorced_separated_control= ifelse(is.na(divorced_separated), 0,
  divorced_separated))|>
  mutate(single_control= ifelse(is.na(single), 0, single))|>
  mutate(widowed_control= ifelse(is.na(widowed), 0, widowed))|>
  mutate(miss_relationship= ifelse(is.na(rel_status_BL), 1, 0))|>
  mutate(miss_cars= ifelse(is.na(cars), 1, 0))|>
  mutate(one_car_control= ifelse(is.na(one_car), 0, one_car))|>
  mutate(mult_cars_control= ifelse(is.na(mult_cars), 0, mult_cars))|>
  mutate(miss_LF_BL= ifelse(is.na(LF_BL), 1, 0))|>
  mutate(LF_BL_control= ifelse(is.na(LF_BL), 0, LF_BL)) |>
  filter(endline_start_w3==1)
```

```
#kable("x") # Jeg må ha med denne for at ting skal fungere.
```

```
s_train <- feols(s_train_bi_w3~ treatment + age_med_BL_control + miss_age_PAP + edu_nohs_BL_control
      + miss_edu_category + married_control + single_control + widowed_control
      + miss_relationship + household_size_control + miss_household_size + one_car_control
      + miss_cars + LF_BL_control + miss_LF_BL |randomization_cohort2, cluster = c("file_nbr"), data = d
```

```
licence <- feols(license_w3~ treatment + age_med_BL_control + miss_age_PAP + edu_nohs_BL_control
      + miss_edu_category + married_control + single_control + widowed_control
      + miss_relationship + household_size_control + miss_household_size + one_car_control
      + miss_cars + LF_BL_control + miss_LF_BL |randomization_cohort2, cluster = c("file_nbr"), data = d



empl <- feols(employed_w3~ treatment +age_med_BL_control + miss_age_PAP + edu_nohs_BL_control
      + miss_edu_category + married_control + single_control + widowed_control
      + miss_relationship + household_size_control + miss_household_size + one_car_control
      + miss_cars + LF_BL_control + miss_LF_BL |randomization_cohort2, cluster = c("file_nbr"), data = d


not_empl <- feols(unemployed_w3~  treatment + age_med_BL_control + miss_age_PAP + edu_nohs_BL_control
      + miss_edu_category + married_control + single_control + widowed_control
      + miss_relationship + household_size_control + miss_household_size + one_car_control
      + miss_cars + LF_BL_control + miss_LF_BL |randomization_cohort2, cluster = c("file_nbr"), data = d

leave_house <- feols(G1_2_abovemed ~ treatment + age_med_BL_control + miss_age_PAP + edu_nohs_BL_control
      + miss_edu_category + married_control + single_control + widowed_control
      + miss_relationship + household_size_control + miss_household_size + one_car_control
      + miss_cars + LF_BL_control + miss_LF_BL |randomization_cohort2, cluster = c("file_nbr"), data = d


make_purchase <- feols(G1_3_abovemed ~ treatment + age_med_BL_control +
                         miss_age_PAP + edu_nohs_BL_control
      + miss_edu_category + married_control + single_control + widowed_control
      + miss_relationship + household_size_control + miss_household_size + one_car_control
      + miss_cars + LF_BL_control + miss_LF_BL |randomization_cohort2, cluster = c("file_nbr"), data = d
```

```
controls <- c("miss_age_PAP","age_med_BL_control", "miss_edu_category", "married_control",
              "widowed_control","miss_relationship", "household_size_control", "miss_household_size",
              "one_car_control", "mult_cars_control", "miss_cars","LF_BL_control", "LF_BL_control",
              "edu_nohs_BL_control", "single_control", "miss_LF_BL", "Constant", "randomization_cohort2


tit <- "Treatment Effects on Individual Outcomes and Intrahousehold Responses"



fitstat_register("control_mean", function(x) mean(x), "Control mean")

fitstat_register("mean", function(x) mean(x, na.rm = T), "control_m")

fitstat_register("pval", function(x) pvalue(x), "p-value b = 0")

fitstat_register("mean_c",

                 function(x){
                   name <- x$fml[2] |> as.character()
                   xer <- df |>
                   filter(treatment == 0) |>
```

```
                    select(name) |> pull()
                    mean(xer,na.rm = T)
                    },
                  "Control mean"
)


fitstat_register("me",

                  function(x){
                    name <- x$fml[2] |> as.character()
                    xer <- df |>
                    filter(treatment == 0) |>
                    select(name) |> pull()

                    x$coefficients[1]/ mean(xer,na.rm = T)
                    },
                  "b/Control mean"
)
```

```
t1pa <- etable(s_train,licence,
       se.below=T,
       drop = controls,
       title =tit,
       digits = "r3",digits.stats = "r3",
       tex = T,
       signif.code = NA,
       dict=c(treatment = "Treatment",
              s_train_bi_w3 = "Started driver`s training",
              license_w3 = "Received license",
              randomization_cohort2 = ""),
       style.tex = style.tex("qje", ),
       fitstat = ~ n + mean_c +me + pval)

t1pb <- etable(empl,not_empl ,
       se.below=T,
       drop = controls,
       title =tit,
       digits = "r3",digits.stats = "r3",
       tex = T,
       signif.code = NA,
       dict=c(treatment = "Treatment",
              employed_w3 = "Employed",
              unemployed_w3 = "Unemployed",
              randomization_cohort2 = ""),
       style.tex = style.tex("qje", ),
       fitstat = ~ n + mean_c +me + pval)

t1pc <- etable(leave_house, make_purchase,
       se.below=T,
       tex = T,
       drop = controls,
       title =tit,
```

```
       digits = "r3",digits.stats = "r3",
       signif.code = NA,
       dict=c(treatment = "Treatment",
              G1_2_abovemed = "Allowed to leave house w/o permission",
              G1_3_abovemed = "Allowed to make purchase w/o permission",
              randomization_cohort2 = ""),
       style.tex = style.tex("qje", ),
       fitstat = ~ n + mean_c +me + pval)



t1pa;t1pb;t1pc
```

Table 1: Treatment Effects on Individual Outcomes and Intrahousehold Responses

	Started driver's training	Received license
	(1)	(2)
Treatment	0.619	0.430
	(0.040)	(0.039)
Observations	467	467
Control mean	0.192	0.102
b/Control mean	3.229	4.221
p-value b = 0	0.000	0.000
fixed effects	✓	✓

Table 2: Treatment Effects on Individual Outcomes and Intrahousehold Responses

	Employed	Unemployed
	(1)	(2)
Treatment	0.085	-0.105
	(0.043)	(0.049)
Observations	488	488
Control mean	0.210	0.569
b/Control mean	0.405	-0.185
p-value b = 0	0.049	0.032
fixed effects	✓	✓

Table 3: Treatment Effects on Individual Outcomes and Intrahousehold Responses

	Allowed to leave house w/o permission (1)	Allowed to make purchase w/o permission (2)
Treatment	0.062	-0.090
	(0.046)	(0.048)
Observations	488	486
Control mean	0.344	0.484
b/Control mean	0.179	-0.186
p-value b = 0	0.184	0.059
fixed effects	✓	✓

Tabell 2

```
hte_var <- c("age_med_BL_control", "edu_nohs_BL_control","LF_BL_control" )
panelb_controls <- c("miss_age_PAP", "miss_edu_category", "married_control",
               "widowed_control","miss_relationship", "household_size_control", "miss_household_size",
               "one_car_control", "mult_cars_control", "miss_cars", "LF_BL_control", "single_control", "


fitstat_register("samlet",
        function(x){
          antall_koeff <- length(x$coefficients)
          b_1 <- x$coefficients[1]
          b_3 <- x$coefficients[antall_koeff]

          round(b_1 + b_3,3)
        },
        "B1 + B3"
        )


fitstat_register("se",

                 function(x){
          antall_koeff <- length(x$coefficients)

          v_b_1 <- x$se[1]**2
          v_b_3 <- x$se[antall_koeff]**2

          cov13 <- x$cov.unscaled[antall_koeff]

          se <- sqrt(v_b_1 + v_b_3 + 2*cov13)

          paste0("(",round(se,3),")")
        },
```

```
        " "
        )

fitstat_register("median_age",
                 function(x){

                   name <- formula(x)[2] |> as.character()

                   df |>
                     filter(treatment == 0 & age_med_BL == 0) |>
                     pull(name) |> mean(na.rm = T) |> round(3)

                 }, "Mean: control, below median age")


fitstat_register("hs",
                 function(x){

                   name <- formula(x)[2] |> as.character()

                   df |>
                     filter(treatment == 0 & edu_nohs_BL == 0) |>
                     pull(name) |> mean(na.rm = T) |> round(3)

                 },
                 "Mean: control, completed HS")
fitstat_register("mcd",
                 function(x){

                   name <- formula(x)[2] |> as.character()
                   df |> filter(treatment == 0 & divorced_separated == 1) |>
                     pull(name) |> mean(na.rm = T) |> round(3)


                 }, "Mean: control, divorced")

fitstat_register("mcm",
                 function(x){

                   name <- formula(x)[2] |> as.character()
                   df |> filter(treatment == 0 & married == 1) |>
                     pull(name) |> mean(na.rm = T) |> round(3)


                 }, "Mean: control, married")
fitstat_register("mcnm",
                 function(x){

                   name <- formula(x)[2] |> as.character()
                   df |> filter(treatment == 0 & single == 1) |>
                     pull(name) |> mean(na.rm = T) |> round(3)
```

```
                }, "Mean: control, never married")

fitstat_register("mcw",
                function(x){

                  name <- formula(x)[2] |> as.character()
                  df |> filter(treatment == 0 & widowed == 1) |>
                    pull(name) |> mean(na.rm = T) |> round(3)


                }, "Mean: control, widowed")


#### Panel A

rl2a <- feols(license_w3 ~ treatment + age_med_BL  + treatment:age_med_BL
      + miss_age_PAP
      + miss_edu_category + married_control + single_control + widowed_control
      + miss_relationship + household_size_control + miss_household_size + one_car_control
      + miss_cars  + miss_LF_BL |randomization_cohort2, cluster = c("file_nbr")
      ,data = df)


emp2a <- feols(employed_w3 ~ treatment +age_med_BL+ treatment:age_med_BL
      + miss_age_PAP
      + miss_edu_category + married_control + single_control + widowed_control
      + miss_relationship + household_size_control + miss_household_size + one_car_control
      + miss_cars  + miss_LF_BL |randomization_cohort2, cluster = c("file_nbr")
      ,data = df)


make_purchase2a <- feols(G1_3_abovemed ~ treatment +age_med_BL  + treatment:age_med_BL
      + miss_age_PAP
      + miss_edu_category + married_control + single_control + widowed_control
      + miss_relationship + household_size_control + miss_household_size + one_car_control
      + miss_cars  + miss_LF_BL |randomization_cohort2, cluster = c("file_nbr")
      ,data = df)




##### Panel b

rl2b <- feols(license_w3 ~ treatment + edu_nohs_BL  + treatment:edu_nohs_BL
      + miss_age_PAP
      + miss_edu_category + married_control + single_control + widowed_control
      + miss_relationship + household_size_control + miss_household_size + one_car_control
      + miss_cars  + miss_LF_BL |randomization_cohort2, cluster = c("file_nbr")
      ,data = df)


emp2b <- feols(employed_w3 ~ treatment + edu_nohs_BL + treatment:edu_nohs_BL
```

```
      + miss_age_PAP
      + miss_edu_category + married_control + single_control + widowed_control
      + miss_relationship + household_size_control + miss_household_size + one_car_control
      + miss_cars  + miss_LF_BL |randomization_cohort2, cluster = c("file_nbr")
      ,data = df)


make_purchase2b <- feols(G1_3_abovemed ~ treatment + edu_nohs_BL + treatment:edu_nohs_BL
      + miss_age_PAP
      + miss_edu_category + married_control + single_control + widowed_control
      + miss_relationship + household_size_control + miss_household_size + one_car_control
      + miss_cars  + miss_LF_BL |randomization_cohort2, cluster = c("file_nbr")
      ,data = df)




####### Panel c


rl2c <- feols(license_w3 ~ treatment + single_control + married_control+ widowed_control
              + treatment:married + treatment:single + treatment:widowed
              + miss_age_PAP + miss_edu_category + miss_relationship
              + household_size_control + miss_household_size + one_car_control
              + miss_cars  + miss_LF_BL
              |randomization_cohort2, cluster = c("file_nbr")
              ,data = df)

emp2c <- feols(employed_w3 ~ treatment + single_control + married_control+ widowed_control
              + treatment:married + treatment:single + treatment:widowed
              + miss_age_PAP + miss_edu_category + miss_relationship
              + household_size_control + miss_household_size + one_car_control
              + miss_cars  + miss_LF_BL
              |randomization_cohort2, cluster = c("file_nbr")
              ,data = df)

make_purchase2c <- feols(G1_3_abovemed~ treatment + single_control + married_control+ widowed_control
              + treatment:married + treatment:single + treatment:widowed
              + miss_age_PAP + miss_edu_category + miss_relationship
              + household_size_control + miss_household_size + one_car_control
              + miss_cars  + miss_LF_BL
              |randomization_cohort2, cluster = c("file_nbr")
              ,data = df)



############


panela <- etable(rl2a,emp2a,make_purchase2a, se.below = T,
       tex = T, title = "Panel A",
       drop = panelb_controls,
       digits = "r3",digits.stats = "r3",
```

```
       signif.code = NA,
       dict=c(treatment = "Treatment",
              license_w3 = "Received license",
              G1_3_abovemed = "Allowed to make purchase w/o permission",
              employed_w3 = "Employed",
              age_med_BL = "Above median age",
              randomization_cohort2 = ""),
       style.tex = style.tex("qje", ),
       fitstat = ~samlet+ se+ n +  median_age);

panelb <- etable(rl2b,emp2b,make_purchase2b, se.below = T,
       tex = T, title = "panel B",
       drop = panelb_controls,
       digits = "r3",digits.stats = "r3",
       signif.code = NA,
       dict=c(treatment = "Treatment",
              license_w3 = "Received license",
              G1_3_abovemed = "Allowed to make purchase w/o permission",
              employed_w3 = "Employed",
              edu_nohs_BL = "Less than HS",
              randomization_cohort2 = ""),
       style.tex = style.tex("qje", ),
       fitstat = ~samlet+ se+ n + hs )

panelc <- etable(rl2c, emp2c, make_purchase2c, se.below = T,
       tex = T, title = "panel c",
       drop = panelb_controls,
       digits = "r3",digits.stats = "r3",
       signif.code = NA,
       dict=c(treatment = "Treatment",
              license_w3 = "Received license",
              G1_3_abovemed = "Allowed to make purchase w/o permission",
              employed_w3 = "Employed",
              single = "never married",
              randomization_cohort2 = ""),
       style.tex = style.tex("qje", ),
       fitstat = ~n+mcd + mcm + mcnm + mcw)



panela;panelb;panelc
```

Oppgave 2

```
kovariater <- c("age_med_BL_control", "miss_age_PAP", "edu_nohs_BL_control",
  "miss_edu_category", "married_control", "single_control", "widowed_control",
  "miss_relationship", "household_size_control", "miss_household_size", "one_car_control","miss_cars",
```

Table 4: Panel A

	Received license (1)	Employed (2)	Allowed to make purchase w/o permission (3)
Treatment	0.526	0.139	0.038
	(0.056)	(0.063)	(0.070)
Above median age	0.150	-0.041	0.274
	(0.062)	(0.075)	(0.089)
Treatment × Above median age	-0.189	-0.105	-0.246
	(0.076)	(0.082)	(0.092)
B1 + B3	0.338	0.034	-0.208
	(0.033)	(0.057)	(0.067)
Observations	467	488	486
Mean: control, below median age	0.092	0.247	0.329
fixed effects	✓	✓	✓

Table 5: panel B

	Received license (1)	Employed (2)	Allowed to make purchase w/o permission (3)
Treatment	0.507	0.078	-0.023
	(0.052)	(0.055)	(0.059)
Less than HS	-0.001	-0.125	0.080
	(0.057)	(0.060)	(0.079)
Treatment × Less than HS	-0.230	0.029	-0.186
	(0.079)	(0.077)	(0.093)
B1 + B3	0.277	0.107	-0.209
	(0.058)	(0.059)	(0.074)
Observations	459	479	477
Mean: control, completed HS	0.129	0.265	0.451
fixed effects	✓	✓	✓

Table 6: panel c

	Received license (1)	Employed (2)	Allowed to make purchase w/o permission (3)
Treatment	0.433	-0.029	-0.221
	(0.059)	(0.072)	(0.072)
Treatment × married	-0.156	0.105	0.058
	(0.101)	(0.111)	(0.131)
Treatment × never married	0.161	0.175	0.344
	(0.087)	(0.107)	(0.108)
Treatment × widowed	-0.234	0.295	0.031
	(0.145)	(0.135)	(0.159)
Observations	463	484	482
Mean: control, divorced	0.083	0.250	0.597
Mean: control, married	0.091	0.171	0.472
Mean: control, never married	0.080	0.246	0.293
Mean: control, widowed	0.208	0.080	0.654
fixed effects	✓	✓	✓

```
data <- df[,c(kovariater, "license_w3", "treatment")] |> na.omit()


Z <- as.matrix(data[, kovariater])

treat <- data$treatment |> as.numeric()

Y <- data$license_w3 |> as.numeric()
```

Bruker fremgangsmetiden som MLr3 har anbefalt på netttsiden deres. Boken “Applied Machine Learning Using mlr3 in R” (Bischl et al., 2024)

Bischl, B., Sonabend, R., Kotthoff, L., & Lang, M. (2024). *Applied Machine Learning using mlr3 in R*.CRC Press https://mlr3book.mlr-org.com/

jeg vet egentlig ikke om Elastic net egner seg. ettersom at vi må standarisere

```
ml_data <- as_task_classif(data, "license_w3")
```

Jeg tror nok det er denne jeg burde bruke

Tuner Elastic Net

Tuner Random Forest

```
as.data.table(instance$archive)[, .(num.trees, mtry ,classif.acc)] |> arrange(desc(classif.acc)) |> hea
```

```
##    num.trees  mtry classif.acc
##        <int> <int>       <num>
## 1:       242     3   0.7197965
## 2:       147     3   0.7176688
## 3:       716     3   0.7176688
## 4:       574     3   0.7176226
## 5:       337     3   0.7175763
## 6:       479     3   0.7154024
```

Tuner neural net

```
as.data.table(instance_nnet$archive)[, .(size, decay ,classif.acc)] |> arrange(desc(classif.acc)) |> hea
```

```
##     size      decay classif.acc
##    <int>      <num>       <num>
## 1:     4 0.10000000   0.7239130
## 2:     1 0.00000000   0.7175763
## 3:     1 0.08947368   0.7175301
## 4:     1 0.07894737   0.7175301
## 5:     3 0.02105263   0.7175301
## 6:     1 0.09473684   0.7175301
```

```
library(GenericML)

set.seed(3110)

learners <- c(
  "mlr3::lrn('cv_glmnet')",
  "mlr3::lrn('ranger')",
  "mlr3::lrn('kknn')"
)

elasticnet_learner <- paste("mlr3::lrn('cv_glmnet', alpha =", as.character(alpha_tune), ")")
rf_learner <- paste("mlr3::lrn('ranger', mtry =", as.character(mtry_tune), ",num.trees =", as.character


nnet_learner <- paste("mlr3::lrn('nnet', size =", as.character(size_tune), ",decay =", as.character(dec

learners <- c(elasticnet_learner, rf_learner, nnet_learner)

diff_clan <- setup_diff(subtract_from = "most", subtracted = 1)

num_splits <- 10 # Default

x <- GenericML(
  Z, treat, Y, learners,
  num_splits = num_splits,
  diff_CLAN = diff_clan,
```

```
  significance_level = 0.025,
  parallel = TRUE)

x$best$overview;paste("Modell best egnet for CLAN:", x$best$CLAN)
```

```
##                                                        lambda lambda.bar
## mlr3::lrn('cv_glmnet', alpha = 0.210526315789474 ) 0.0006017464  0.1797731
## mlr3::lrn('ranger', mtry = 3 ,num.trees = 242 )    0.0115340336  0.1971496
## mlr3::lrn('nnet', size = 4 ,decay = 0.1 )          0.0034209867  0.1923983

## [1] "Modell best egnet for CLAN: mlr3::lrn('ranger', mtry = 3 ,num.trees = 242 )"
```

```
name <- c()
d1 <- c()
d2 <- c()
d3 <- c()
d4 <- c()
difference <- c()
c <- 1

for (i in kovariater)
{
  name[c] <- i
  svar <- get_CLAN(x, variable = i, plot = FALSE)[1]$estimate
  d1[c] <- svar[1]
  d2[c] <- svar[2]
  d3[c] <- svar[3]
  d4[c] <- svar[4]
  difference[c] <- svar[5]
  c <- c +1
}


#estimater <- data.frame(name,d1,d2,d3,d4,difference)

estimater <- data.frame(name,d1,d4,difference)

estimater <- estimater |> mutate(difference = d4 -d1)

estimater[,2:ncol(estimater)] <- estimater[,2:ncol(estimater)] |> round(3)

## Runde litt av-

kable(estimater)
```

name	d1	d4	difference
age_med_BL_control	1.000	1.000	0.000
miss_age_PAP	0.000	0.000	0.000
edu_nohs_BL_control	0.712	0.000	-0.712
miss_edu_category	0.000	0.016	0.016
married_control	0.288	0.000	-0.288

name	d1	d4	difference
single_control	0.017	0.763	0.746
widowed_control	0.254	0.000	-0.254
miss_relationship	0.000	0.016	0.016
household_size_control	5.259	5.651	0.392
miss_household_size	0.017	0.000	-0.017
one_car_control	0.525	0.300	-0.225
miss_cars	0.008	0.000	-0.008
LF_BL_control	0.695	0.945	0.250
miss_LF_BL	0.000	0.000	0.000

Generic ML gir kausalt fordi RCT.l