

```
*****  
*****
```

```
**** Aineiston valmistelu - 2. vaihe
```

```
*****
```

```
*****  
*****
```

```
** Luetaan aineisto
```

```
import delimited "${INFILE}", clear stringcols(_all) encoding(UTF-8)
```

```
drop if mi(alv_ilmoitusten_myynti_2021)
```

```
global N_master = _N
```

```
** Tarkistus
```

```
duplicates report ${id}
```

```
if $N_master != `r(unique_value)' {
```

```
    di "Master-tiedostossa tuplahavaintoja - keskeytetään ajo"
```

```
    exit
```

```
}
```

```
** Poistetaan 1. vaiheen koe-ja kontrolliryhmien yritykset
```

```
merge 1:1 ${id} using "${WORKFOLDER_VAIHE1}\output\treatment_group.dta", generate(merge2)  
keepusing(_merge)
```

```
keep if merge2 == 1
```

```
drop merge2
```

```
drop _merge
```

```
merge 1:1 ${id} using "${WORKFOLDER_VAIHE1}\output\control_group.dta", generate(merge2)  
keepusing(_merge)
```

keep if merge2 == 1

drop merge2

drop _merge

** 2-numerotason toimialakoodi

gen toimiala2 = substr(toimialakoodi,1,2)

** Liikevaihtomuuttuja numeeriseksi

destring alv_ilmoitusten_myynti_2021, replace dpcomma

** Liitetään kirjain-tason toimiala ja yhdistetään pienet luokat suurempiin

merge m:1 toimiala2 using "\${INFOLDER}\avain_tol08_2numero_kirjain" , keep(match master)
keepusing(tol08_kirjain)

replace tol08_kirjain = "C" if tol08_kirjain == "B" // yhdistetään kaivostoiminta ja louhinta teollisuuteen

replace tol08_kirjain = "F" if tol08_kirjain == "D" // yhdistetään sähkö-, kaasu-, ja lämpöhuolto rakentamiseen

replace tol08_kirjain = "F" if tol08_kirjain == "E" // yhdistetään vesi-, viemäri- ja jätevesihuolto rakentamiseen

replace tol08_kirjain = "N" if tol08_kirjain == "O" // yhdistetään julkinen hallinto hallintoon

replace tol08_kirjain = "N" if tol08_kirjain == "U" // yhdistetään kansainväliset organisaatiot hallintoon

replace tol08_kirjain = "S" if tol08_kirjain == "T" // yhdistetään kotitalouksien toimiminen työnantajina muuhun palvelutoimintaan

replace tol08_kirjain = "S" if tol08_kirjain == "X" // yhdistetään toimiala tuntematon muuhun palvelutoimintaan

replace tol08_kirjain = "S" if mi(tol08_kirjain) // yhdistetään puuttuva toimialatieto muuhun palvelutoimintaan

** Muodostetaan stratifioidussa otannassa käytettävät luokat

```

count
global mincellsize = ceil(10/(31000/`r(N)'))
global nsceil=8
gen strata = ""
levels tol08_kirjain, local(toimialat)
foreach j in `toimialat' {

    count if tol08_kirjain == "`j'"
    global nsmax_tol = floor(`r(N)')/$mincellsize)
    gen aux = alv_ilmoitusten_myynti_2021 if tol08_kirjain == "`j'"
    if `=min($nsceil,$nsmax_tol)'>=2 {
        xtile auxqnt = aux if tol08_kirjain == "`j'", nq(`=min($nsceil,$nsmax_tol)')
    }
    if `=min($nsceil,$nsmax_tol)'<2 {
        gen auxqnt = 1 if tol08_kirjain == "`j'"
    }

    replace strata = tol08_kirjain + string(auxqnt) if tol08_kirjain == "`j'"
    drop aux
    drop auxqnt

}

gen stratacell = strata

** Järjestetään y-tunnuksen mukaan

sort ${id}

** Tallennetaan aineisto

```

```
save "${INFOLDER}\full_targetgroup.dta", replace
```