

Yksinyrittäjien rekrytointituen kokeilu- ja verrokkiryhmä – poiminta

**** Poiminta

**** Asetetaan satunnaislukugeneraattorin siemenluku koneajasta

```
global system_time = clock("`c(current_date)`c(current_time)", "DMYhms")
```

```
global seed=substr("$system_time",1,10)
```

```
set seed $seed
```

**** 1. Koeryhmän poiminta

** Ladataan aineisto

```
use "${INFOLDER}\full_targetgroup.dta", clear
```

** Stratifioitu otanta

```
count
```

```
global n1p=100*(3500/`r(N)')
```

```
sample ${n1p}, by(stratacell)
```

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

```
global NT_initial = _N
```

** Stratifioitu otanta voi antaa pyöristyvirheestä johtuen hieman ryhmän tavoitekoosta poikkeavan havaintomäärän.

** Täsmäys tehdään tarvittaessa suoralla otannalla.

** Jos stratifioitu otanta antoi täsmälleen 3500 havaintoa, tallennetaan taulu

```
if $NT_initial == 3500 {
```

```
    gen treatment=1
```

```
    save "${OUTFOLDER}\treatment_group.dta", replace}
```

** Jos stratifioitu otanta antoi yli 3500 havaintoa, poistetaan havainnot suoralla otannalla

```
if $NT_initial > 3500 & $NT_initial!=. {
```

```
    sample 3500, count
```

```
    gen treatment=1
```

```
    save "${OUTFOLDER}\treatment_group.dta", replace }
```

** Jos stratifioitu otanta antoi alle 3500 havaintoa, täydennetään havainnot suoralla otannalla jäljelle jääneistä yritysistä

```
if $NT_initial < 3500 {
```

```
    use "${INFOLDER}\full_targetgroup.dta", clear
```

```

merge 1:1 ${id} using "${INFOLDER}\treatment_group_initial.dta", gen(match)
keep(match master)
drop if match==3
drop match
global Nadd1 = 3500 - $NT_initial
sample $Nadd1 , count
append using "${INFOLDER}\treatment_group_initial.dta"
gen treatment=1
save "${OUTFOLDER}\treatment_group.dta", replace}

* Poistetaan väliaikainen tiedosto
erase "${INFOLDER}\treatment_group_initial.dta"

**** 2. Kontrolliryhmän poiminta

** Ladataan aineisto
use "${INFOLDER}\full_targetgroup.dta", clear

** Poistetaan koeryhmään poimitut
merge 1:1 ${id} using "${OUTFOLDER}\treatment_group.dta", nogen keep(match master)
keepusing(treatment)
drop if treatment==1
drop treatment

** Stratifioitu otanta
count
global n0p=100*(20000/'r(N)')
sample $n0p, by(stratacell)
save "${INFOLDER}\control_group_initial.dta", replace
global NC_initial = _N

** Jos stratifioitu otanta antoi täsmälleen 20000 havaintoa, tallennetaan taulu
if $NC_initial == 20000 {
    gen treatment=0
    save "${OUTFOLDER}\control_group.dta", replace}

** Jos stratifioitu otanta antoi yli 20000 havaintoa, poistetaan havainnot suoralla otannalla
if $NC_initial > 20000 & $NC_initial!=. {
    sample 20000, count
    gen treatment=0
    save "${OUTFOLDER}\control_group.dta", replace}

** Jos stratifioitu otanta antoi alle 20000 havaintoa, täydennetään havainnot suoralla otannalla jäljelle
jääneistä yrityksistä
if $NC_initial < 20000 {
    use "${INFOLDER}\full_targetgroup.dta", clear

```

```
merge 1:1 ${id} using "${OUTFOLDER}\treatment_group.dta", gen(match) keep(match
master) keepusing(treatment)
drop if match==3
drop match treatment
merge 1:1 ${id} using "${INFOLDER}\control_group_initial.dta", gen(match) keep(match
master)
drop if match==3
drop match
global Nadd2 = 20000 - $NC_initial
sample $Nadd2 , count
append using "${INFOLDER}\control_group_initial.dta"
gen treatment=0
save "${OUTFOLDER}\control_group.dta", replace}
** Poistetaan väliaikainen tiedosto
erase "${INFOLDER}\control_group_initial.dta"
```