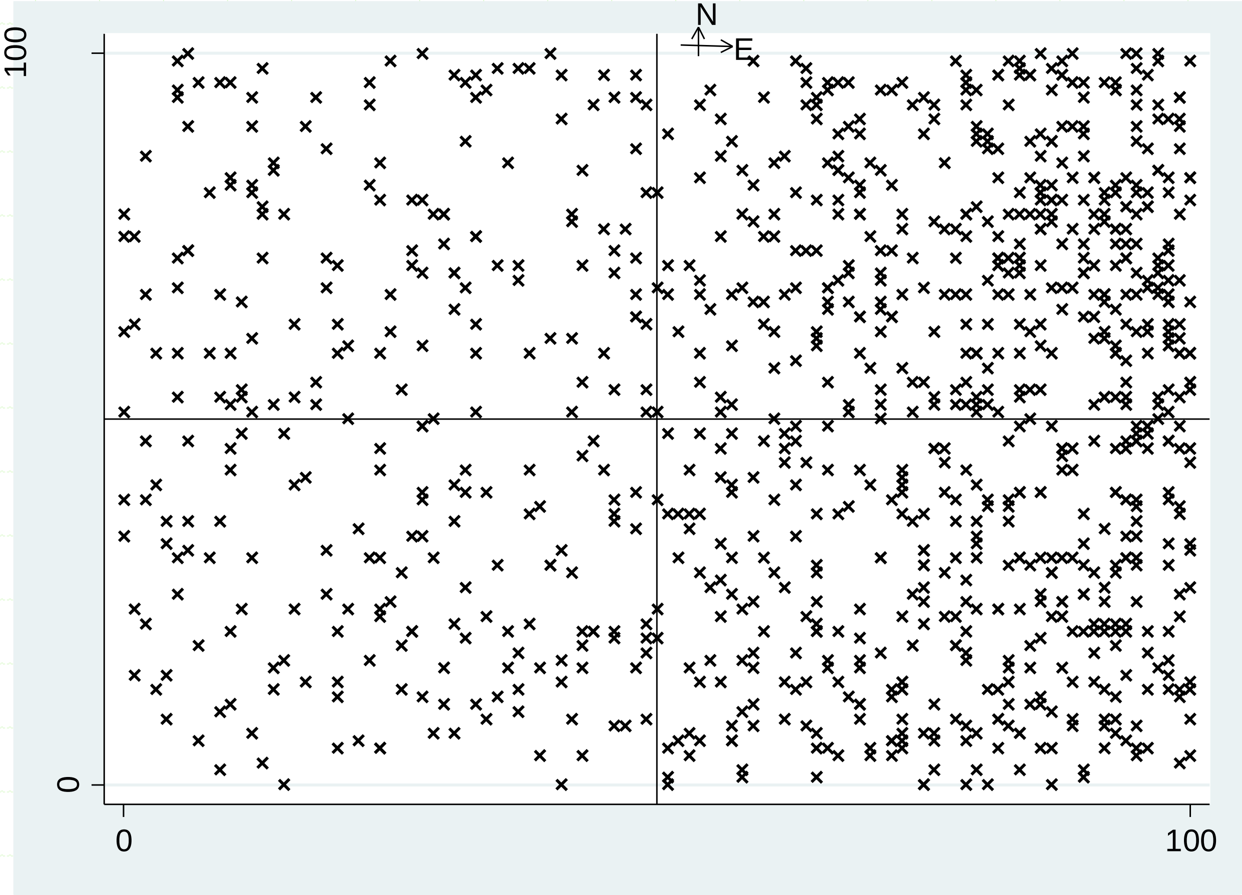
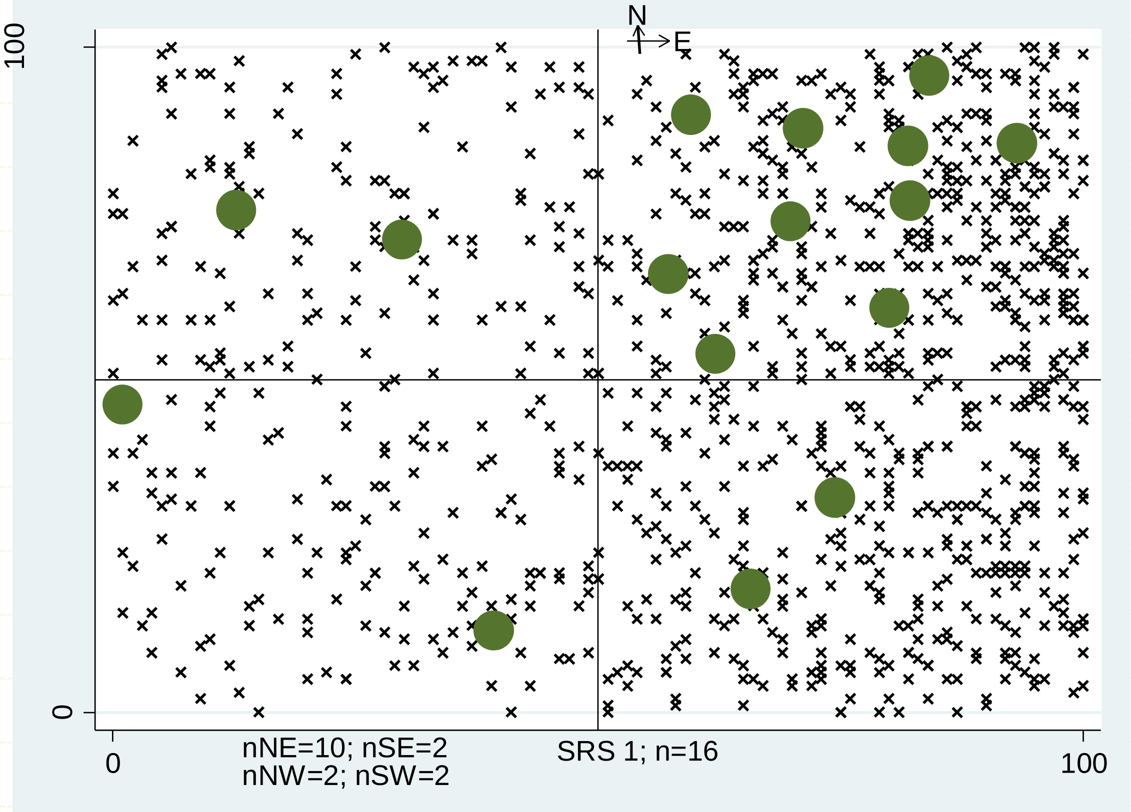
Population

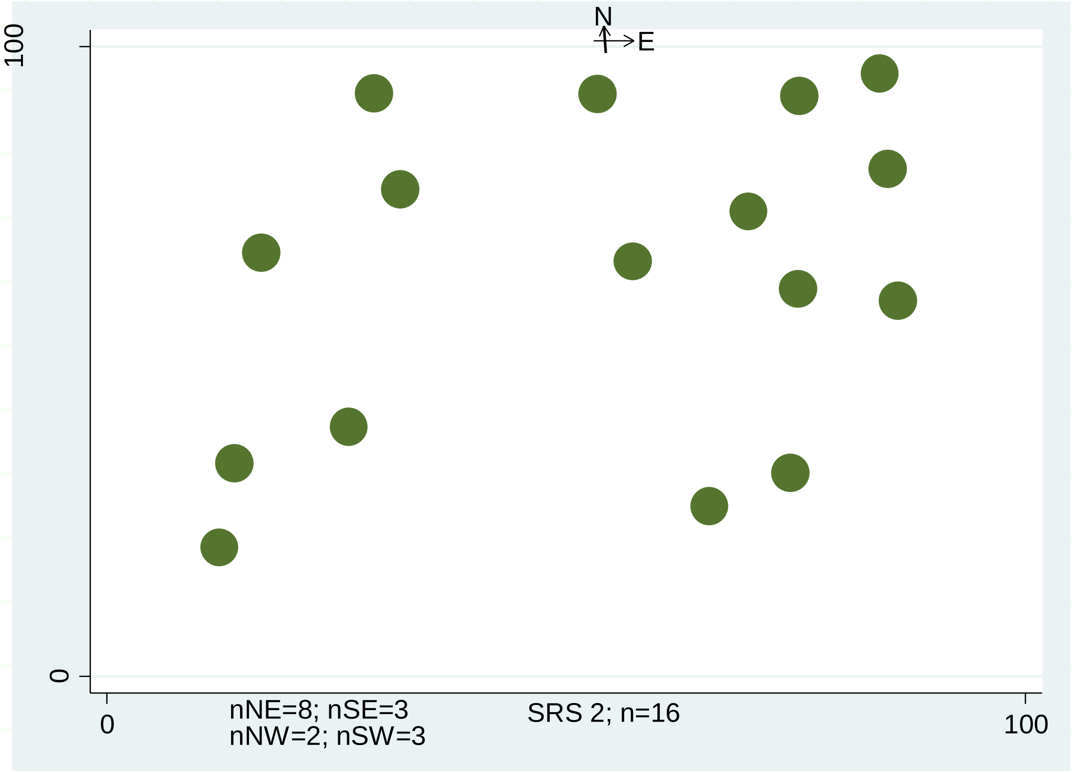


SRS1



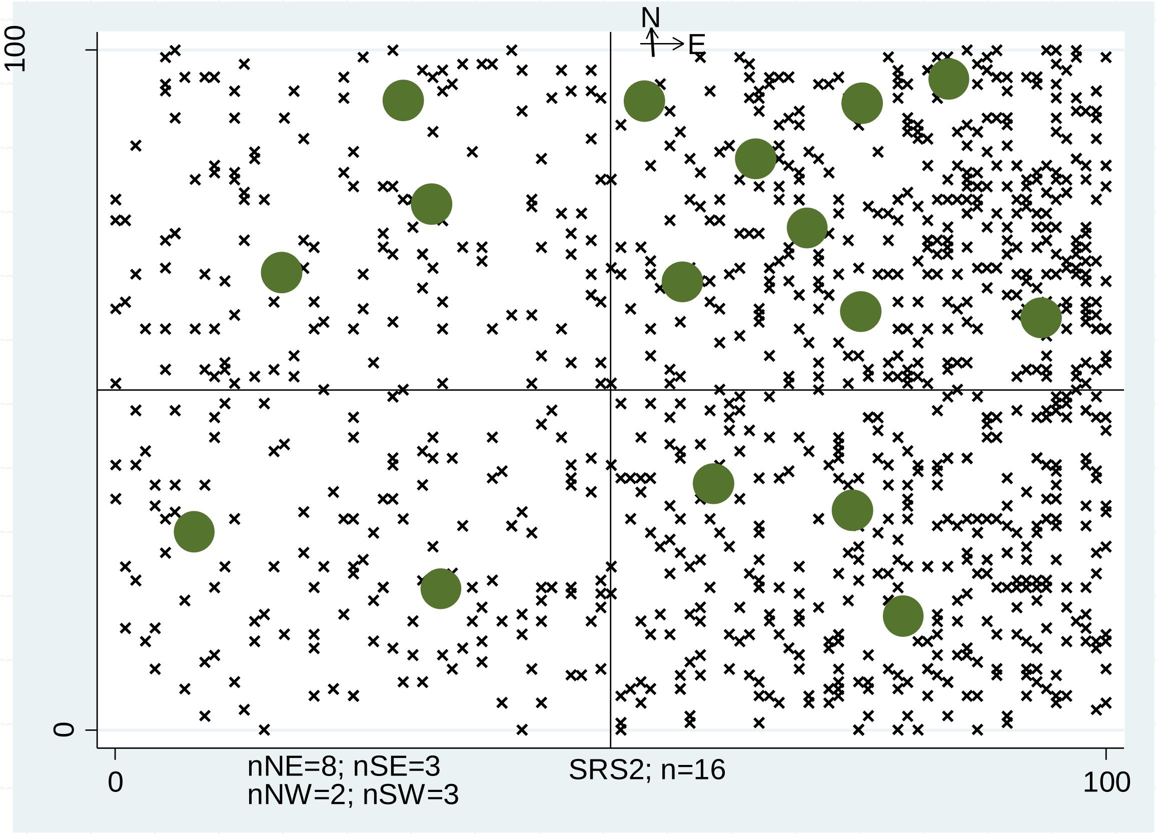
$$\bar{y}_{SRS1} \approx \bar{Y}_{POP}$$

SRS2



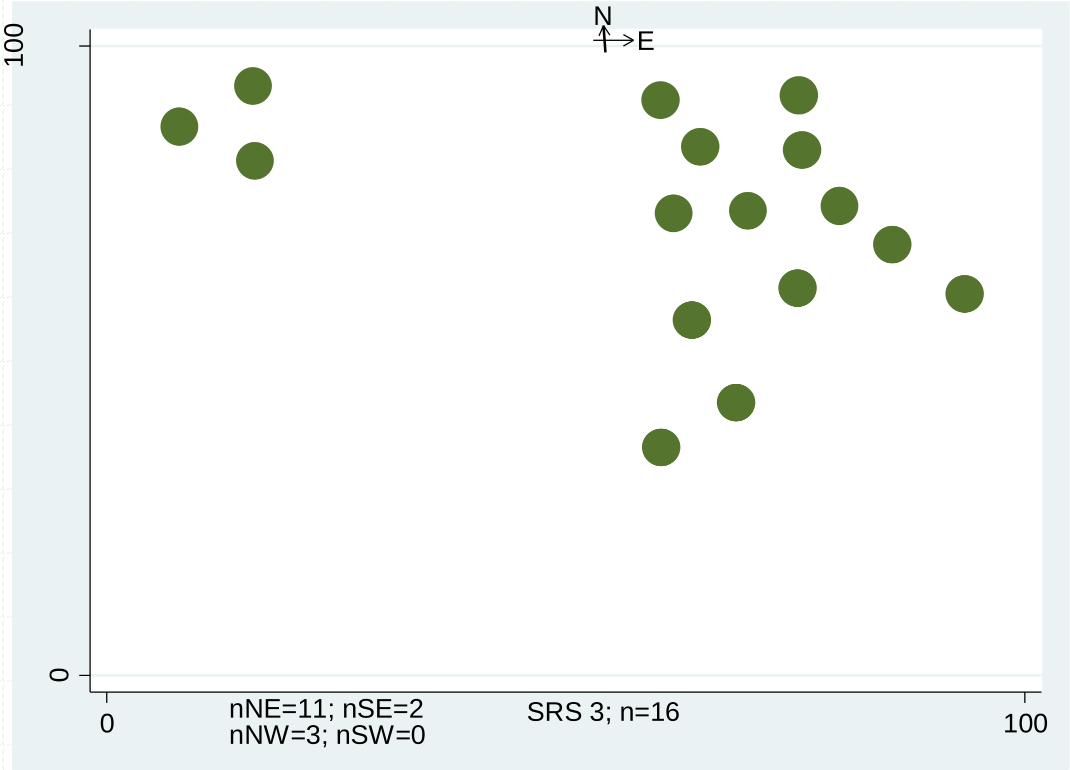
$$\bar{y}_{SRS2} < \bar{Y}_{POP}$$

SRS2



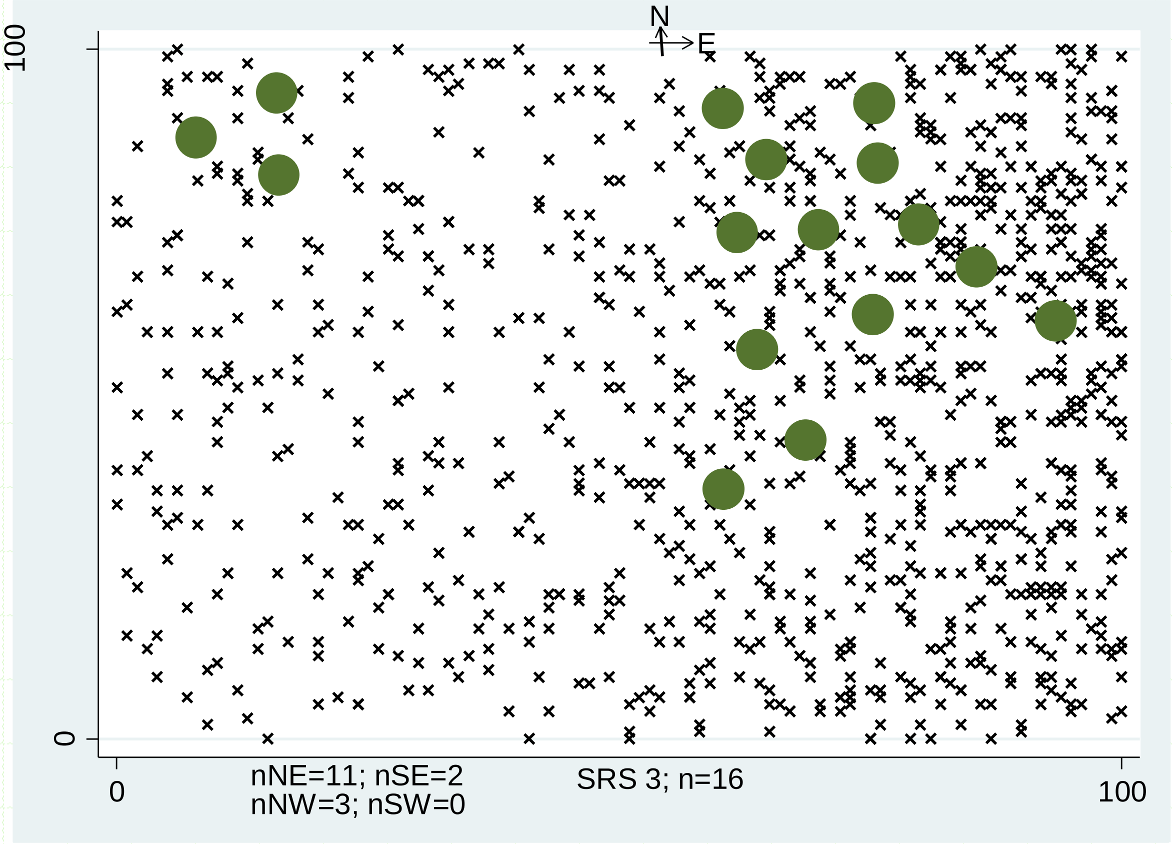
$$\bar{y}_{SRS2} < \bar{Y}_{POP}$$

SRS3



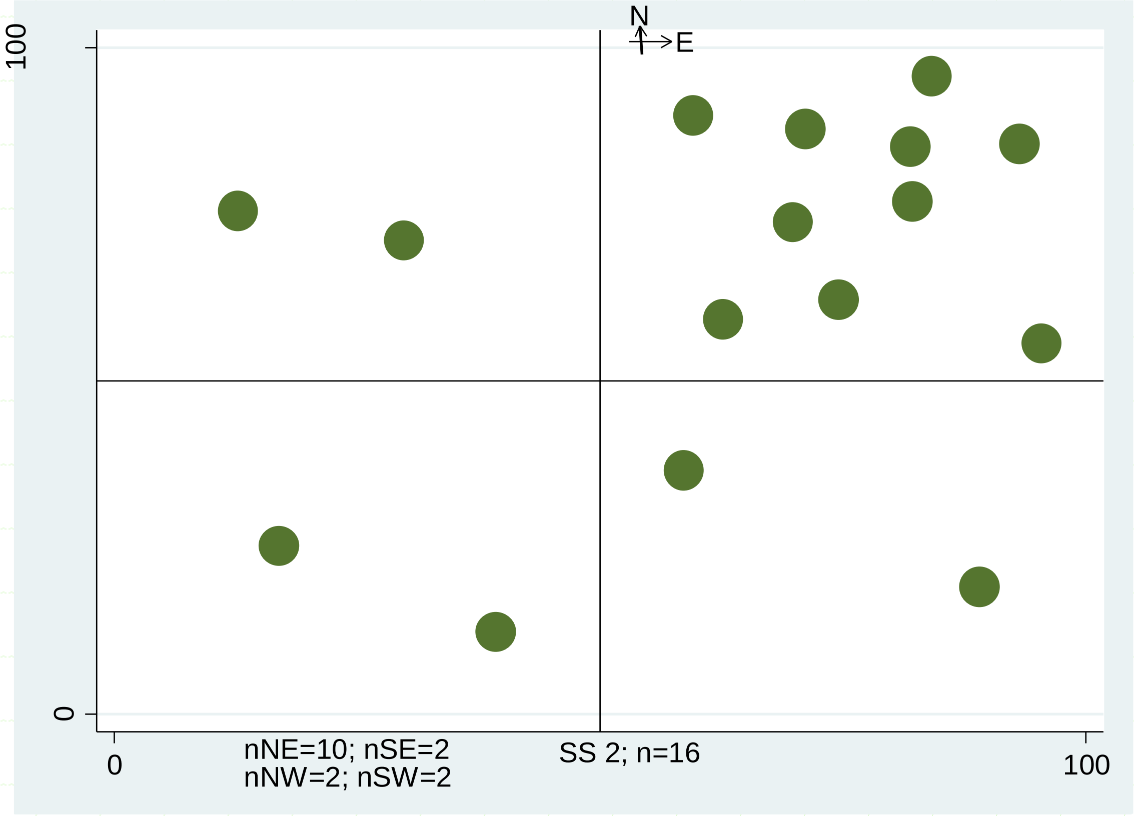
$$\bar{y}_{SRS3} > \bar{Y}_{POP}$$

SRS3



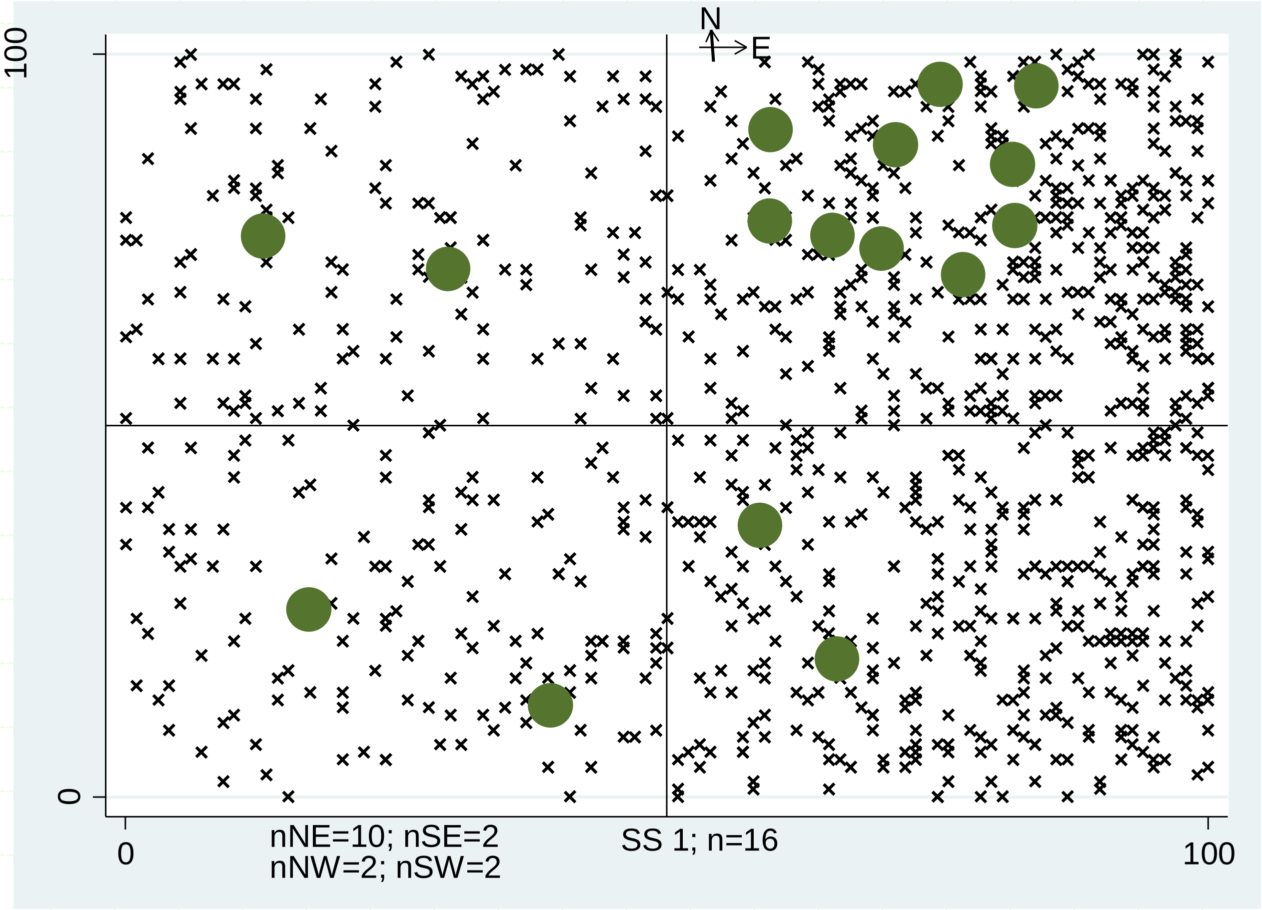
$$\bar{y}_{SRS3} > \bar{Y}_{POP}$$

SS1



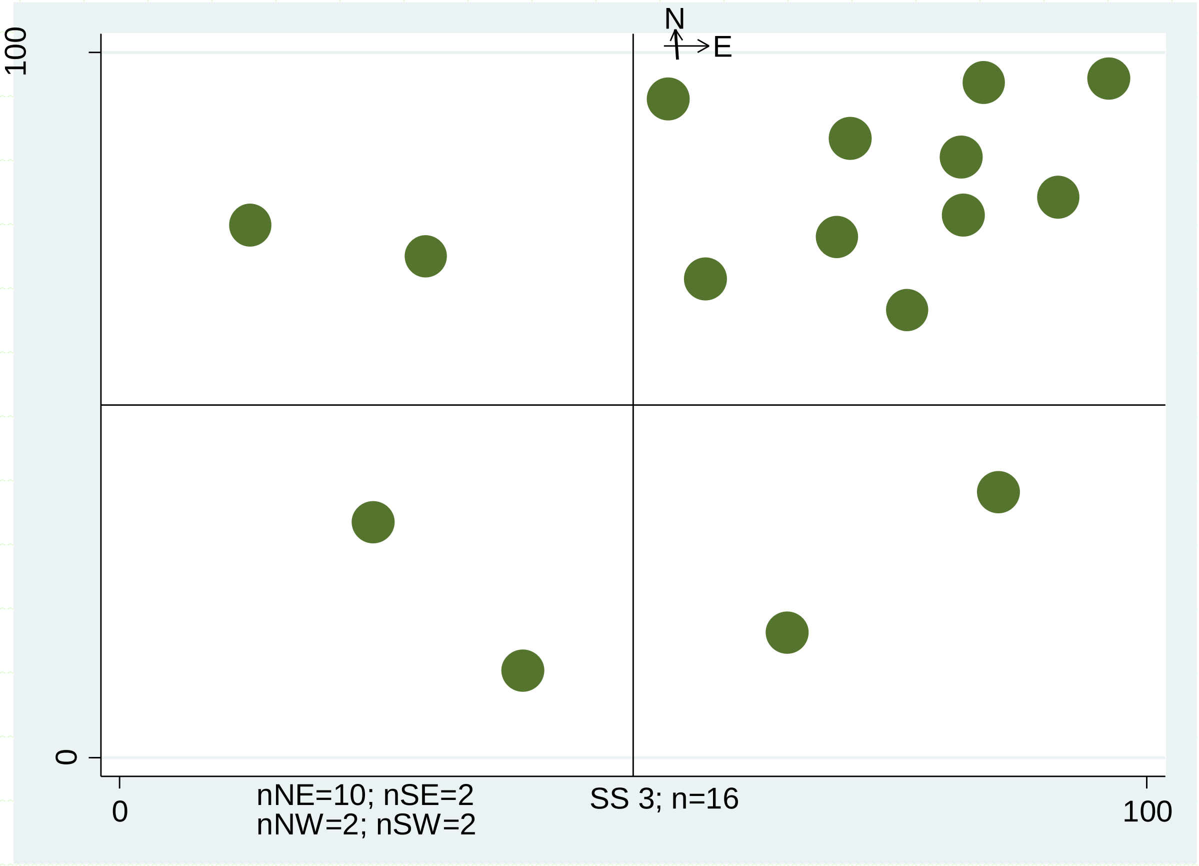
$$\bar{y}_{SS1} \approx \bar{Y}_{POP}$$

SS1



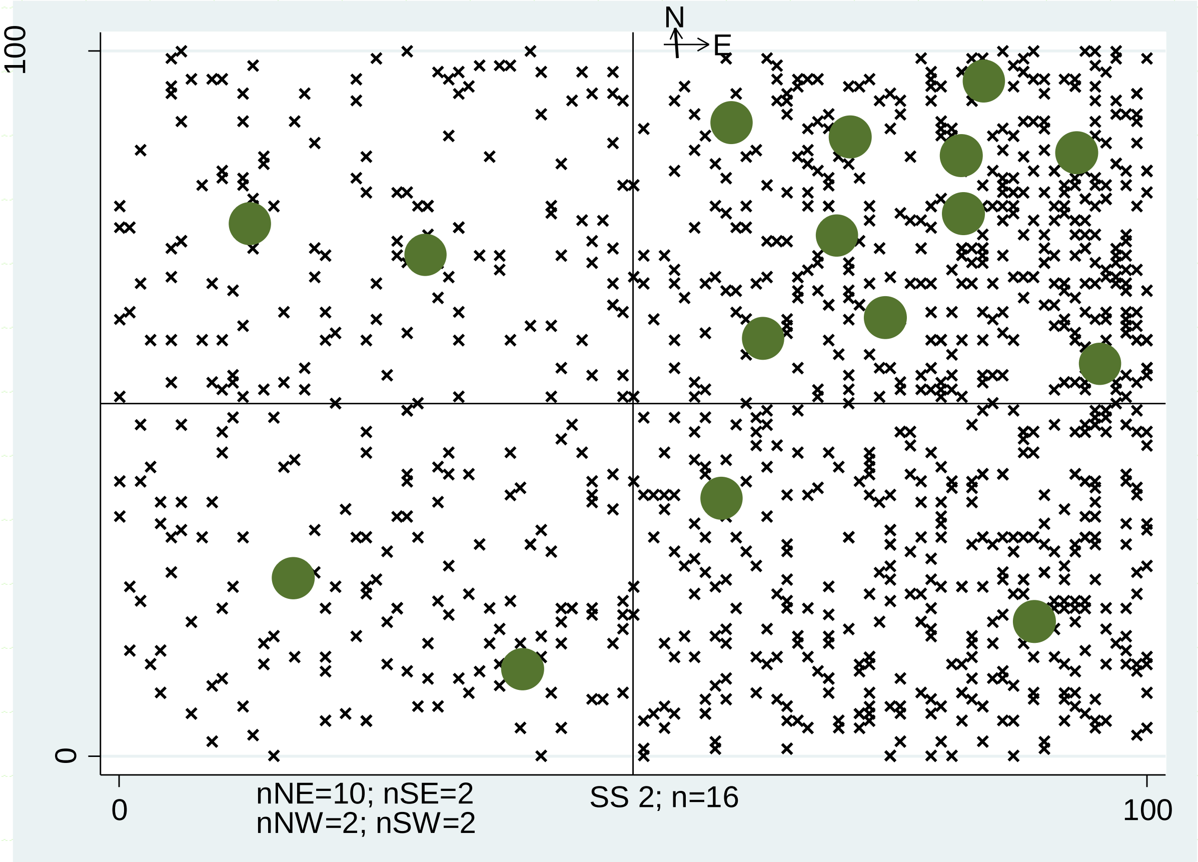
$$\bar{y}_{SS1} \approx \bar{Y}_{POP}$$

SS2



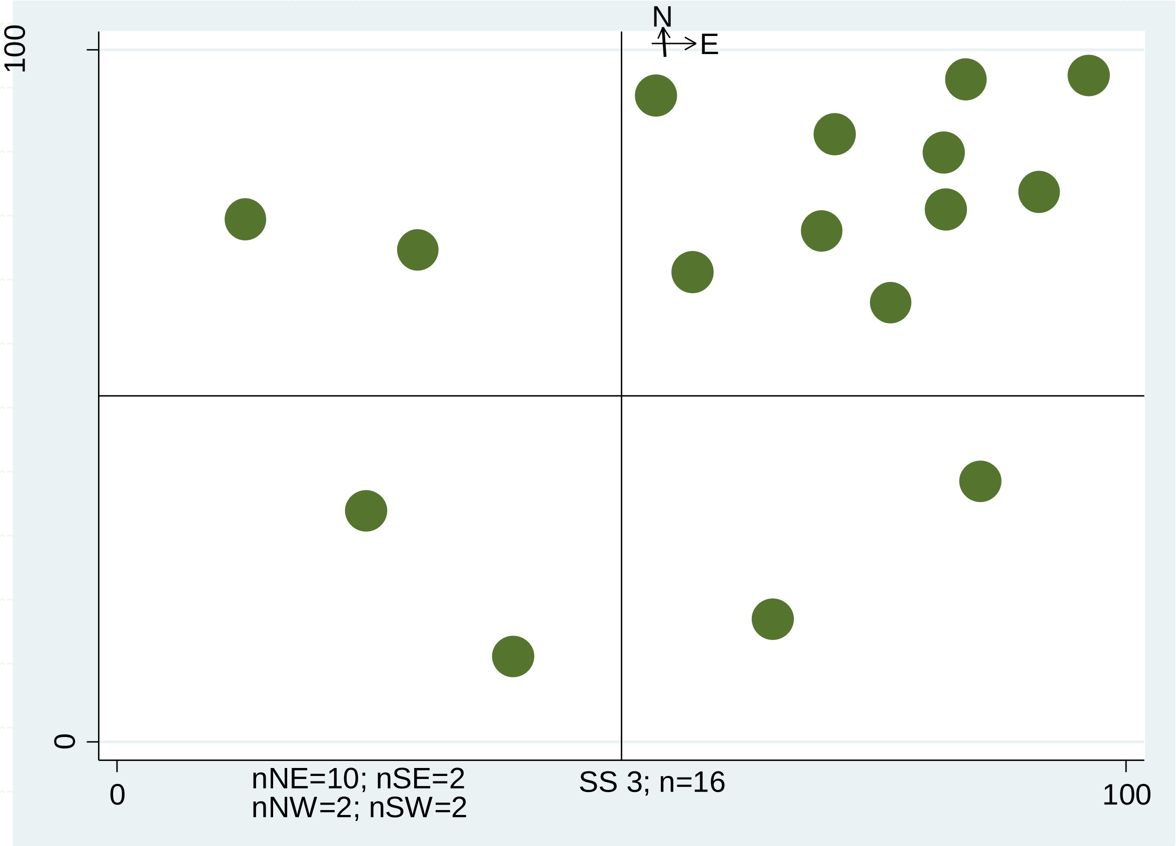
$$\bar{y}_{SS2} \approx \bar{Y}_{POP}$$

SS2



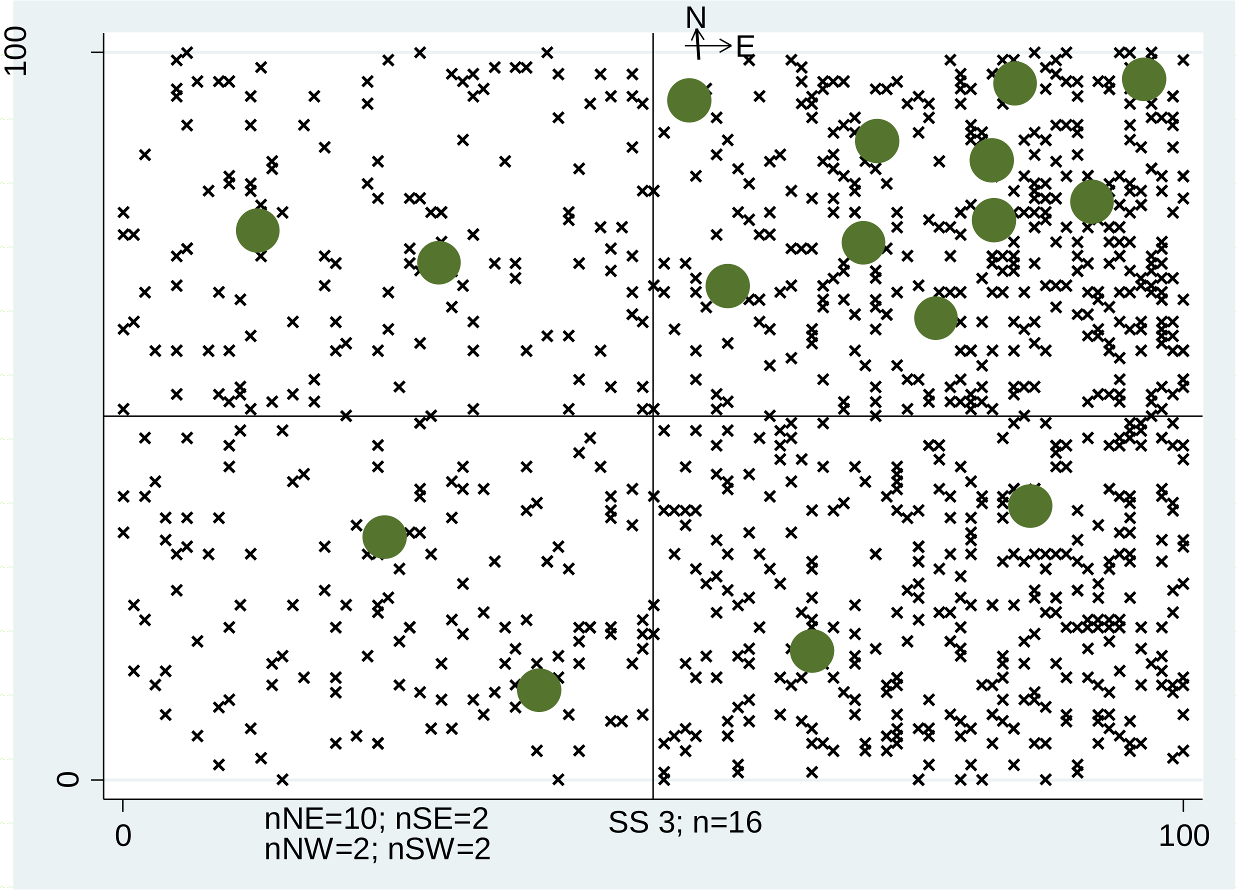
$$\bar{y}_{SS2} \approx \bar{Y}_{POP}$$

SS3



$$\bar{y}_{SS3} \approx \bar{Y}_{POP}$$

SS3



$$\bar{y}_{SS3} \approx \bar{Y}_{POP}$$

Repeat SRS and SS 100 times

Estimate $\mathrm{Var}(\bar{y}_{SRS})$ and $\mathrm{Var}(\bar{y}_{SS})$

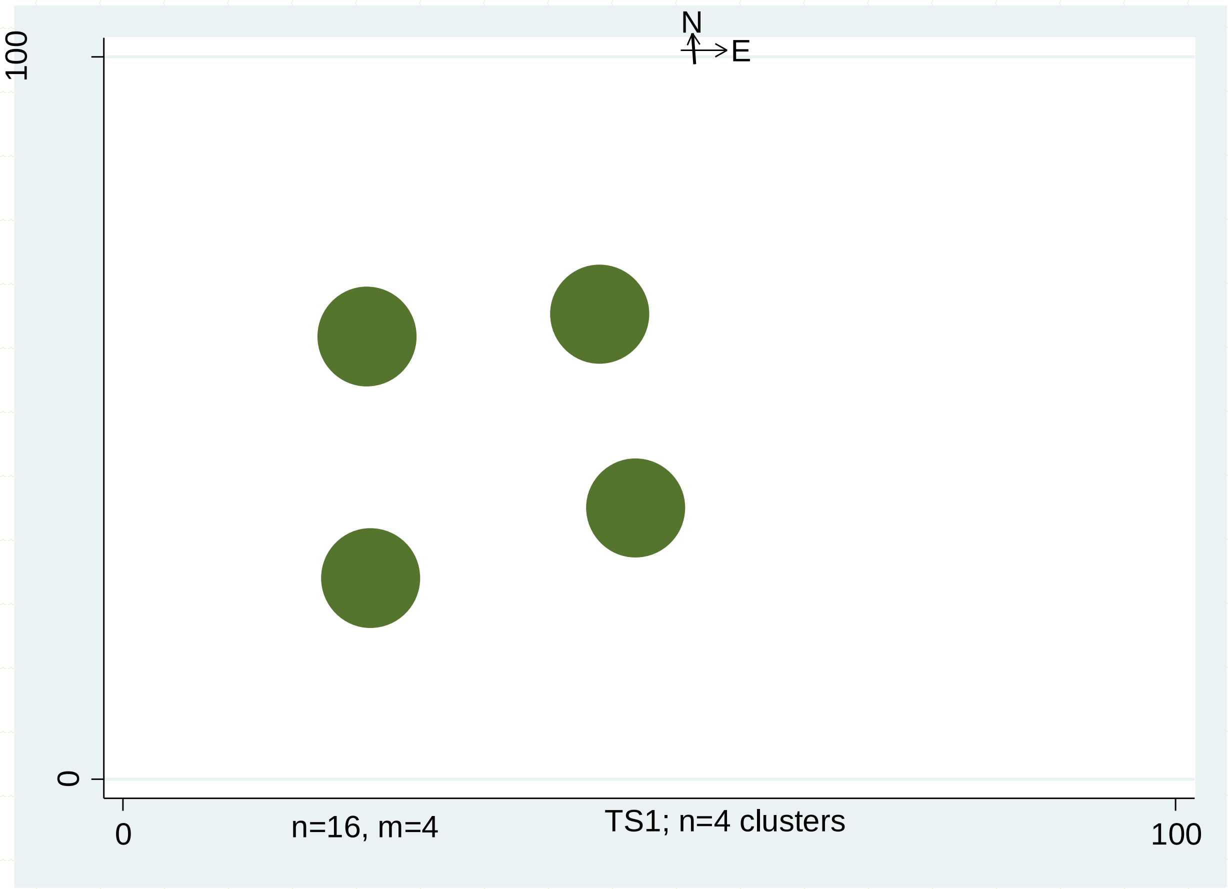
$\mathrm{Var}(\bar{y}_{SRS})$ and $\mathrm{Var}(\bar{y}_{SS})$

$$\mathrm{Var}(\bar{y}_{SRS}) = \frac{1}{(n-1)}\sum_{i=1}^{100}(\bar{y}_{SRS_i} - \bar{Y})^2$$

$$\mathrm{Var}(\bar{y}_{SS}) = \frac{1}{(n-1)}\sum_{i=1}^{100}(\bar{y}_{SS_i} - \bar{Y})^2$$

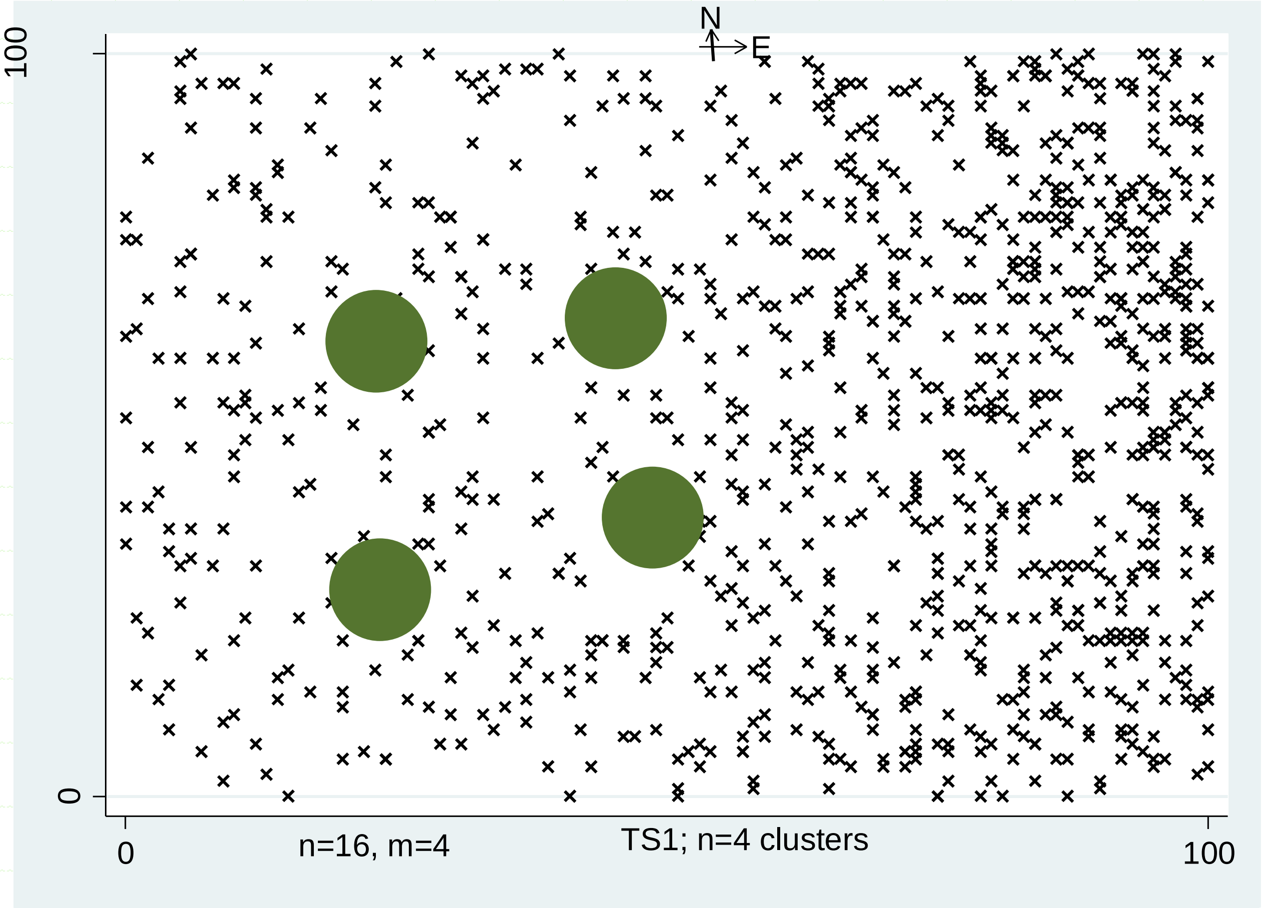
$$\mathrm{Var}(\bar{y}_{SS}) \overset{?}{>=<} \mathrm{Var}(\bar{y}_{SRS})$$

Two-stage TS1



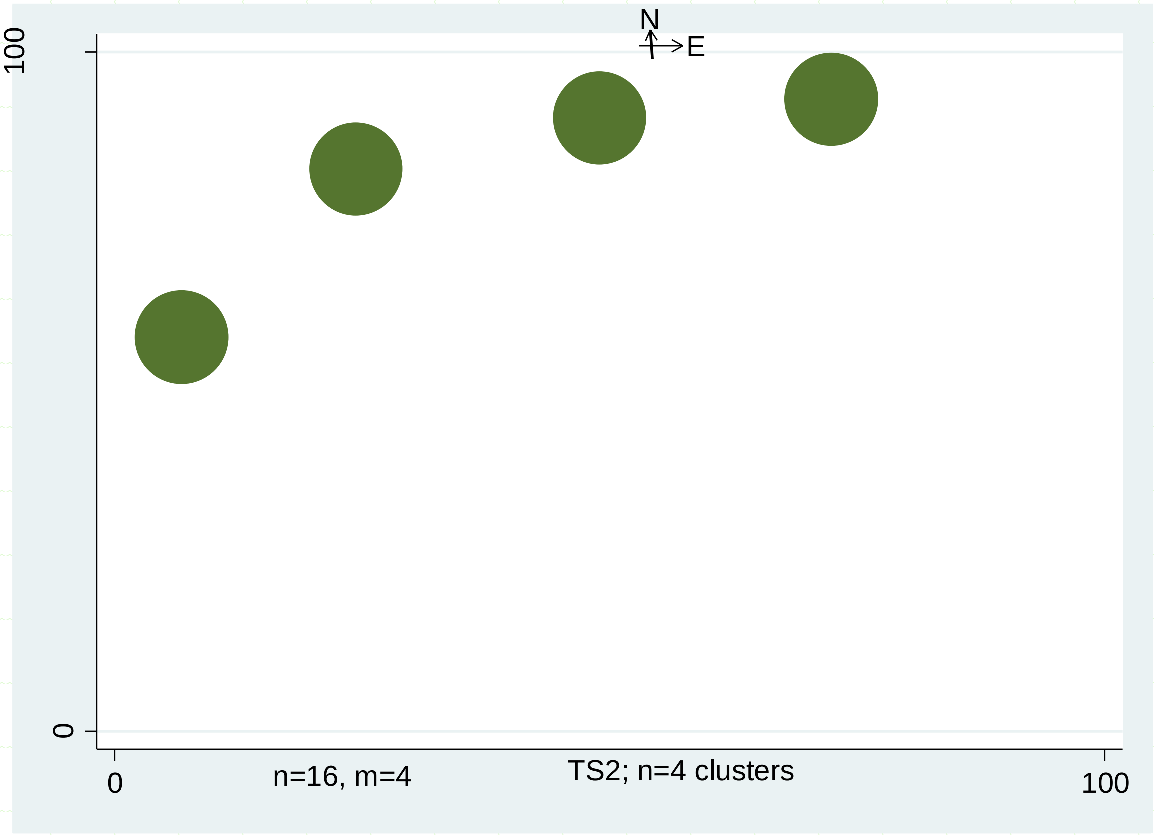
$$\bar{y}_{CS1} < \bar{Y}_{POP}$$

Two-stage TS1



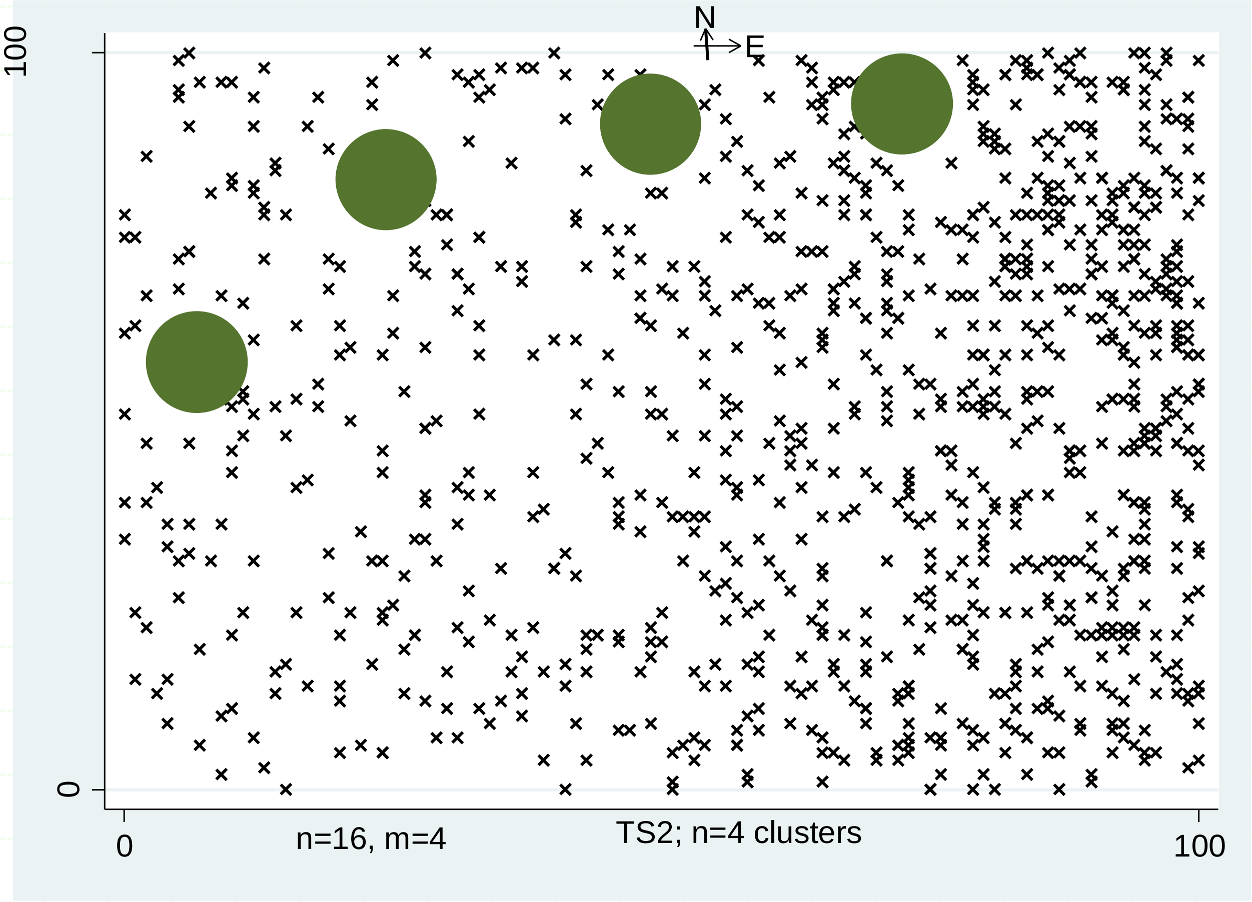
$$\bar{y}_{CS1} < \bar{Y}_{POP}$$

Two-stage TS2



$$\bar{y}_{CS2} > \bar{Y}_{POP}$$

Two-stage TS2



$$\bar{y}_{CS2} > \bar{Y}_{POP}$$

Repeat SRS and CS 100 times

Estimate $\mathrm{Var}(\bar{y}_{SRS})$ and $\mathrm{Var}(\bar{y}_{cS})$

$\mathrm{Var}(\bar{y}_{SRS})$ and $\mathrm{Var}(\bar{y}_{cS})$

$$\mathrm{Var}(\bar{y}_{SRS}) = \frac{1}{(n-1)}\sum_{i=1}^{100}(\bar{y}_{SRS_i} - \bar{Y})^2$$

$$\mathrm{Var}(\bar{y}_{cS}) = \frac{1}{(n-1)}\sum_{i=1}^{100}(\bar{y}_{cS_i} - \bar{Y})^2$$

$$\mathrm{Var}(\bar{y}_{cS}) \overset{?}{>=<} \mathrm{Var}(\bar{y}_{SRS})$$

Clustering - loss of precision (m)

$$Var(\bar{y}_{CS}) = Var(\bar{y}_{SRS})[1+(m-1)\rho]$$

$$\Rightarrow \quad Var(\bar{y}_{CS}) \geq Var(\bar{y}_{SRS})$$

Precision is decreased as ....

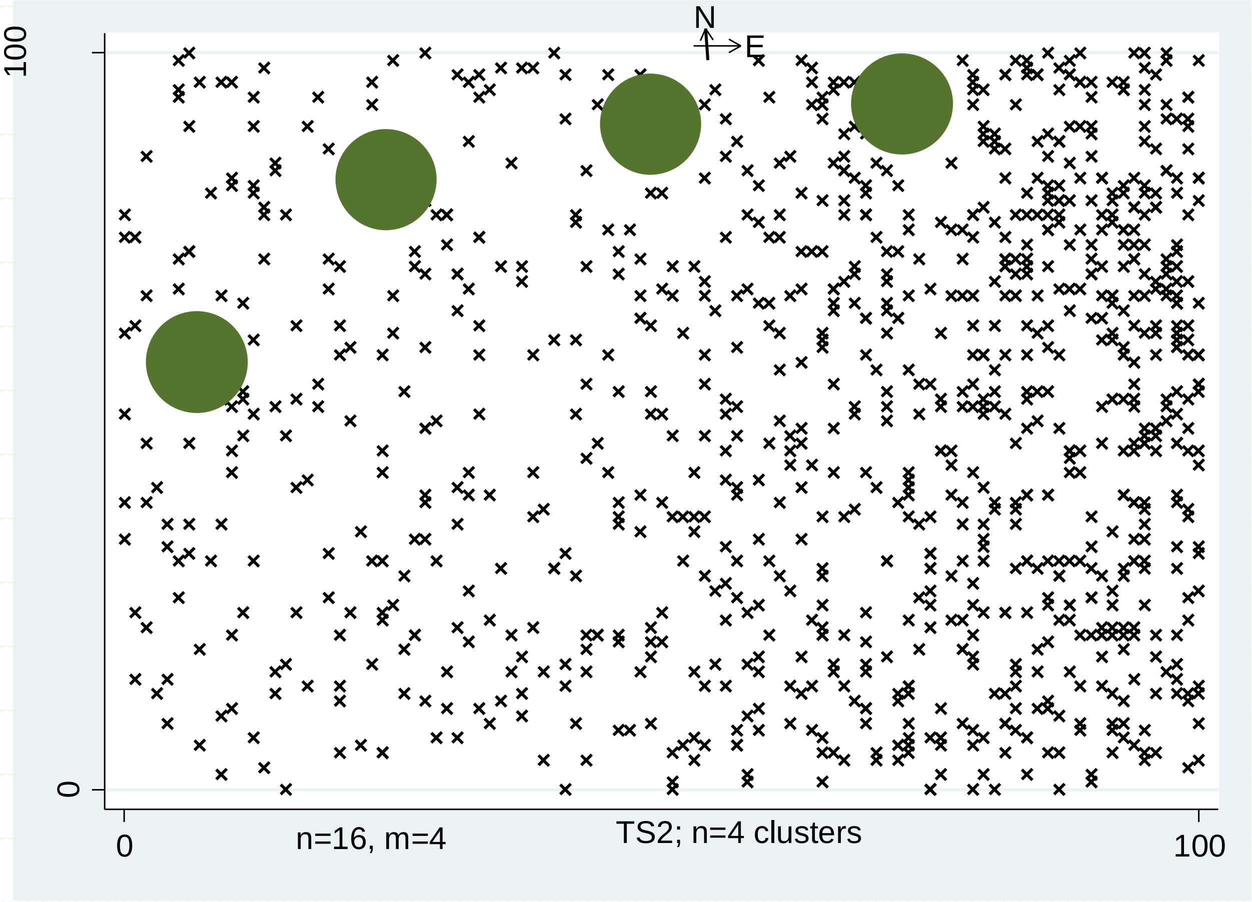
... ρ increases, or in words, as the correlation of the $y_{i,c}$ obs increases within each cluster. The more similar the observations are within clusters.

... m increases. For a fixed sample size, as the number of observations within a cluster increases.

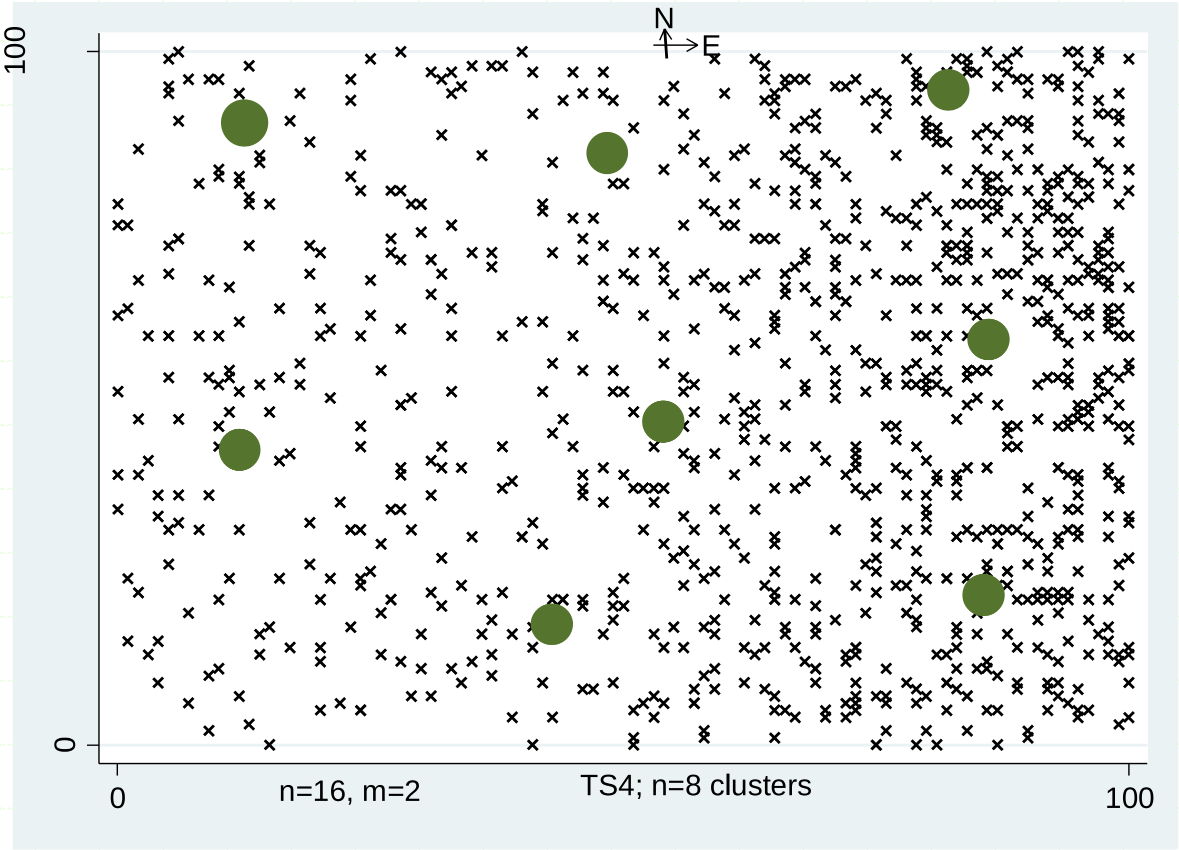
Why? Less coverage of the variation in the frame.

As m shrinks to 1, CS moves to SRS.

Two-stage TS2

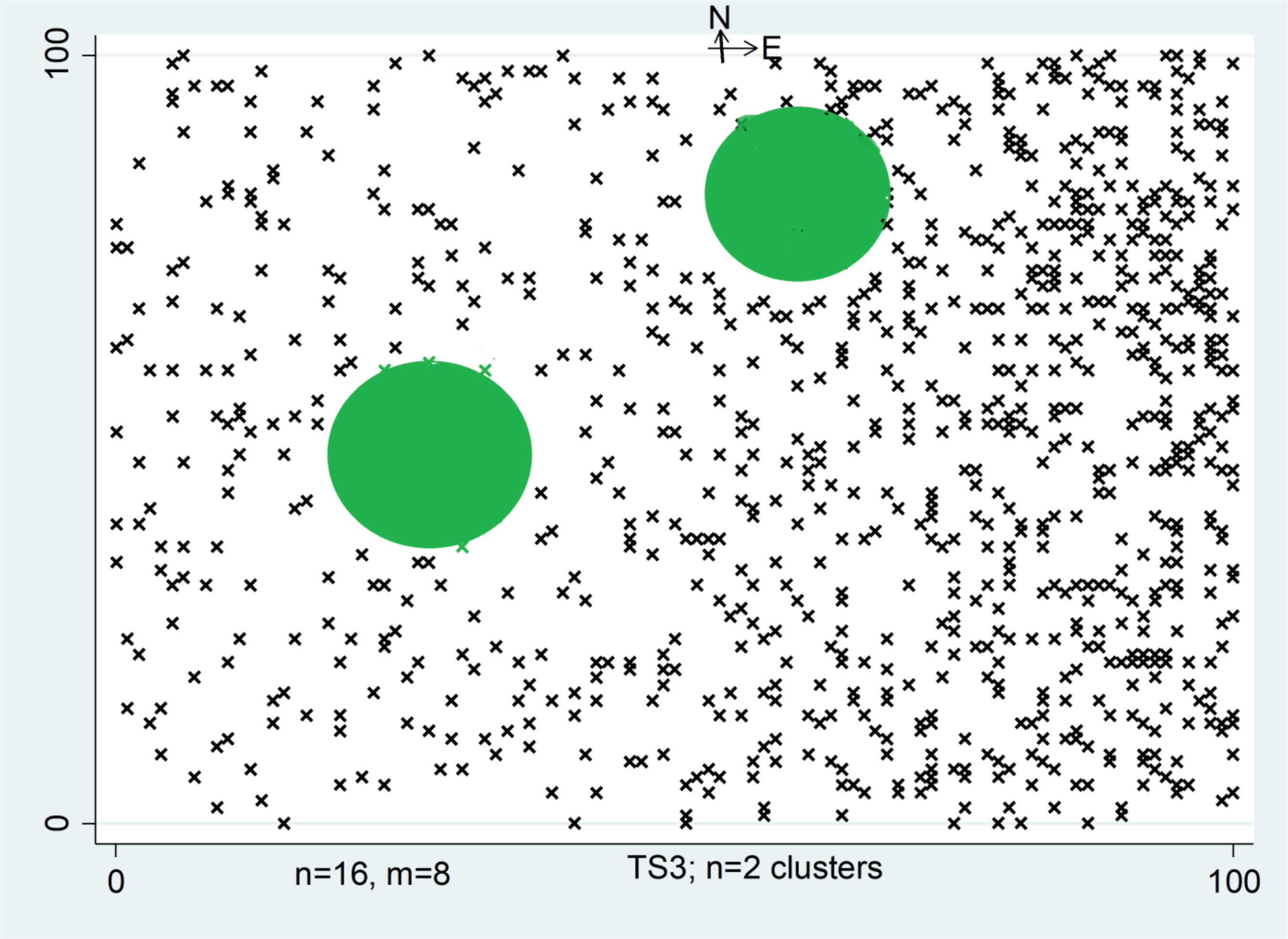


Two-stage TS4



$$\bar{y}_{cS3} \approx \bar{Y}_{POP}$$

Two-stage TS3



$$\bar{y}_{cS3} > \bar{Y}_{POP}$$