

# Spatial Sampling in R

Maria Michela Dickson

Dip. Scienze Statistiche, UniPD



# Il campionamento

La **teoria del campionamento** è l'ambito delle scienze statistiche che si occupa dei criteri da seguire nella progettazione della selezione di un campione da una popolazione finita o non finita, con l'obiettivo di fare inferenza su di essa, al costo più contenuto possibile.

Definita una popolazione di  $N$  unità, selezioniamo un **campione**  $s$  quale sottoinsieme di  $n \leq N$  unità della popolazione, in accordo ad una funzione di probabilità.

**Disegno di campionamento:** distribuzione di probabilità sul sottoinsieme di tutti i possibili campioni  $\mathcal{S}$ :

$$P(S) \geq 0, \forall S \in \mathcal{S}$$

# Il campionamento

**Obiettivo:** stima di un parametro della popolazione (totale o media).

**Stimatore:**  $\hat{\theta}$  di un parametro  $\theta$  è qualsiasi statistica campionaria scelta per assegnare un valore al parametro considerato (uno stimatore è una funzione che associa un valore  $\theta$  a ciascun campione  $S$ ).

**Stima:** il valore numerico dello stimatore nel campione selezionato.

**Errore di stima:** la differenza tra la stima e il valore reale del parametro.

$Y$  è la variabile oggetto di stima  $\rightarrow y_1, y_2, y_3, \dots, y_N \rightarrow Y = \sum y_j$

# Il campionamento spaziale

Quando la variabile di interesse è distribuita su uno spazio geografico e sono note le posizioni delle unità della popolazione, allora possiamo implementare un disegno di **campionamento spaziale**.

La variabile spazialmente distribuita mostra dipendenza spaziale: sarebbe ridondante prelevare due campioni da luoghi molto vicini tra loro.

# Il campionamento spaziale

Possono essere campionati diversi tipi di popolazioni:

- distribuite su uno spazio continuo (ad es. la popolazione dei valori di inquinamento atmosferico in un'area urbana)
- un insieme di oggetti discreti (ad es. l'insieme di tutte le famiglie o l'insieme di tutti i distretti censuari)

# Disegni di campionamento spaziale

- Metodi basati sulle unità contigue (e.g. Hedayat *et al.*, 1988; Wright e Stufken, 2008)
- Generalized Random Tessellation Stratified (Stevens e Olsen, 2004)
- **Pivot locale (Grafström *et al.*, 2012)**
- **Spatial Correlated Poisson Sampling (Grafström, 2012)**
- **Local cube (Grafström e Tillé, 2013)**

# In R: il pacchetto “BalancedSampling”

## Balanced Sampling

Select balanced and spatially balanced probability samples in multi-dimensional spaces with any prescribed inclusion probabilities. The local pivotal method and spatially correlated Poisson sampling (for spatially balanced sampling) are included. Also the cube method (for balanced sampling) and the local cube method (for doubly balanced sampling) are included.

Current version: 1.5.4

*By Anton Grafström*

Swedish University of Agricultural Sciences

# In R: il pacchetto “BalancedSampling”

```
#  
#####  
#   Installazione e caricamento   #  
#           pacchetti              #  
#####  
#  
library(BalancedSampling)  
#  
data(meuse.all)  
dim(meuse.all)  
str(meuse.all)  
#  
# Definizione dimensione di popolazione e campionaria  
N <- length(meuse.all[,1])  
n <- 50
```

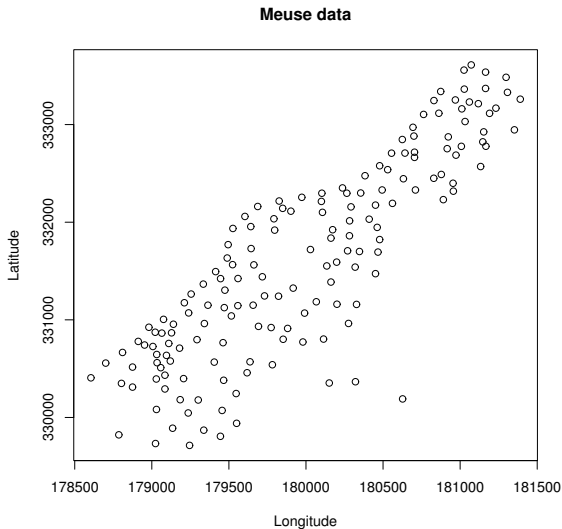


# In R: il pacchetto "BalancedSampling"

R Data: meuse.all

	sample	x	y	cadmium	copper	lead	zinc	elev	dist.m	om	ffreq	soil	lime	landuse
1	1	181072	333611	11.7	85	299	1022	7.909	50	13.6	1	1	1	Ah
2	2	181025	333558	8.6	81	277	1141	6.983	30	14.0	1	1	1	Ah
3	3	181165	333537	6.5	68	199	640	7.800	150	13.0	1	1	1	Ah
4	4	181298	333484	2.6	81	116	257	7.655	270	8.0	1	2	0	Ga
5	5	181307	333330	2.8	48	117	269	7.480	380	8.7	1	2	0	Ah
6	6	181390	333260	3.0	61	137	281	7.791	470	7.8	1	2	0	Ga
7	7	181165	333370	3.2	31	132	346	8.217	240	9.2	1	2	0	Ah
8	8	181027	333363	2.8	29	150	406	8.490	120	9.5	1	1	0	Ab
9	9	181060	333231	2.4	37	133	347	8.668	240	10.6	1	1	0	Ab
10	10	181232	333168	1.6	24	80	183	9.049	420	6.3	1	2	0	W
11	11	181191	333115	1.4	25	86	189	9.015	400	6.4	1	2	0	Fh
12	12	181032	333031	1.8	25	97	251	9.073	300	9.0	1	1	0	Ag
13	13	180874	333339	11.2	93	285	1096	7.320	20	15.4	1	1	1	W
14	14	180969	333252	2.5	31	183	504	8.815	130	8.4	1	1	0	Ah
15	15	181011	333161	2.0	27	130	326	8.937	220	9.1	1	1	0	Ah
16	16	180830	333246	9.5	86	240	1032	7.702	10	16.2	1	1	1	W
17	17	180763	333104	7.0	74	133	606	7.160	10	16.0	1	1	1	W
18	18	180694	332972	7.1	69	148	711	7.100	10	16.0	1	1	1	W
19	19	180625	332847	8.7	69	207	735	7.020	10	13.7	1	1	1	W

# In R: il pacchetto “BalancedSampling”



# In R: il pacchetto “BalancedSampling”

```
#
#####
#      Campionamento LPM      #
#####
#
# Definizione probabilità di inclusione
prob <- rep(n/N, N)
# Definizione coordinate geografiche
x <- cbind(meuse.all$x, meuse.all$y)
#
# Setto il seme generatore
set.seed(2025)
#
# Selezione del campione
s_lpm <- lpm(prob,x)
# Campione selezionato
s_lpm
#
```

# In R: il pacchetto “BalancedSampling”

```
#
#####
#      Campionamento SCPS      #
#####
#
# Setto il seme generatore
set.seed(2025)
# Selezione del campione SCPS
s_scps <- scps(prob,x)
# Campione selezionato
s_scps
#
#
```

# In R: il pacchetto “BalancedSampling”

```
#
#####
#           Campionamento Lcube           #
#####
#
# Setto il seme generatore
set.seed(2025)
# Selezione del campione Local cube
Xspread <- cbind(meuse.all$x, meuse.all$y)
Xbal <- cbind(meuse.all$x, meuse.all$y)
s_lcube <- lcube(prob, Xspread, Xbal)
# Campione selezionato
s_lcube
#
-
```

# In R: il pacchetto “BalancedSampling”

```
# Confronto tra stime #  
  
#####  
# Campionamento casuale semplice #  
# senza ripetizione #  
#####  
#  
library(sampling)  
# Setto il seme generatore  
set.seed(2025)  
# Selezione del campione  
s1 <- srswor(n,N)  
# Indicatore di inclusione  
s1  
# Aggancio alle variabili  
sample1 <- getdata(meuse.all, s1)  
#
```

## In R: il pacchetto “BalancedSampling”

```
#####  
# Stimatore di Horvitz-Thompson #  
#           del totale           #  
#           (CCS)                #  
#####  
#  
# Calcolo probabilità di inclusione  
pik <- rep(n/N,N)  
#  
# Definizione variabile sotto stima "zinc"  
y <- meuse.all$zinc  
# Calcolo delle stime H-T  
estTOTl <- HTestimator(y[s1==1],rep(n/N, n))  
#  
# Errore relativo  
errRelccs <- (estTOTl-sum(y))/sum(y)  
#  
#
```

# In R: il pacchetto “BalancedSampling”

```
#
#####
# Stimatore di Horvitz-Thompson #
#   del totale                 #
#   (LPM)                      #
#####
#
# Calcolo delle stime H-T
estTOTspat1 <- HTestimator(y[s_lpm],prob[s_lpm])
#
# Errore relativo
errRel1pm <- (estTOTspat1-sum(y))/sum(y)
#
#####
# Stimatore di Horvitz-Thompson #
#   del totale                 #
#   (SCPS)                     #
#####
#
# Calcolo delle stime H-T
estTOTspat2 <- HTestimator(y[s_scps],prob[s_scps])
#
# Errore relativo
errRelscps <- (estTOTspat2-sum(y))/sum(y)
#
#####
# Stimatore di Horvitz-Thompson #
#   del totale                 #
#   (Lcube)                    #
#####
#
# Calcolo delle stime H-T
estTOTspat3 <- HTestimator(y[s_lcube],prob[s_lcube])
#
# Errore relativo
errRelcube <- (estTOTspat3-sum(y))/sum(y)
#
```



## In R: il pacchetto “BalancedSampling”

```
>
> #
> risultati <- cbind(errRelccs, errRelldpm, errRelscps, errRelcube)
> risultati
      [,1]      [,2]      [,3]      [,4]
[1,] 0.1497262 -0.01609471 -0.08281912 0.05278211
>
>
>
>
>
>
> |
```

# In R: il pacchetto “BalancedSampling”

```
#
par(mfrow=c(2,2))
# Plot con CCS
plot(meuse.all$x, meuse.all$y, main="Selezione CCS", ylab="Latitude", xlab="Longitude")
points(meuse.all$x[s1==1], meuse.all$y[s1==1], col="green", pch=16)
# Plot con LPM
plot(meuse.all$x, meuse.all$y, main="Selezione LPM", ylab="Latitude", xlab="Longitude")
points(meuse.all$x[s_lpm], meuse.all$y[s_lpm], col="violet", pch=16)
# Plot con SCPS
plot(meuse.all$x, meuse.all$y, main="Selezione SCPS", ylab="Latitude", xlab="Longitude")
points(meuse.all$x[s_scps], meuse.all$y[s_scps], col="blue", pch=16)
# Plot con Lcube
plot(meuse.all$x, meuse.all$y, main="Selezione Lcube", ylab="Latitude", xlab="Longitude")
points(meuse.all$x[s_lcube], meuse.all$y[s_lcube], col="red", pch=16)
#
```

# In R: il pacchetto “BalancedSampling”

