

# TEORIA E TECNICA DELL'INDAGINE STATISTICA E DEL CAMPIONAMENTO (MATR. DISPARI)

## CAMPIONAMENTO A STADI

MANUELA SCIONI

Dipartimento di Scienze Statistiche

[manuela.scioni@unipd.it](mailto:manuela.scioni@unipd.it)



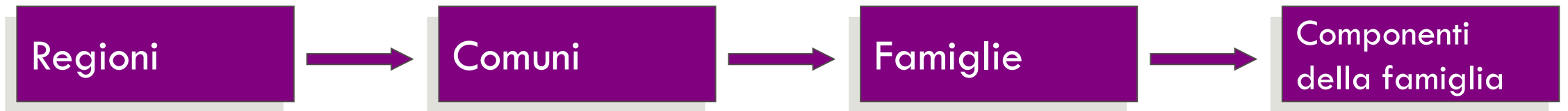
UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA



# CAMPIONAMENTO SU PIÙ STADI

Si basa sul concetto di *popolazioni gerarchiche*: la popolazione finale di unità è contenuta in un aggregato di unità di livello (o stadio) superiore, le quali possono essere contenute in unità sempre più ristrette in numero e ampie in dimensione

ESEMPIO



Procedimento: si campionano le unità di I stadio, poi quelle di II stadio contenute nelle unità di I stadio selezionate, e così via.

# PERCHÉ CAMPIONAMENTO A STADI

- Non si dispone di un'unica lista della popolazione statistica, e crearla a partire dalle liste disponibili è troppo oneroso
- La popolazione è per sua natura organizzata in cluster (es. scuole)
- La popolazione è distribuita in un territorio vasto, e vincoli organizzativi ed economici impediscono di effettuare un CCS o un Campionamento stratificato

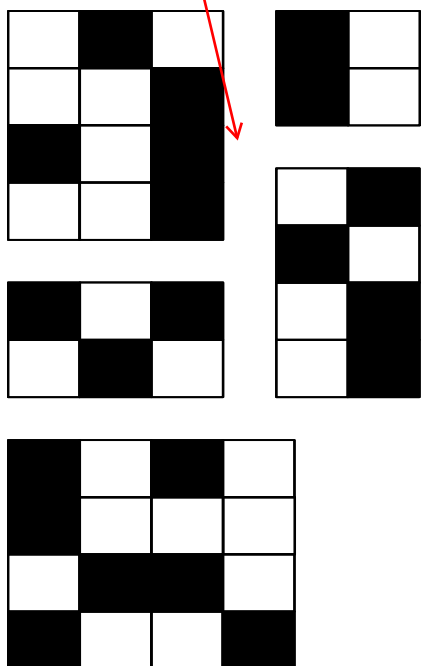
SCELTE N S O  
E O CENTRO

RIPARTIZIONI GEOGRAFICHE  
NORD SUD OVEST EST

A STADI ad esempio meno  
efficace xke solo NORD , SUD

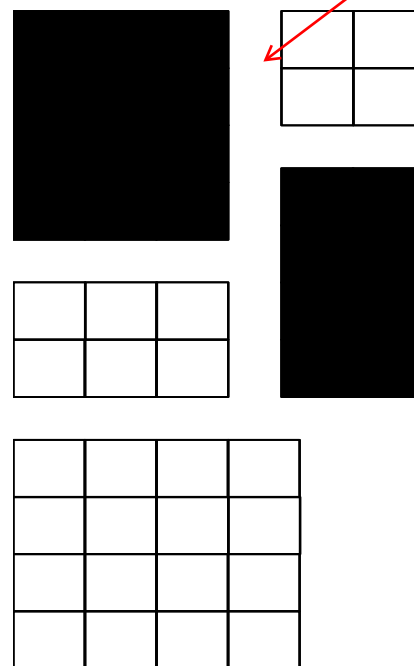
# CAMPIONAMENTO STRATIFICATO VS CAMP. A STADI

CUMUNI SCELTI NEL NORD CHE è STATO  
SCELTO AL PRIMO STADIO UPS

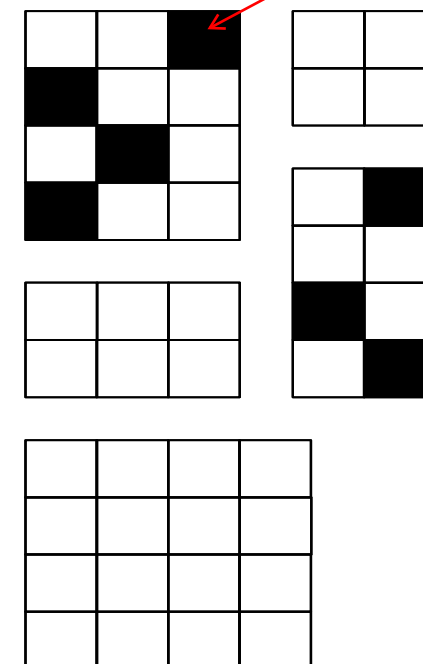


Stratificato

maggiore efficienza delle  
stime rispetto al ccs



1 stadio (grappoli)



2 stadi

minore efficienza rispetto al  
ccs

# CAMPIONAMENTO STRATIFICATO VS CAMP. A STADI

Non confondiamo gli **strati** con gli **stadi**:

La stratificazione porta quasi sempre ad una maggiore efficienza delle stime, mentre il camp. a stadi porta quasi sempre ad una minore efficienza

Perché?

Camp. stratificato: si procede a selezionare unità da **OGNI** strato, per cui il disegno è tanto più efficiente quanto più le unità dentro gli strati si assomigliano

Camp. a stadi: si procede a selezionare solo **ALCUNE** unità di primo stadio (UPS), che devono rappresentare tutta la popolazione, anche quella contenuta nelle UPS non selezionate. Per cui le unità di secondo stadio (USS) contenute nelle UPS selezionate devono essere il più possibile distribuite come la popolazione.

es: comuni  
delle regioni

regione

se correlazione intraclasse bassa allora il campionamento a stadi puo dare buoni risultati , cmq la varianza triplica rispetto al ccs, per questo sicuramente il camp a stadi è quello meno efficiente, a maggior ragione di uno stratificato, ma quello più economico

## PERCHÉ IL CAMPIONAMENTO A STADI È MENO EFFICIENTE?

- Le unità all'interno di uno stesso cluster tendono ad assomigliarsi, poiché condividono dei fattori, generalmente non misurabili. Anche la sola residenza nella stessa area geografica implica la condivisione di fattori socio-economici tipici del territorio.
- Poiché però noi abbiamo selezionato solo ALCUNE UPS, e a queste chiediamo di rappresentare TUTTA la popolazione, vorremmo invece che fossero eterogenee.
- Più le USS contenute nella stessa UPS si assomigliano, più aumenta l'inefficienza del camp. a stadi, perché ogni USS aggiuntiva spiega poco più delle altre USS già selezionate.
- La correlazione delle USS contenute nelle UPS è detta **correlazione intraclasse**. Nel campionamento su più stadi l'efficienza delle stime è funzione della correlazione intraclasse

di solito si considerano 2,3 stadi al max, xke se di più aumenta la correlazione

# SCELTE PER FORMARE UN CAMPIONE SU DUE STADI

quante unita estrarre al primo e successivi

Individuazione  
del numero  
appropriato  
di stadi

Determinazione  
della numerosità  
del campione

Individuazione  
delle caratteristiche  
per stratificare  
le unità di  
primo stadio

Quante unità  
selezionare al  
primo e ai  
successivi stadi

- Assegnazione di probabilità di inclusione alle unità
- Come selezionare le unità al primo e ai successivi stadi

quale le unita di ups e  
uss  
meno stadi meglio  
sarà di solito 3 al max

le unita di primo stadio non si campionano con un ccs ma  
con un campionamento stratificato, 'stadistrati' quindi devo  
definire le variabili di stratificazione per il primo stadio.

es: comp delle famiglie culla per le unita di primo stadio

# SCELTE PER FORMARE UN CAMPIONE SU PIÙ STADI

unità al primo stadio

unità secondo stadio

Numero appropriato di stadi: limitare il numero di stadi allo stretto indispensabile, perché più stadi = più correlazione intraclasse

Campionamento al primo stadio: il disegno al primo stadio è quasi sempre stratificato, es. le regioni sono gli strati e i Comuni le UPS

Numerosità campione: scelta del numero di UPS e di USS

- Il criterio è selezionare più UPS, ognuna con un numero limitato di USS, perché così si riduce la correlazione intraclasse
- Quante UPS e USS selezionare è di fatto vincolato a criteri economici ed organizzativi
- Processo di andata e ritorno: determinata la numerosità campionaria, si calcola la precisione delle stime; se è accettabile, il processo termina, altrimenti si prosegue per aggiustamenti successivi.
- In questo processo si possono utilizzare delle funzioni che legano costi e numerosità campionaria (Lohr, pagg.156-157)



# TEORIA E TECNICA DELL'INDAGINE STATISTICA E DEL CAMPIONAMENTO (MATR. DISPARI)

## CAMPIONAMENTO A STADI NOTAZIONE

MANUELA SCIONI

Dipartimento di Scienze Statistiche

[manuela.scioni@unipd.it](mailto:manuela.scioni@unipd.it)



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA



# NOTAZIONE: POPOLAZIONE

$K = n^{\circ} \text{ UPS}$

$M_i = n^{\circ} \text{ USS nell'i-ma UPS}$

$N = \sum_i M_i = n^{\circ} \text{ totale di USS nella pop}$

$t_i = \sum_{j=1}^{M_i} Y_{ij} = \text{totale nell'i-ma UPS}$

$t = \sum_{i=1}^K t_i = \sum_{i=1}^K \sum_{j=1}^{M_i} Y_{ij} = \text{totale nella popolazione}$

$S_t^2 = \frac{1}{K-1} \sum_{i=1}^K \left( t_i - \frac{t}{K} \right)^2 = \text{varianza del totale delle UPS}$

$\bar{Y} = \sum_{i=1}^K \sum_{j=1}^{M_i} \frac{Y_{ij}}{N} = \text{media generale}$

$\bar{Y}_i = \sum_{j=1}^{M_i} \frac{Y_{ij}}{M_i} = \frac{t_i}{M_i} = \text{media nell'i-ma UPS}$

$S^2 = \sum_{i=1}^K \sum_{j=1}^{M_i} \frac{(Y_{ij} - \bar{Y})^2}{N-1} = \text{varianza USS}$

$S_i^2 = \sum_{j=1}^{M_i} \frac{(Y_{ij} - \bar{Y}_i)^2}{M_i-1} = \text{varianza dell'i-ma UPS}$

# NOTAZIONE CAMPIONE

$k$  = n° di UPS selezionate all'interno del mio campione

$m_i$  = n° USS selezionate nell' $i$ -ma UPS

$\bar{y}_i = \sum_{j \in S_i} \frac{y_{ij}}{m_i}$  = media campionaria nell' $i$ -ma UPS

$\hat{t}_i = \sum_{j \in S_i} \frac{M_i}{m_i} y_{ij}$  = stima del totale nell' $i$ -ma UPS

$\hat{t}_{unb} = \frac{K}{k} \sum_{i \in S} \hat{t}_i$  = stima non distorta del totale nella pop

$s_t^2 = \frac{1}{k-1} \sum_{i \in S} \left( \hat{t}_i - \frac{\hat{t}_{unb}}{K} \right)^2$  = stima della varianza del totale

$s_i^2 = \frac{1}{m_i-1} \sum_{j \in S_i} (y_{ij} - \bar{y}_i)^2$  = varianza campionaria nell' $i$ -ma UPS

# TEORIA E TECNICA DELL'INDAGINE STATISTICA E DEL CAMPIONAMENTO (MATR. DISPARI)

## CAMPIONAMENTO A STADI CAMPIONAMENTO A GRAPPOLI

MANUELA SCIONI

Dipartimento di Scienze Statistiche

[manuela.scioni@unipd.it](mailto:manuela.scioni@unipd.it)



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA



# CAMPIONAMENTO AD UNO STADIO (A GRAPPOLI)

- $M_i = m_i \rightarrow$  ha senso se non costa nulla rilevare tutte le USS in un'UPS (ma aumenta la correlazione intraclasse)
- Dal punto di vista delle stime è il caso più semplice:
  - trattiamo le medie e i totali di ogni UPS come  
se fossero osservazioni indipendenti e ignoriamo le singole USS
- È come lavorare con un campione (CCS o Stratificato) di  $k$  unità