

## Tipologie di campionamento

I fattori che intervengono sono due:



**Ruolo di chi forma il campione**

*La scelta delle unità avviene con un meccanismo consapevole*



**Attitudine delle unità a comparire nel campione**

*Le persone si limitano a costituire la frame e la selezione è lasciata a meccanismi di pura sorte*

## L'attitudine

E' misurata su di una scala a rapporti e con valori nell'intervallo (0,1)

Attitudine dell'unità  $U_j$  a comparire in posizione  $i$  - esima:  $a_{ij}$  con  $0 \leq a_{ij} \leq 1$

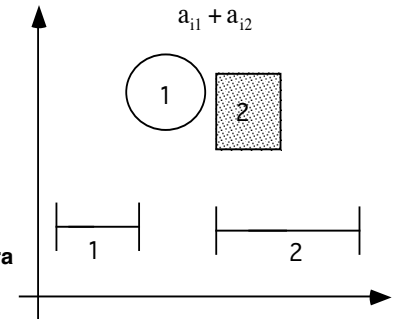
E' propria dell'unità e non condivisa con altre perciò è assimilabile ad una grandezza fisica estensiva: area, lunghezza, etc.

Tendenza dell'una o dell'altra:

Poiché per ogni posizione del campione si ha

$$\sum_{i=1}^N a_{ij} = 1$$

L'attitudine di una non può essere aumentata senza diminuire quella di un'altra



## L'attitudine/2



Si ha  $a_{ij}=0$  se l'unità "j" non ha alcuna possibilità di entrare nella posizione "i" del campione.

*Ciò succede quando l'unità è già inclusa e non è consentita la ripetizione oppure quando l'unità subisce la deliberata esclusione (cut-off)*



Si ha  $a_{ij}=1$  se l'unità "j" deve per forza entrare nella posizione "i" del campione.

*Si tratta di unità AUTORAPPRESENTATIVE delle quali un campione che sia rappresentativo non può non includere. Di solito non è consentita la ripetizione di tali unità*



Se l'attitudine -per la posizione "i" - è uguale per tutte le unità della popolazione avrà:

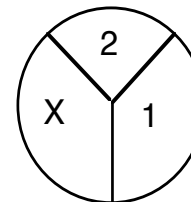
$$\sum_{j=1}^N a_{ij} = \sum_{j=1}^N a = Na = 1 \Rightarrow a = \frac{1}{N}$$

## Valori dell'attitudine

L'attitudine può essere legata ad una variabile continua o ad una discreta.

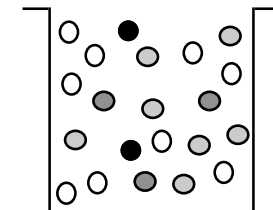
Può coincidere con una proprietà fisica delle unità (ad esempio la superficie per le unità areali) o una caratteristica attribuita dall'esterno (ad esempio distanza dal luogo dell'indagine)

Metodo per una variabile CONTINUA



Settori proporzionali alla frequenza attesa

Metodo per una variabile DENSA O DISCRETA



Biglie colorate in proporzione alla attitudine

## Agon et Alea



La selezione delle unità è frutto di due principi opposti:



**AGON:** forza attiva derivante dalla attitudine delle unità che competono per entrare nel campione



**ALEA:** forza passiva né giusta né meritocratica che favorisce e sfavorisce le unità allo stesso modo

I due principi interagiscono e non si oppongono nella formazione del campione purché l'attitudine sia correttamente misurata ed attribuita.

## Metodi di campionamento

Considerando i due fattori -persone e attitudini- le scelte possibili sono:

| RUOLO DELL'ATTITUDINE DELLE UNITA' |          |                                              |
|------------------------------------|----------|----------------------------------------------|
| RUOLO DI CHI FORMA IL CAMPIONE     | Naturale | Modificata                                   |
|                                    | Attivo   | Campioni quasi-casuali<br>Campioni Ragionati |
|                                    | Neutrale | Campioni casuali<br>Campioni di convenienza  |

Tutte le tipologie sono applicate.

Alcune sono più diffuse di altre

Se sono in contrapposizione si decide in base ai costi ed al contenimento degli errori campionari e non campionari.

## Campionamento ragionato

Si scelgono le unità in base alle opinioni di chi conduce l'indagine che individua le unità "tipiche" o "rappresentative" in base a certe variabili ed avarie considerazioni di opportunità.

*L'ipotesi è che se le unità sono rappresentative rispetto a queste lo saranno anche rispetto alle altre*

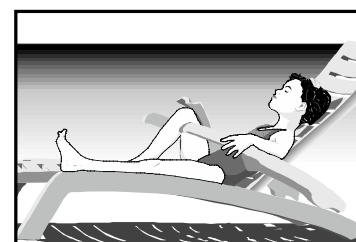
Lo studio di casi, molto utilizzato nelle discipline aziendalistiche, è un tipico esempio di campione ragionato.

La scelta di un paniere di beni e servizi per un indice dei prezzi ricalca tale metodologia

*E' una tecnica rischiosa perchè la formazione del campione deriva dal solo giudizio delle persone, ma in molte occasioni è l'unica strada percorribile*

## Esempi

La tipica casalinga italiana



La dieta delle top model più famose



Mezzo di trasporto più adoperato dagli studenti



## Tecniche di camp. ragionato



### Purposive sampling (campionamento mirato)

La scelta avviene in base a caratteristiche anche negative mostrata dalle unità.

*Produttori di un giocattolo interrogano i bambini che non appena l'hanno avuto lo hanno rotto*



### Focus group (gruppo di studio)

Intervista approfondita ad un gruppo di persone che conoscono aspetti diversi di uno stesso problema.

*Riscontrare nuovi bisogni che i clienti non sanno nemmeno di avere*

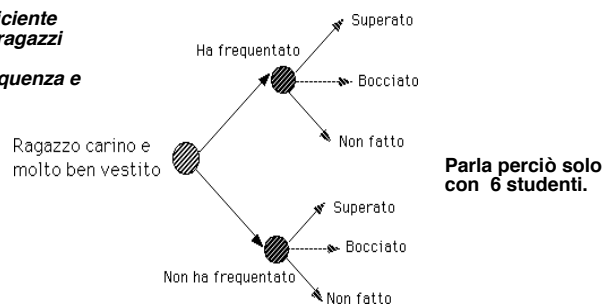
## Esempio

Letizia Aiello studia nella Facoltà di Economia ed è indecisa se frequentare o no un certo corso di statistica avanzata tenuto da un docente notoriamente severo.

Prima di effettuare chiede notizie a chi conosce il corso.

Poiché ha poco tempo da perdere (è molto impegnata!) vuole limitare al minimo il numero di persone con cui parlare. Come effettuare una buona scelta?

*La Aiello ritiene sufficiente interrogare solo bei ragazzi che vestono firmato distinguendo per frequenza e per esito dell'esame.*



Non è certo che il campione sia buono, ma forse ci scappa un invito a cena.

## Campionamento di convenienza

### Accessibility Sampling

il campione è formato includendo le unità della lista più facili o più rapide da raggiungere o più disposte ad essere raggiunte.

*L'idea è che un buon assortimento di unità dia luogo ad un campione rappresentativo*

*Esempi:*

*Clienti disposti a seguire nuove organizzazioni di vendita*

*Pazienti di uno studio medico che non contestano un cambio di sede*

il campione di convenienza è soggetto al rischio di sostituzione arbitraria delle unità e da "convenienza" indotta da particolari condizioni delle unità scelte

La rappresentatività di questi campioni è molto dubbia. Servono per INDAGINI PILOTA (pre-testing) preludio di indagini più complesse.

## Tecniche di camp. di convenienza



### Network sampling (campionamento a cascata)

Individuata una unità, le altre sono rilevate in base ad indicazioni fornite dalle altre già esaminate.

*Studio di gruppi etnici dispersi nella popolazione*



### Campioni autoselezionati

Le unità finiscono nel campione per loro deliberata volontà, ma sulle quali può essere esercitato un controllo selettivo

*Volontari per la sperimentazione di un farmaco*



### Campioni fortuiti

Le unità finiscono nel campione indipendentemente dalla volontà di chi forma il campione, ma anche per effetto di una attitudine sconosciuta

*Edifici risparmiati da una catastrofe*

## Estrazione con e senza reimmissione

Dopo che l'unità è stata selezionata ci sono due possibilità



**Campionamento con reimmissione:** le attitudini delle unità rimangono tutte inalterate dopo la selezione (anche quella della unità prescelta)

*Esempio: nella roulette esce il 12 e poi riesce il 12*

Campioni possibili:  $N^n$



**Campionamento senza reimmissione (o in blocco):** le attitudini delle unità si modificano (quella della unità prescelta è posta pari a zero).

*Esempio: sulla ruota di Genova esce il 12 al 1° estratto che non Potrà più uscire nelle successive estrazioni della ruota*

Campioni possibili:  $\binom{N}{n}$

La scelta della tecnica è questione di opportunità: se "N" è grande rispetto ad "n" la differenza si riduce.

## Revisione delle attitudini

Se l'estrazione avviene **SENZA** reimmissione occorre modificare le attitudini.

Se è uscita l'unità "j" con attitudine "a<sub>j</sub>" le attitudini delle unità rimaste è

$$\tilde{a}_i = \frac{a_i}{1 - a_j}; \quad i = 1, \dots, N; i \neq j$$

Si nota che ora la misura delle attitudini non può prescindere dall'ampiezza del campione:

Se n=2 una data unità può entrare in 1° o in 2° posizione e qui l'attitudine cambia secondo chi è entrata in prima posizione.

Si dimostra che:

$$\tilde{a}_i = a_i \left( A + \frac{1 - 2a_i}{1 - a_j} \right) \quad i = 1, 2, \dots, N; i \neq j; \quad \text{con } A = \sum_{j=1}^N \left( \frac{a_j}{1 - a_j} \right)$$

## Revisione delle attitudini/2

Campioni di ampiezza n=2

Nella tabella a destra sono riportate due diverse serie di attitudini per una popolazione di N=4 unità.

Nel primo caso le attitudini per il campione di ampiezza n=2 sono abbastanza simili a quelle per n=1, ma nel secondo sono invece molto diverse.

| i | Caso 1         |               | Caso 2         |               |
|---|----------------|---------------|----------------|---------------|
|   | a <sub>i</sub> | $\tilde{a}_i$ | a <sub>i</sub> | $\tilde{a}_i$ |
| 1 | 0.4            | 0.3579        | 0.70           | 0.4691        |
| 2 | 0.3            | 0.3042        | 0.25           | 0.2623        |
| 3 | 0.2            | 0.2206        | 0.05           | 0.0905        |
| 4 | 0.1            | 0.1173        | 0.10           | 0.1781        |

Per n>2 la revisione delle attitudini è questione difficile e gestita con metodi ingegnosi, ma complessi.

E' per questo che tale approccio, pur essendo quello più naturale, è il meno seguito e si presuppongono unità con eguale attitudini anche quando ciò è poco ragionevole

## Campionamento stratificato

Muovendo da alcune conoscenze a priori sulla popolazione si migliora l'efficienza del piano di campionamento riducendo l'arbitrio della Sorte

La popolazione di "N" unità è suddivisa in sottopopolazioni (strati) all'interno dei quali le unità siano omogenee.

A questo provvedono le variabili classificatorie

Da ogni strato si estrae un campione e si ottiene così il Campione Stratificato

|   |   |   |   |
|---|---|---|---|
| c | c | c | c |
| c | c | c | c |
| c | c | c | c |
| c | c | c | c |
| c | c | c | c |

La forte somiglianza tra le unità dello strato fa sì che, a livello di strato, il campione "rappresenti meglio" le caratteristiche della sottopopolazione.

*Dall'unione di sottocampioni singolarmente più rappresentativi si dovrebbe ottenere un campione complessivo pure più rappresentativo.*

## Esempio

Consideriamo un corso formato da 30 persone: 18 donne e 12 maschi con 12 fumatrici e 4 fumatori. Si deve ottenere un campione di ampiezza  $n=5$

|                                               |   |    |
|-----------------------------------------------|---|----|
| 1° variabile classificatoria: Sesso           | U | D  |
| 2° variabile classificatoria: fumo di tabacco | F | NF |
|                                               | 4 | 11 |
|                                               | 8 | 7  |

POPOLAZIONE

Per stratificare occorre precisare: numero, tipo e modalità delle variabili classificatorie che hanno legami significativi con le unità

La composizione della popolazione rispetto agli strati deve essere NOTA o almeno conosciuta con buona approssimazione.

Di solito si tratta di aspetti generali come sesso, professione, residenza, etc.

Se si usano "m" variabili classificatorie di cui ciascuna ha "k<sub>i</sub>" modalità, il numero degli strati possibili è

$$K = \prod_{i=1}^m K_i = k_1 * k_2 * \dots * k_m \quad K_i \geq 2$$

## Campionamento stratificato/2

indichiamo con  $N_i$  il numero di unità dello strato i-esimo e con  $n_i$  l'ampiezza del campione estratto dallo strato i-esimo.

Supponendo di aver formato "k" strati valgono le relazioni:  $\sum_{i=1}^k N_i = N$ ;  $\sum_{i=1}^k n_i = n$

Il numero totale dei campioni possibili è dato da

Con reimmissione  $\prod_{i=1}^K (N_i)^{n_i}$

Non è necessario utilizzare la stessa tecnica in tutti gli strati.

Senza reimmissione  $\prod_{i=1}^K \binom{N_i}{n_i}$

La forma più diffusa di scelta è il campione casuale semplice, ma non è obbligatoria

a)  $30^8 = 656,100,000,000$ ;  $\binom{30}{8} = 5,852,925$ ;

b)  $12^2 * 11^2 * 7^2 * 5^2 = 21,344,400$ ;  $\binom{12}{2} * \binom{11}{2} * \binom{7}{2} * \binom{5}{2} = 1,016,400$

## Scelta delle variabili classificatorie



Debbono essere immuni da legami funzionali

*E' inutile stratificare le automobili per cilindrata e cavalli fiscali*



Debbono essere indicatori disponibili, sicuri, stabili

*Caratteri geomorfologici, confini amministrativi, sesso, età professione*



Debbono essere collegate con la o le variabili da analizzare

*Paradossalmente, se dobbiamo spiegare la "X" la migliore stratificazione si ha utilizzando proprio la "X"*

## Stratificazione proporzionale

In che modo deve essere ripartita l'ampiezza "n" tra i vari strati?

Esistono tecniche statistiche che ottimizzano la spesa ed il rendimento del campione, ma risultano ancora ostiche a questo livello del corso.

Una tecnica più semplice è quella di scegliere le unità secondo l'importanza dello strato nella popolazione

Supponiamo di misurare l'importanza dello strato "S<sub>i</sub>" attribuendo un peso non negativo e con somma dei pesi unitaria

$$W_i \geq 0; \quad \sum_{i=1}^K W_i = 1$$

con una scelta

$$n_i = [W_i * n] + \delta_i$$

dove  $\delta_i$  è uguale ad uno solo per la frazione con il più alto resto

## Esempi

Peso in funzione della numerosità dello strato:

$$w_i = \frac{N_i}{N};$$



Peso in funzione dell'ammontare di una data variabile X

$$w_i = \frac{\sum_{j \in S_i} X_j}{\sum_{j=1} X_j};$$

POPOLAZIONE

|    | U | D  |
|----|---|----|
| F  | 4 | 11 |
| NF | 8 | 7  |

CAMPIONE

|    | U | D |
|----|---|---|
| F  | 1 | 3 |
| NF | 2 | 1 |

$$n_1 = \left[ \frac{4 \cdot 8}{30} \right] = 1 + 0.06 \quad n_3 = \left[ \frac{11 \cdot 8}{30} \right] = 2 + 0.93$$

$$n_2 = \left[ \frac{8 \cdot 8}{30} \right] = 2 + 0.13 \quad n_4 = \left[ \frac{7 \cdot 8}{30} \right] = 1 + 0.86$$

## Applicazione

Indagine nell'ambito del corso sul grado di conoscenza della Legge sul lavoro *part-time* degli studenti nelle strutture universitarie

Estrazione di un campione casuale stratificato di ampiezza n=100

|            | Sesso |        |       |
|------------|-------|--------|-------|
| Residenza  | Donne | Uomini |       |
| In sede    | 9500  | 7000   | 17500 |
| Fuori sede | 1300  | 2200   | 3500  |
|            | 10800 | 9200   | 20000 |

POPOLAZIONE

campione

|            | Donne | Uomini | Totale |
|------------|-------|--------|--------|
| In sede    | 48    | 35     | 83     |
| Fuori sede | 7     | 10     | 17     |
| TOTALE     | 55    | 45     | 100    |

RILEVAZIONE CAMPIONARIA

| Grado di conoscenza | frequenza |
|---------------------|-----------|
| Ottimo              | 9         |
| Medio               | 13        |
| Scarso o nullo      | 78        |
| TOTALE              | 100       |

## Selezione sistematica

Sia data una popolazione di "N" unità numerate da 1 a N dalla quale si deve estrarre un campione di ampiezza "n"

Si ricordi che l'intervallo (o passo) di campionamento è la frazione  $f=N/n$

Caso di "f" intero

Si estrae un numero casuale "c" tra 1 e f "c". La 1ª unità estratta è quella in posizione  $r=c+1$ . La 2ª è in posizione  $r+f$ , la 3ª in  $r+2f$ . In generale

L'estratta i-esima è in posizione  $r+(i-1)*f$

Ad esempio se  $N=80$  e  $n=8$  (quindi con  $f=10$ ) ed  $c=2$  le unità estratte sono quelle nelle posizioni: {3, 13, 23, 33, 43, 53, 63, 73}

*Da notare che il campionamento sistematico non va usato se il passo è legato all'ordine: se ogni 30 soldati c'è un sergente ed il passo è 30 il campione sarà formato da soli sergenti*

## Esempio

120 studenti di un corso sono elencati in ordine alfabetico e con numerazione progressiva. Si estrae un campione sistematico di 15 unità per partecipare ad un viaggio-studio

L'intervallo di campionamento è  $120/15=8$  unità. Scegliamo "r" nel blocco F-2:  $r=2$

le unità prescelte sono

2 10 18 26 34 42 50 58 66 74 82 90 98 106 114

|   |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |
|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|

E' ovvio che se "f" è pari ed "r" è pari non sarà mai estratta una posizione dispari per cui la allocazione delle unità sulla lista non deve avere sistematicità almeno a livello di pari e dispari

Non si deve per forza avviare il conteggio dalla prima. E' equivalente partire dalla posizione N-esima ovvero numerare le unità da 1 ad N, ma da destra verso sinistra.

## Caso di "f" non intero

Quando il passo risulta un numero frazionario ci sono varie possibilità

- Si può approssimare f all'intero inferiore (difetto) ma le unità finali non hanno probabilità di essere scelte.

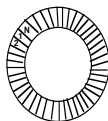
**Esempio:**  $N=120, n=13, f = \frac{120}{13} = 9.23 \approx 9$   
 $\{1, 10, 19, 28, 37, 46, 55, 64, 73, 82, 91, 100, 109\}$   
 $\{9, 18, 27, 36, 45, 54, 63, 72, 81, 90, 99, 108, 117\}$   
 le unità 118, 119, 120 non usciranno

- Si può approssimare all'intero superiore (eccesso), ma la numerosità del campione è inferiore a quella prefissata

**Esempio:**  $N=120, n=13, f = \frac{120}{13} = 9.23 \approx 10$   
 $\{1, 11, 21, 31, 41, 51, 61, 71, 81, 91, 101, 111, 121\}$   
 $\{9, 19, 29, 39, 49, 59, 69, 79, 89, 99, 109, 119, 129\}$

Per risolvere il problema si può selezionare a caso una unità tra tutte quelle non uscite e colmare la lacuna

Si può anche pensare ad una disposizione circolare



## Campionamento a quota

La popolazione è suddivisa in sottogruppi omogenei: secondo il sesso, la residenza, l'età, il reddito, istruzione, occupazione, etc.

Dai dati censuari o da altra fonte ufficiale o comunque pubblica si ricava la percentuale nella popolazione di ogni gruppo.

L'ampiezza del campione è suddivisa proporzionalmente in quote, una per ogni gruppo.

Ad ogni intervistatore è assegnata una parte o l'intera quota e la scelta delle unità, purché ricadenti nel gruppo, è del tutto libera.

Il rischio è che si scelgano le unità più raggiungibili (campionamento di convenienza) trascurando le altre rischiando di "infioccare" il campione.



## Esempio

La dott.ssa Anna Bartelli, disoccupata, collabora alla segreteria regionale di un partito di nuova fondazione.

il coordinamento nazionale vuole sapere rapidamente e con una spesa irrisoria che atteggiamento hanno i calabresi rispetto alle riduzioni IRPEF per le persone anziane.

Come si dovrà regolare?

Dall'ultimo censimento ricava la tabella regionale doppia su classi di età e condizione professionale e la rielabora ottenendo ad esempio:

|       | Lav. Aut. | Lav. Dip. | Disocc. | Pens. | Cas. Std. |
|-------|-----------|-----------|---------|-------|-----------|
| 18-30 |           |           |         |       |           |
| 30-40 |           |           |         |       |           |
| 40-50 |           |           |         |       |           |
| 50-60 |           |           |         |       |           |
| 61-65 |           |           |         |       |           |
| >65   |           |           |         |       |           |

Stabilito che di ogni cella deve interrogare l'1% si mette al lavoro con un breve questionario.

Inserirà: genitori, parenti, inquilini, amici, conoscenti, colleghi, fornitori e chiunque incontra e si mostra disponibile a riempire un questionario.

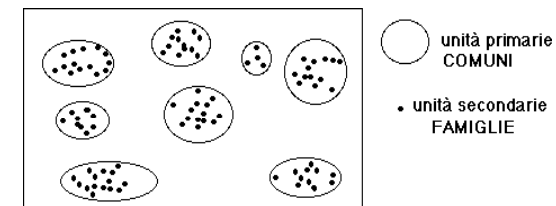
L'unico vincolo è di completare le interviste in ogni casella. Prima finisce e prima dà i risultati richiesti.

Una buona idea è di far completare ogni casella da una persona diversa

## Campionamento a 2 stadi

Se la popolazione può essere collocata su vari livelli gerarchici il piano di campionamento si può realizzare seguendo i vari livelli.

Esempio: l'indagine ISTAT sulla forza lavoro forma il campione selezionando prima i comuni e poi all'interno dei comuni, le varie famiglie



il piano di campionamento a 2 stadi procede scegliendo un campione di unità primarie (dette *cluster* o grappoli). All'interno di ogni primaria sceglie un campione di unità secondarie.

Se si considerassero TUTTE le unità dei grappoli scelti si avrebbe il piano detto campionamento casuale a grappoli (*cluster sampling*)

## Campionamento a più stadi/2

Supponiamo che la popolazione sia organizzata su 2 livelli gerarchici e contenga "m" grappoli con numerosità

$$N_i \quad i = 1, 2, \dots, m$$

Si devono scegliere "k" grappoli tra gli "m".

La scelta dei grappoli può avvenire nei modi già studiati:

**CASUALE SEMPLICE - SISTEMATICO - STRATIFICATO**

La scelta ideale sarebbe in base all'attitudine o importanza, ma se avviene.....



Con reimmissione può capitare di esaminare più volte lo stesso cluster



senza reimmissione ci sono problemi per rivedere e calcolare le attitudini

## Selezione dei cluster

Una tecnica che dà buoni risultati è la seguente:

Si definisce il livello cumulato di attitudine  $A_i = \sum_{j=1}^i a_j$ ;  $i = 1, 2, \dots, m$

Si assegna ad ogni cluster un intervallo frazionario

Si estrae una frazione casuale "q" di 4 cifre decimali dalla tavola premettendo "0." e si calcola

$$A_{i-1} \leq qA_m \leq A_i$$

L'unità per cui la disequazione è vera è la prima prescelta. Le altre seguono questa equispartizione di  $[m/k]$  unità.

| unità | attitudine | $A_i$                    | Scelta           |
|-------|------------|--------------------------|------------------|
| 1     | $a_1$      | $A_1 = a_1$              | $[0, A_1]$       |
| 2     | $a_2$      | $A_2 = a_1 + a_2$        | $[A_1, A_2]$     |
| 3     | $a_3$      | $A_3 = a_1 + a_2 + a_3$  | $[A_2, A_3]$     |
| :     | :          | :                        | :                |
| N     | $a_n$      | $A_m = \sum_{i=1}^N a_i$ | $[A_{m-1}, A_m]$ |

## Esempio

m=8 cluster

k=3 unità di 1° livello nel campione

Frazione casuale:  $q=0.9462$

$$0.9462 \cdot 1825 = 1726.8$$

Quindi la prima scelta è Mordano.

Frazione di campionamento:

$$\frac{8}{3} = 2.6 \approx 2$$

Supponendo una organizzazione circolare la scelta ricade su Caslaflumanese e su Castel Guelfo.

Con un campione one-stage il campione finale sarà formato da

$$166 + 106 + 176 = 448 \text{ famiglie}$$

| Comuni           | Famiglie | $A_i$ | scelta         |
|------------------|----------|-------|----------------|
| Borgo Tossignano | 60       | 60    | $[0, 60]$      |
| Caslaflumanese   | 106      | 166   | $[61, 166]$    |
| Castel del Rio   | 61       | 227   | $[167, 227]$   |
| Castel Guelfo    | 176      | 403   | $[228, 403]$   |
| Dozza            | 97       | 500   | $[404, 500]$   |
| Fontanelice      | 80       | 580   | $[501, 580]$   |
| Imola            | 1079     | 1659  | $[581, 1659]$  |
| Mordano          | 166      | 1825  | $[1660, 1825]$ |

## Campionamento a più stadi/3

L'ampiezza del campione è variabile poiché dipende dal numero di unità incluse nelle unità primarie e poi finite nel campione

$$n = \sum_{i=1}^h n_i$$

dove  $n_i$  è il numero di unità del grappolo "i" scelte al 2° stadio.

i rapporti  $\frac{n_i}{N_i}$  sono detti FRAZIONI DI SONDAGGIO e sono determinate con opportune tecniche

Spesso si adotta la stessa percentuale per ogni grappolo

$$\frac{n_1}{N_1} = \frac{n_2}{N_2} = \dots = \frac{n_h}{N_h} = \alpha\%$$



## Esempio

Indagine sul numero di corsi seguiti dagli iscritti a Statistica. Si ritiene la popolazione suddivisa in due livelli:

I) Anno di corso; II) Sesso

| Grappoli    | Denominazione | Popolazione nel corso |
|-------------|---------------|-----------------------|
| 1           | 2° Anno       |                       |
| 2           | 3° Anno       |                       |
| 3           | 4° Anno       |                       |
| 4           | 1° FC         |                       |
| 5           | >1 FC         |                       |
| TOTALE POP. |               |                       |

Ci sono 5 unità di I livello; si decide di estrarre un campione di ampiezza 3

Al livello I sono possibili

$$\binom{5}{3} = 10$$

campioni di unità primarie cioè di anni di corso

## Esempio/continua

Estrazione senza reimmissione del grappolo: (3,4,1)

il campione sarà formato da tutti o parte degli iscritti al 4°, al 1°FC ed al 2°

Si stabilisce una frazione di sondaggio del 20%

| Grappoli     | Denominazione | Elementi nel campione (20% pop.) |
|--------------|---------------|----------------------------------|
| 3            | 4° Anno       |                                  |
| 4            | 1° FC         |                                  |
| 1            | 1° Anno       |                                  |
| TOTALE CAMP. |               |                                  |

Non rimane poi che scegliere le unità nei vari grappoli

## Doppio campionamento

In caso di:

- ☐ stratificazione per variabili poco note
- ☐ Frame incomplete
- ☐ Analisi settorializzate

Si ricorre al doppio campionamento (double sampling o two-phase sampling)

Si effettua una selezione preliminare (screening) delle unità per variabili molto ampie e poco costose da esaminare.

Solo a questo punto si applica una vera e propria tecnica di campionamento

Esempio

Per ridurre noncoverages e nonresponses si seleziona un campione di unità così configurate e si produce lo sforzo massimo per ottenere le risposte.

## Campionamento per confronti a coppia

La scelta delle unità di 2° livello nelle campionamento a due stadi può avvenire in un modo ragionato:

- 1) Si identifica una variabile dimensionale che misura l'importanza ai fini della rappresentatività e che sia presente in modo univoco nelle unità
- 2) Si dispongono le unità in ordine crescente di grandezza secondo tale variabile dimensionale.
- 3) Per formare il campione si scelgono la prima e l'ultima unità, la seconda e la penultima fino a quella (o quelle se sono in numero pari) del centro dello ordinamento

Esempio:

Nel verificare l'età media di una popolazione si sceglie al primo stadio un campione di famiglie e per il secondo stadio finiscono nel campione la più anziana e la più giovane

