

PIANO NAZIONALE
LAUREE SCIENTIFICHE

2016-2017



Tecniche campionarie

Prof. Francesco D. d'Ovidio



UNIVERSITÀ
DEGLI STUDI DI BARI
ALDO MORO

VANTAGGI e SVANTAGGI

della raccolta di informazioni tramite campioni

VANTAGGI

- riduzione dei costi
- maggiore rapidità
- maggiore accuratezza
- assenza di alternative

SVANTAGGI

- errore campionario
- discriminazioni all'interno della popolazione
- eccessiva dimensione del campione in caso di eventi rari

Le tecniche statistiche

Le **tecniche statistiche** sono quegli strumenti che consentono di:

- acquisire in modo razionale, ordinato ed efficiente informazioni su un collettivo attraverso il prelievo e la misura di campioni,
- elaborare tali informazioni, valutandone il grado di precisione,
- effettuare **stime** circa le caratteristiche del sistema dal quale il **campione** è stato estratto, sui fattori coinvolti, sui parametri e su altre grandezze che consentono di conoscere meglio il sistema stesso e migliorarlo.

Il campionamento

- Si definisce *campionamento* un procedimento attraverso il quale da un insieme di unità (detto “*collettivo*”, “*popolazione*” o “*universo*”), che costituiscono l’oggetto dello studio, si estrae un numero ridotto di casi scelti con criteri (le *tecniche di campionamento*) tali da consentire la generalizzazione dei risultati ottenuti all’intera popolazione.
- Le unità estratte costituiscono il *campione*.

Proprietà del campionamento

Rappresentatività: è la capacità di fornire, in piccolo ma senza distorsioni, un'immagine accurata della popolazione a cui si riferisce.

Sufficienza: misura l'attendibilità dei dati, ossia la probabilità che essi siano validi per l'universo entro termini statisticamente determinabili.

Eterogeneità: presenza di caratteristiche e qualità diverse nel campione, collegate alla variabilità dei comportamenti.

Teoria del campionamento

Definizione della **popolazione** (universo)

↳ identifica popolazione obiettivo (**target population**)

popolazione d'indagine (**survey / study population**)

N
↓
 n

- ◆ insieme delle unità
- ◆ lista delle unità

disegno di campionamento

insieme delle **regole** per la corretta
formazione del campione

probabilistico

non probabilistico

Il disegno di campionamento

Il disegno di campionamento è l'insieme delle decisioni prese per formare il campione.

Le sue fasi sono:

1. definizione della struttura del campione (*richiede la conoscenza della lista delle unità che compongono l'universo che si intende osservare, o di un adeguato equivalente*)
2. selezione delle unità campionarie
3. probabilità di inclusione delle singole unità
4. determinazione della numerosità del campione

Le tecniche di campionamento

Le tecniche di campionamento consentono di selezionare alcune unità statistiche del collettivo secondo criteri prefissati.

CAMPIONAMENTO PROBABILISTICO (O CASUALE):

parte da una probabilità nota, o calcolabile, di ogni unità della popolazione di entrare a far parte del campione.

CAMPIONAMENTO NON PROBABILISTICO:

quando non è nota, né è ricavabile, la probabilità di inclusione di una unità nel campione.

N.B.: solo i campionamenti probabilistici assicurano la possibilità di conoscere la misura dell'impresione che si commette inevitabilmente nel selezionare le unità.

CAMPIONAMENTO PROBABILISTICO

- ✓ Campionamento casuale semplice
*Campionamento bernoulliano (con reinserimento),
Campionamento senza reinserimento, Campion. in blocco,
Campion. sistematico, Campion. con probabilità diseguali*
- ✓ Campionamento casuale stratificato
Campion. stratif. proporzionale, Campion. stratif. uniforme
- ✓ Campionamento a due o più stadi
- ✓ Campionamenti per gruppi (“cluster”)
Campionamento a grappoli, Campionamento aereo
- ✓ Campionamento ruotato

CAMPIONAMENTO NON PROBABILISTICO

- Campionamento per quote
- Campionamento a scelta ragionata
- Campionamento con adesione volontaria
- Campionamento bilanciato
- Campionamento “snowball”.

Storicamente, campioni di questo tipo sono stati concepiti e applicati prima quelli probabilistici, ma non sono altrettanto rappresentativi; inoltre non consentono il calcolo dell'errore dovuto al campionamento e quello della bontà delle stime.

Il campionamento probabilistico

Le unità sono scelte in modo casuale, utilizzando in modo appropriato le diverse tecniche di selezione.

La casualità interviene nella *selezione delle unità e si ottiene attribuendo ad ogni unità della popolazione una probabilità nota (e diversa da zero)* di essere selezionata.

Quando la probabilità di estrazione, oltre ad essere nota, in ciascuna estrazione è posta *uguale per tutte le unità della popolazione*, si parla di *campionamento casuale semplice*.

Una prima riflessione

Campione casuale a E' un campione estratto da una popolazione in cui tutte le unità hanno probabilità non nulla di essere estratte.

Un campione è **rappresentativo quando è estratto in modo casuale** (e **non semplicemente quando è grande**)

È molto meglio avere un campione piccolo, ma estratto in modo casuale, che un campione grande ma estratto “a casaccio”.

Il campionamento casuale semplice

Avendo determinato un'ampiezza campionaria pari a n , il c.c.s. corrisponde concettualmente a inserire in un'urna (o in un computer) i primi N numeri interi associati alle N unità della popolazione, e quindi estrarre casualmente dall'urna n volte uno di questi numeri (o dal computer tramite la generazione di n numeri aleatori), riportandolo in una apposita lista di campionamento e poi reinsertendolo nell'urna. La lista dei numeri identificherà dunque le unità campionate.

Questo metodo è anche detto campionamento **bernoulliano** (dal nome del matematico Bernoulli, che lo formalizzò), oppure **campionamento con reinserimento** (con ripetizione).

Il rapporto n/N è detto *frazione di campionamento* (cioè “quota del collettivo da inserire nel campione”).

- Nel campionamento casuale semplice tutte le unità della popolazione hanno uguale probabilità di essere estratte.
- La probabilità (a priori) che una unità sia scelta in ogni estrazione è pari a $1/N$, per cui la probabilità complessiva che tale unità faccia parte del campione è pari a n/N (pari alla frazione di campionamento).

- Lo *spazio campionario*, indicato con il simbolo $|\Omega|$ (ossia il numero di campioni estraibili), nel caso di campionamento casuale semplice è dato dal prodotto di n volte la numerosità N della popolazione, cioè N^n .

Per es., avendosi sempre N campioni diversi con 1 solo elemento, i campioni di 2 elementi sono costruiti associando a ciascuno di essi tutti gli N elementi della popolazione, ottenendo N^2 campioni.

I campioni di 3 elementi, a loro volta, sono ancora costruiti associando tutti gli N elementi del collettivo a ciascuno dei campioni di 2 elementi, ottenendo in tal modo N^3 campioni.

E così via...

**Supponiamo di avere una popolazione di 6 unità:
quanti e quali campioni di dimensione 2 possono
essere estratti con il C.C.S. ?**

Indicando ciascuna unità della popolazione con lettere da A a F, i possibili campioni (con reinserimento) sono dati dalle coppie:

AA AB AC AD AE AF BA BB BC BD BE BF CA CB CC CD CE CF
DA DB DC DD DE DF EA EB EC ED EE EF FA FB FC FD FE FF

I campioni possibili sono dunque 36 (6^2), e ognuno di essi ha una probabilità di essere estratto pari, ovviamente, a $1/36$.

Ogni unità della popolazione ha la stessa probabilità di essere estratta. Nell'esempio, ogni unità appare in 12 diversi campioni, per cui la probabilità che un soggetto sia estratto è $12/36=1/3$, ancora una volta pari alla frazione di campionamento $2/6=1/3$.

MA...

“Il campionamento casuale semplice è raramente applicato nelle indagini statistiche, sia perché la selezione è completamente affidata al caso e *non considera le informazioni note a priori* sulla popolazione, sia perché nelle indagini su vasta scala comporta un piano di rilevazione costoso e di difficile realizzazione dal punto di vista organizzativo, *necessitando inoltre della lista completa della popolazione*, che spesso non è disponibile”
(Corbetta, 1999)

- Il campionamento casuale semplice è **inapplicabile** nelle indagini “*distruttive*” e **distorsivo** nelle indagini svolte in popolazioni poco numerose, poiché è alta la probabilità di campionare più volte la stessa unità.

Il campionamento senza reinserimento

Questa tecnica differisce dalla precedente perché il numero estratto e inserito nella lista NON viene reinserito nell'urna (nel caso di lista su computer, si farà in modo di cancellare il numero estratto o di inibirne una nuova estrazione da parte del generatore di numeri casuali).

Di conseguenza, ogni unità della popolazione associata a tali numeri avrà $1/N$ probabilità di essere campionata alla prima estrazione, $1/(N-1)$ alla seconda, $1/(N-2)$ alla terza e così via, fino a una probabilità di $1/(N-n+1)$ alla n.esima estrazione.

La probabilità totale “a priori” che una unità faccia parte del campione sarà quindi pari a

$$\frac{1}{N} + \frac{1}{N-1} + \dots + \frac{1}{N-n+1} = \frac{N + (N-1) + \dots + (N-n+1)}{N \cdot (N-1) \cdot \dots \cdot (N-n+1)} \geq \frac{n}{N}$$

Si noti che tale probabilità sarà pari a n/N solo nel caso $n=1$, essendo qui $(N-n+1) = N$, per cui l'espressione si ferma a $1/N$.

Lo spazio campionario $|\Omega|$ (ossia il numero di possibili campioni che possono essere estratti dalla popolazione data) è fornito, dunque, dal prodotto posto a denominatore della frazione che indica la probabilità totale “a priori”: $N \cdot (N-1) \cdot (N-2) \cdot \dots \cdot (N-n+1)$.

Una formula generale che utilizza una funzione matematica nota (anche nei fogli elettronici) si ottiene moltiplicando e dividendo proprio il prodotto $N \cdot (N-1) \cdot (N-2) \cdot \dots \cdot (N-n+1)$ per la quantità $(N-n)!$, che si legge *(N-n) fattoriale*, e che corrisponde al prodotto dei primi (N-n) numeri interi.

Si avrà:

$$|\Omega| = (N)! / (N-n)!,$$

mentre la probabilità di estrazione di un qualsiasi campione è data dall'inverso:

$$\pi = (N-n)! / (N)! .$$

Ad esempio, quanti e quali campioni di dimensione 2 possono essere estratti con campionamento senza reinserimento da una popolazione di 6 unità?

Indicando sempre ciascuna unità con lettere da A a F, tutti i possibili campioni senza reinserimento sono dati dalle coppie:

AB AC AD AE AF BA BC BD BE BF CA CB CD CE CF
DA DB DC DE DF EA EB EC ED EF FA FB FC FD FE

I campioni possibili sono evidentemente 30 e ognuno di essi ha probabilità di essere estratto pari, ovviamente, a $\pi = 1/30$.

Ma tale conclusione equivale a dividere tra loro il fattoriale di 6 (pari a $6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 720$) e quello di $6-2=4$ (ossia $4 \cdot 3 \cdot 2 \cdot 1 = 24$): infatti $|\Omega| = 720/24 = 30$; reciprocamente, $\pi = 24/720 = 1/30$.

E quanti campioni di dimensione 3 possono essere estratti con lo stesso metodo da un collettivo di 6 unità?

Indicando sempre ciascuna unità con lettere da A a F, tutti i possibili campioni senza reinserimento sono dati dalle terne:

ABC ABD ABE ABF ACB ACD ACE ACF ADB ADC ADE ADF AEB
AEC AED AEF AFB AFC AFD AFE BAC BAD BAE BAF BCA BCD
BCE BCF BDA BDC BDE BDF BEA BEC BED BEF BFA BFC BFD BFE
CAB CAD CAE CAF CBA CBD CBE CBF CDA CDB CDE CDF CEA
CEB CED CEF CFA CFB CFD CFE DAB DAC DAE DAF DBA DBC
DBE DBF DCA DCB DCE DCF DEA DEB DEC DEF DFA DFB DFC
DFE EAB EAC EAD EAF EBA EBC EBD EBF ECA ECB ECD ECF EDA
EDB EDC EDF EFA EFB EFC EFD FAB FAC FAD FAE FBA FBC FBD
FBE FCA FCB FCD FCE FDA FDB FDC FDE FEA FEB FEC FED

Ossia 120 possibili campioni, con probabilità di estrazione $1/120$.

Ciò corrisponde a dividere il fattoriale di 6 ($6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 720$) per il fattoriale di $6-3$ ($3 \cdot 2 \cdot 1 = 6$): ossia $|\Omega| = 720/6 = 120$.

Il campionamento in blocco

Gli n elementi numerici vengono estratti dall'urna (o dalla lista nel computer) tutti insieme: appunto **in blocco**.

Ciò implica che, come nel campionamento senza reinserimento, **una singola unità non può essere presente nel campione più di una volta**.

Però, a differenza della precedente tecnica (che forma campioni ordinati), qui non ha senso distinguere il campione $\{a, b\}$ dal campione $\{b, a\}$, perché non esiste un ordine di estrazione, né una differente probabilità di estrazione dipendente da tale ordine.

Di conseguenza, ogni unità della popolazione associata a tali numeri avrà, a priori, n/N probabilità di essere estratta e, conseguentemente, di far parte del campione.

Ma... attenzione: quanti campioni di dimensione 2 possono essere estratti con campionamento IN BLOCCO da una popolazione di 6 unità?

Indicando sempre ciascuna unità con lettere da A a F, tutti i possibili campioni in blocco sono dati dalle coppie:

AB AC AD AE AF BC BD BE BF CD CE CF DE DF EF

I campioni possibili sono 15 (la metà esatta di quelli possibili con il campionamento senza reinserimento) e ognuno di essi ha probabilità di essere estratto pari a $\pi = 1/15$.

Tale conclusione equivale a dividere tra loro il fattoriale di 6 (ossia $6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 720$) e il doppio del fattoriale di $6-2$, cioè $2 \cdot (4 \cdot 3 \cdot 2 \cdot 1) = 48$, anche leggibile come $(2 \cdot 1) \cdot (4 \cdot 3 \cdot 2 \cdot 1) = 48$, da cui, appunto, $|\Omega| = 720/48 = 15$ e $\pi = 48/720 = 1/15$.

E quanti campioni di dimensione 3 possono essere estratti con lo stesso metodo da un collettivo di 6 unità?

Indicando sempre ciascuna unità con lettere da A a F, tutti i possibili campioni in blocco sono dati dalle terne:

ABC ABD ABE ABF ACD ACE ACF ADE ADF AEF
BCD BCE BCF BDE BDF BEF CDE CDF CEF DEF

Cioè 20 possibili campioni, con probabilità di estrazione $1/20$.

Ciò equivale a dividere il fattoriale di 6 (cioè $6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 720$) per il **sestuplo** del fattoriale di $6-3=3$, il quale equivale però al **prodotto del fattoriale di 3 per quello di $6-3$** : $(3 \cdot 2 \cdot 1) \cdot (3 \cdot 2 \cdot 1) = 36$: da cui $|\Omega| = 720/36 = 20$ e, reciprocamente, $\pi = 36/720 = 1/20$.

Generalizzando, $|\Omega| = N!/[n!(N-n)!]$ e $\pi = n!(N-n)!/N!$

Il campionamento sistematico

Questo metodo può essere considerato un caso particolare del campionamento in blocco, ove, però, le unità non vengono estratte **tutte** mediante sorteggio.

Date **N** unità della popolazione, presumibilmente elencate in maniera casuale in un' unica lista numerata (si escludono, quindi, le liste ordinate in base a caratteristiche importanti, come, ad es., il reddito), se si vuole estrarre un campione di numerosità **n** da tale sequenza casuale di unità **si determina innanzitutto il passo di campionamento $k = N/n$** (reciproco della frazione di campionamento).

Si estrae casualmente un numero j compreso fra 1 e k , che identifica la prima unità da selezionare,; a partire dalla j -esima unità, si selezionano poi le unità successive in base al passo di campionamento: una ogni k unità della lista.

Probabilità di estrazione “a priori” e spazio campionario si calcolano in modo analogo al campionamento in blocco.

- ✓ Il campionamento sistematico consente di ottenere campioni casuali anche nella situazione in cui manchi la lista della popolazione e N sia ignoto ma stimabile in base a dati precedenti: per es., sapendo che un supermercato ha circa 1800 clienti al giorno (giornata di 10 ore continuate: 3 ogni minuto), volendo ottenere un campione rappresentativo di 200 clienti ($k=1800/200=9$) se ne intervista uno ogni 9, o, in alternativa, uno ogni 3 minuti.
- ✓ Deve comunque essere evitata ogni selezione diversa da quella predeterminata dal passo di campionamento.

Il campionamento con probabilità diseguali

Questo metodo di campionamento si applica, soprattutto in caso di campionamenti senza reinserimento, se il collettivo di riferimento non è grande e se è suddiviso in unità statistiche che hanno (o dovrebbero avere ai fini dell'indagine) differenti probabilità di essere comprese nel campione.

Ad esempio, quando si vuole estrarre un campione casuale di comuni per uno studio sulla produzione di rifiuti urbani, sarebbe opportuno che i capoluoghi di provincia, che sono presumibilmente maggiori produttori, abbiano una maggiore probabilità di essere estratti rispetto a comuni meno popolosi.

Ponendo p_i la probabilità di inclusione di ogni i .esima unità ($i=1, 2, \dots, N$) del collettivo, ovviamente diversa da quella delle altre unità, identificata la probabilità più piccola (che immaginiamo qui essere la probabilità p_j associata alla j .esima unità) e trovato il minor numero intero K tale che $Kp_j \geq 1$, si identifichino gli interi k_i più prossimi a ciascun valore Kp_i , con la condizione che $k_1+k_2+\dots+k_N = K$.

A questo punto, si inseriscono nell'urna (o nella lista sul computer) k_1 numeri "1", k_2 numeri "2" e via dicendo (fino a k_N numeri "N"), per poi estrarre normalmente, con uno dei metodi precedentemente descritti, n ($<N$) numeri a cui sono associate altrettante unità della popolazione.

Con campionamento senza reinserimento, logicamente, sarà necessario eliminare i duplicati via via estratti, poi reiterando l'estrazione in modo da ottenere n unità diverse.

Esempio:

Volendo estrarre $n=5$ comuni dall'insieme dei 10 comuni della provincia BAT, si ha la seguente distribuzione delle popolazioni comunali (prime colonne della tabella):

ponendo $N=58$ (poiché per Spinazzola $p_i=0,0172$, che moltiplicato per 58 dà 1,00017) si imbussolano 21 palline col n.1, 3 col n.2, ecc., si mescolano tutte le palline e si estraggono, poi, 5 di esse senza ripetizione (escludendo, ovviamente, i comuni che dovessero essere stati già estratti).

Esempio:

Volendo estrarre $n=5$ comuni dall'insieme dei 10 comuni della provincia BAT, si ha la seguente distribuzione delle popolazioni comunali:

n	Comuni	Popolazione censim. 2011	Distrib. % p_i	Kp_i	k_i
1	Andria	100.052	25,54%	14,81	15
2	Barletta	94.239	24,06%	13,95	14
3	Bisceglie	54.678	13,96%	8,10	8
4	Canosa di Puglia	30.422	7,77%	4,50	5
5	Margherita di Savoia	12.193	3,11%	1,81	2
6	Minervino Murge	9.333	2,38%	1,38	1
7	San Ferdinando di Puglia	13.916	3,55%	2,06	2
8	Spinazzola	6.755	1,72%	1,00	1
9	Trani	55.842	14,26%	8,27	8
10	Trinitapoli	14.293	3,65%	2,12	2
	Provincia BAT	391.723	100,00%	<i>totale k_i</i>	58

ponendo $N=58$ (si ha per Spinazzola $p_i=0,0172$, che moltiplicato per 58 dà 1,00017) si trovano gli interi k_i e poi si inseriscono nella lista (o nell'urna) 15 n."1", 14 n. "2", ecc.; si estraggono, poi, 5 numeri senza ripetizione (non contando, ovviamente, i numeri che dovessero essere stati già estratti).

Il campionamento stratificato

Con popolazione distinguibile in gruppi omogenei al loro interno, soprattutto rispetto all'indagine da svolgere, per aumentare l'efficienza del campione l'estrazione casuale non è effettuata sull'intera popolazione ma all'interno dei gruppi, come se si trattasse di tanti collettivi distinti da cui estrarre altrettanti campioni casuali.

- a) Suddividere la popolazione in sottopopolazioni (*strati*) il più possibile omogenee rispetto all'indagine, utilizzando una o più variabili ad essa correlate (ad es., per indagini sul reddito, si stratifica per regione di residenza, classe di età e professione).
- b) Estrarre da ogni strato un campione casuale (con o senza reinserimento, a blocchi o sistematico).
- c) Unire i campioni dei singoli strati per ottenere il campione totale.

Quando si stratifica

La stratificazione si usa quando si vuole:

- evidenziare insiemi di unità significative per la ricerca;
- separare sottopopolazioni con caratteristiche speciali;
- utilizzare informazioni note, mantenendo tuttavia la casualità dell'estrazione;
- e, soprattutto, individuare sottopopolazioni omogenee rispetto alla variabile da studiare, ottenendo così stime più efficienti di quelle che si possono avere con un campione casuale semplice.

- A parità di numerosità, infatti, il campionamento stratificato assicura un minore errore di campionamento rispetto al c.c.s. (maggiore efficienza).
- Il campione stratificato può essere **proporzionale** o **non proporzionale**. Un campione proporzionale riproduce la stessa composizione degli strati della popolazione.

Se operai, impiegati, lavoratori autonomi e liberi professionisti sono rispettivamente il 35, 45, 15 e 5% della popolazione, si potrà costruire un campione di $n=1000$ casi selezionando **350 operai, 450 impiegati, 150 lavoratori autonomi e 50 liberi professionisti.**

Per determinare quante unità campionare in ogni strato, ci si può basare sulle frequenze relative $f_1=N_1/N, f_2=N_2/N, \dots, f_s=N_s/N$ degli s strati del collettivo: **$n_1 = n \cdot f_1, n_2 = n \cdot f_2, \dots, n_s = n \cdot f_s,$** oppure calcolarle moltiplicando l'ammontare degli strati per la frazione di campionamento $k=n/N$: **$n_1 = k \cdot N_1, \dots, n_s = k \cdot N_s .$**

- Un campione stratificato non proporzionale invece è indipendente del tutto o in parte dalla composizione degli strati nella popolazione.
- Del tutto indipendente dalla composizione degli strati nel collettivo è il **campione stratificato uniforme**, in cui occorre solo conoscere la numerosità del campione da estrarre e il numero degli strati: le unità da campionare in ogni strato saranno semplicemente:

$$n_1 = n/s , \quad n_2 = n/s , \quad \dots , \quad n_s = n/s .$$

- Si usa il campionamento stratificato uniforme quando vi sono alcuni strati della popolazione che fanno rilevare **numerosità piccole**, così che il normale campionamento proporzionale fornirebbe sub-campioni con numerosità troppo esigua per trarre statistiche affidabili.

- Il problema del campionamento stratificato uniforme è ovviamente la **sovrastima delle statistiche** tratte dagli strati che nella popolazione hanno il minor peso: ad esempio, se lo scopo d'indagine è quello di stimare il reddito medio della popolazione, nell'esempio precedente l'utilizzo di questa tecnica di campionamento produrrebbe **strati tutti uguali di 250 unità**: in tal modo il reddito dei liberi professionisti concorrerebbe al calcolo del reddito medio in misura **5 volte maggiore** di quanto avviene nella popolazione.
- La soluzione di questo problema consiste (dopo l'indagine) nel moltiplicare le stime ottenute in ogni strato per i rispettivi «**coefficienti di espansione**» $k_1 = k \cdot N_1 / n_1, \dots, k_s = k \cdot N_s / n_s$ (ove k è ancora la frazione di campionamento n/N).
Nell'esempio, i coefficienti di espansione sarebbero: $k_1 = 350/250 = 1,4$ per gli operai, $k_2 = 450/250 = 1,8$ per gli impiegati, $k_3 = 150/250 = 0,6$ per i lavoratori autonomi, $k_4 = 50/250 = 0,2$ per i liberi professionisti.

- Solo parzialmente indipendente dalla composizione degli strati nella popolazione è invece il **campione stratificato quasi-proporzionale**, che nasce quando si vuole evitare di avere sub-campioni numericamente esigui, ma si vuol mantenere la maggiore importanza degli strati più numerosi (tipicamente, si usa quando solo una minima parte dei sub-campioni risulterebbe inadeguata).
- Questo campionamento richiede di fissare la minima numerosità **m** accettabile per uno strato del campione, effettuare la stratificazione proporzionale e, se uno o più sub-campioni risultano numericamente inferiori al minimo, portarne l'ampiezza fino a tale livello **m**.

- Ciò implica che l'ampiezza totale del campione potrebbe essere superiore al numero n predeterminato. Laddove le risorse non consentano di andare oltre tale limite, occorrerà ricalcolare la frazione di campionamento escludendo sia da n che da N gli strati che sono stati corretti, e applicando nuovamente il campionamento stratificato proporzionale agli strati restanti.
- Logicamente, se dal nuovo campionamento risultasse qualche altro strato numericamente inferiore al minimo m , sarebbe necessario ripetere l'intera procedura di **aggiustamento** e poi **esclusione** di tale strato da una nuova stratificazione.

- Se nell'esempio precedente si fosse scelto di costruire un campione di soli **500** casi, le numerosità degli strati campionari sarebbero state, rispettivamente, 175, 225, 75 e 25. Ove si ponga (per motivi inferenziali che saranno chiariti nel seguito) a **50** il limite **m**, si dovrebbero dunque rilevare 525 casi invece di 500.
- Se, invece, il limite di 500 unità fosse **invalidabile**, si fissa un numero di 50 professionisti e si riproporzionano le restanti 450 unità (semplicemente **dividendo** le numerosità degli altri singoli strati per $500-25=475$ e poi **moltiplicando** ogni risultato per $500-50=450$), ottenendo in tal modo, rispettivamente, strati da **166**, **213** e **71** casi.
N.B.: i risultati esatti del calcolo sarebbero 165,8 operai, 213,2 impiegati e 71,1 lavoratori autonomi: è ovviamente necessario arrotondare tali risultati all'unità più vicina, con il vincolo di mantenere il previsto totale di 450 unità (in caso di necessità, arrotondare per eccesso il numero minore o per difetto quello maggiore).

Regola n° 1: Le variabili di stratificazione

Non esistono criteri assoluti o oggettivi per la scelta delle variabili di stratificazione, ma solo indicazioni di massima.

Suggerimenti

- Le variabili scelte per la stratificazione devono essere correlate con la variabile, o le variabili, di interesse per la rilevazione, ma tra loro devono essere indipendenti.
- Una buona variabile di stratificazione è, generalmente, la *suddivisione territoriale*.
- Nelle indagini *multiscopo*, la scelta delle variabili di stratificazione non è più finalizzata alla massima efficienza, ma ad una migliore suddivisione della popolazione sulla base delle conoscenze che si hanno sul fenomeno.

Regola n° 2: Il numero di strati

Nemmeno per la scelta del numero di strati esistono criteri assoluti o oggettivi.

Suggerimenti

1. L'efficienza delle stime aumenta con il numero di strati;
2. Tuttavia, in linea di tendenza, dopo un certo numero di suddivisioni della popolazione il beneficio in termini di efficienza è modesto;
3. Inoltre, all'aumentare del numero di strati crescono i costi della rilevazione
4. Un numero elevato di strati è auspicabile quando il campionamento è *su base territoriale*, poiché si controlla la dispersione delle unità e si rende più agevole l'organizzazione e l'esecuzione del lavoro sul campo.

Campionamenti per *cluster*

In questo tipo di campionamento, invece di estrarre un campione casuale di soggetti (unità elementari), viene estratto un campione casuale di **gruppi di soggetti**, chiamati “*cluster*”, contraddistinti da un elemento comune (aree geografiche, edifici, scuole); in seguito si cerca di selezionare tutte le unità comprese al loro interno, senza ricorrere a procedure casuali.

Attenzione:

I campionamenti per *cluster* sono adeguati ed utilizzabili solo quando la variabilità è **minima** tra i cluster e **massima** al loro interno, ossia quando i cluster sono omogenei tra loro (potendosi così ritenere solo suddivisioni arbitrarie di un unico collettivo), mentre al loro interno si rilevano tutte le differenziazioni che si potrebbero rilevare nella popolazione.

Campionamenti per *cluster*

Vantaggi:

- ✓ Non occorre avere una lista di tutti gli individui della base di campionamento, ma basta enumerare tutti i cluster selezionati per far sì che tutti gli individui al loro interno entrino a far parte del campione.
- ✓ Si elimina il rischio dei rifiuti legati ai meccanismi di inclusione/esclusione dal campione
- ✓ E' molto più facile raggiungere persone aggregate in un cluster piuttosto che disperse sul territorio.

Svantaggi

- ✓ Il campionamento a cluster non è utilizzabile quando la variabilità è maggiore tra i cluster che non al loro interno.

Campionamenti per *cluster*: il campionamento a grappoli

E' una procedura che viene utilizzata quando la popolazione è naturalmente suddivisa in gruppi di unità simili o contigue (famiglie, classi scolastiche, utenti di trasporti pubblici, ecc.), ma eterogenei al loro interno come la popolazione nel suo complesso.

E' utile quando manca la lista delle unità elementari.

Dato un insieme di gruppi o **grappoli** di un determinato tipo (per esempio, le scuole di una provincia), se ne estrae casualmente un determinato numero (numero che va definito dividendo la numerosità campionaria prefissata per la dimensione media dei grappoli stessi) e poi si considerano **tutte** le “unità finali” appartenenti a tali grappoli.

Campionamenti per *cluster*: il campionamento per aree

In questo tipo di campionamento l'unità da rilevare è un'area geografica (per es., una grande città). La **lista** delle **unità di primo livello** è qui un insieme di aree dai contorni inequivocabili su una mappa, per es., gruppi di isolati urbani o precise delimitazioni poligonali (quadrati) sulla mappa, definite in modo da non sovrapporsi.

Da questa lista, si estraggono casualmente alcune di queste aree e poi si selezionano tutte le unità statistiche in esse contenute (famiglie, aziende ecc.), oppure solo parte di esse, eventualmente con procedimento sistematico (per es., solo i residenti ai piani dispari dei numeri civici pari). In questo secondo caso il campionamento non è più solo “per cluster”.

Il campionamento a più stadi

- ✓ Questo metodo (caso particolare di campionamento a cluster) prevede una selezione delle unità campionarie in cascata, in due o più stadi successivi.
- ✓ Non comporta un aumento di efficienza rispetto al c.c.s, ma una semplificazione della procedura di estrazione ed una diminuzione dei costi di rilevazione quando la popolazione è molto numerosa e dispersa sul territorio.
- ✓ Rappresenta una scelta forzata quando manca la lista completa delle unità della popolazione.
- ✓ Le unità vengono divise in unità primarie, unità secondarie, terziarie e così via, fino alle unità finali; in genere, solo per queste ultime sarà opportuno o necessario disporre di liste nominative per l'estrazione.

- Si tratta di *popolazioni gerarchiche*, in cui una popolazione di unità statistiche è contenuta in un insieme di unità di livello superiore (che a sua volta può essere contenuto da un altro: ad es., studenti in un c.d.l., c.d.l. in un Dipartimento, Dipartimenti in un Ateneo, Atenei nell'intero sistema nazionale).
- Per selezionare un campione a stadi è generalmente necessaria la *lista delle unità finali*.
- Ad ogni stadio sono necessarie soltanto le singole liste delle sub-popolazioni contenute nelle unità selezionate a livello superiore (trattandosi in genere di raggruppamenti di tipo sociale o territoriale delle unità di livello inferiore, sono quasi sempre facilmente disponibili o ricavabili).

I passi del campionamento a più stadi

1. Determinare il numero di stadi su cui effettuare la selezione in base a:
 - ✓ Accessibilità delle liste;
 - ✓ Costi;
 - ✓ Reperibilità delle informazioni.
2. Individuare le caratteristiche per stratificare le unità di primo stadio.
3. Decidere quante unità selezionare al primo e ai successivi stadi;
4. Decidere come selezionare le unità;

Campionamento a stadi:

Vantaggi

1. Flessibilità e adattabilità

- L'estrazione si può effettuare con criteri differenti a ogni stadio.

2. Riduzione dei costi

- La rilevazione dei dati è concentrata sui punti selezionati al primo stadio;
- L'organizzazione del lavoro (formazione delle liste, selezione del campione, reclutamento del personale, esecuzione della rilevazione, supervisione sul campo, ...) risulta quindi facilitata.

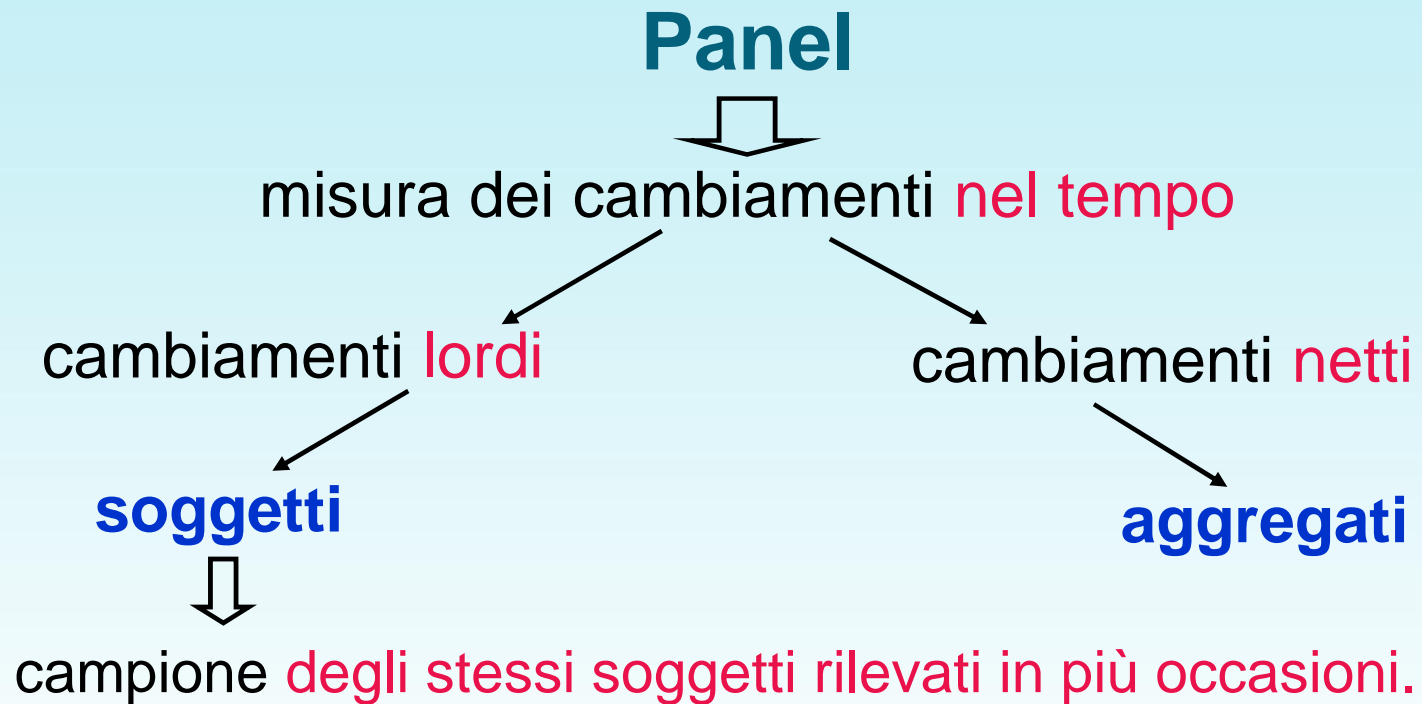
Svantaggi

Complessità della stima e rischio di stime inefficienti

Le unità appartenenti ad un insieme molto omogeneo tendono ad assomigliarsi e quindi le loro risposte risultano penalizzate in termini di variabilità e di rappresentatività.

Campionamento ruotato

Questo tipo di campionamento, utile nelle rilevazioni per *panel* (cioè fatte in tempi diversi sulle unità del medesimo campione), serve sia a ridurre gli inconvenienti dei *panel* (difficoltà di trovare soggetti disposti a collaborare per molto tempo, “mortalità da *panel*” e condizionamento da *panel*) e sia a studiare i flussi di unità che cambiano stato.



- Problemi:
- Difficoltà di reclutamento
 - Mobilità in uscita
 - Condizionamento da intervista

Possibili alternative: **rotazione** del campione



sostituzione di alcune unità statistiche con altre appositamente selezionate per essere utilizzate

La procedura del campionamento ruotato prevede che, dopo prefissati intervalli di tempo, alcune unità statistiche vengano sostituite *a rotazione* con altre appositamente previste, selezionate in modo che siano il più possibile omogenee con quelle sostituite, ma solo rispetto alle finalità dell'indagine (per es., nell'indagine sulle Forze di Lavoro si sostituiscono le unità con altre residenti nello stesso comune-campione).

Schema base di rotazione trimestrale delle sezioni di rilevazione

Gruppi di rotazione	<i>Anno t-2</i>	<i>Anno t-1</i>				Anno t				Anno t+1				Anno t+2				
	<i>Trimestre IV</i>	<i>I</i>	<i>II</i>	<i>III</i>	<i>IV</i>	<i>I</i>	<i>II</i>	<i>III</i>	<i>IV</i>	<i>I</i>	<i>II</i>	<i>III</i>	<i>IV</i>	<i>I</i>	<i>II</i>	<i>III</i>	<i>IV</i>	
A	*	*			*	*												
B		*	*			*	*											
C			*	*			*	*										
D				*	*			*	*									
A'					*	*			*	*								
B'						*	*			*	*							
C'							*	*			*	*						
D'								*	*			*	*					
A''									*	*			*	*				
B''										*	*			*	*			
C''											*	*			*	*		
D''												*	*			*	*	
A'''													*	*			*	*
B'''														*	*			
C'''															*	*		
D'''																*	*	
A''''																		*

La scelta del tipo di campionamento da effettuare (ma anche della tecnica di rilevazione) dipende dunque dal confronto fra la dimensione massima data dal budget disponibile e la dimensione ottimale del campione ottenuta in base a limiti (probabilistici) di precisione delle stime. Un criterio di scelta, nel caso che tale dimensione ottimale n sia maggiore di quella massima «da budget» può essere espresso come segue:

- se la tecnica di indagine è ritenuta fondamentale (ad esempio, se occorre forzatamente utilizzare la CATI), si può ridurre n a un livello compatibile con il valore massimo riducendo la precisione delle stime (cioè aumentando il semi-intervallo di confidenza $d/2$) oppure utilizzando un tipo di campionamento più efficiente (come il *campionamento stratificato*);

- se, invece, l'elemento più importante è la precisione dell'indagine oppure la sua copertura, va incrementato il valore massimo n' (o n^*) riducendo i costi variabili:

per esempio, l'indagine Almalaurea sui laureati, che ha come scopo la massima copertura possibile (e, quindi, è tendenzialmente un'indagine censuaria sul *dominio statistico* dei laureati), per ridurre i costi variabili utilizza una tecnica di **rilevazione mista**, dapprima invitando i laureati presenti nel database a partecipare a un'indagine CAWI (ove viene registrata l'identità di chi accede all'indagine, anche se resta anonimo l'insieme delle risposte), e poi contattando telefonicamente coloro che non hanno risposto al precedente invito, avviando in tal modo un'indagine CATI.

Un esempio: l'Indagine ISTAT sulle forze di lavoro

Dal 1959: rilevazione dei principali aggregati dell'offerta di lavoro mediante un **questionario** somministrato ad un **campione di famiglie**, intervistate presso le loro abitazioni.

Obiettivo: Stima dei principali aggregati dell'offerta di lavoro.

La rilevazione è detta “**continua**” in quanto le informazioni sono rilevate con riferimento a tutte le settimane dell'anno, tenuto conto di un'opportuna distribuzione nelle tredici settimane di ciascun trimestre del campione complessivo.

La rilevazione è progettata per garantire stime trimestrali a livello regionale e stime provinciali in media d'anno. Le stime trimestrali rappresentano lo stato del mercato del lavoro nell'intero trimestre.

Indagine sulle forze di lavoro: definizioni

Forze di lavoro: comprendono le persone occupate e quelle disoccupate.

Occupati: comprendono le persone di 15 anni e più che nella settimana di riferimento:

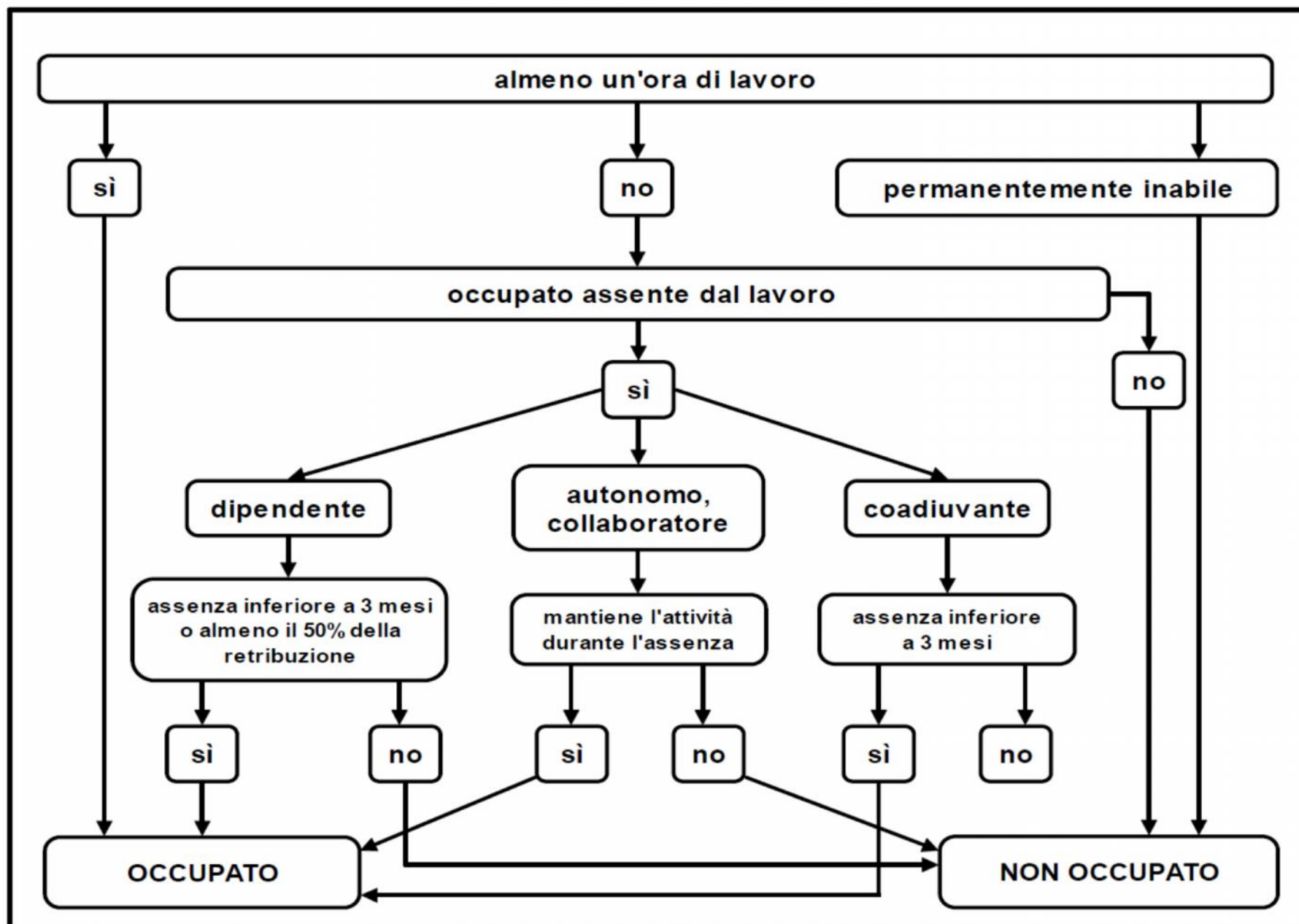
- hanno svolto almeno un'ora di lavoro in una qualsiasi attività che preveda un corrispettivo monetario o in natura;
- hanno svolto almeno un'ora di lavoro non retribuito nella ditta di un familiare nella quale collaborano abitualmente;
- sono assenti temporaneamente dal lavoro (ad esempio, per ferie o malattia).

Persone in cerca di occupazione: comprendono le persone non occupate tra 15 e 74 anni che:

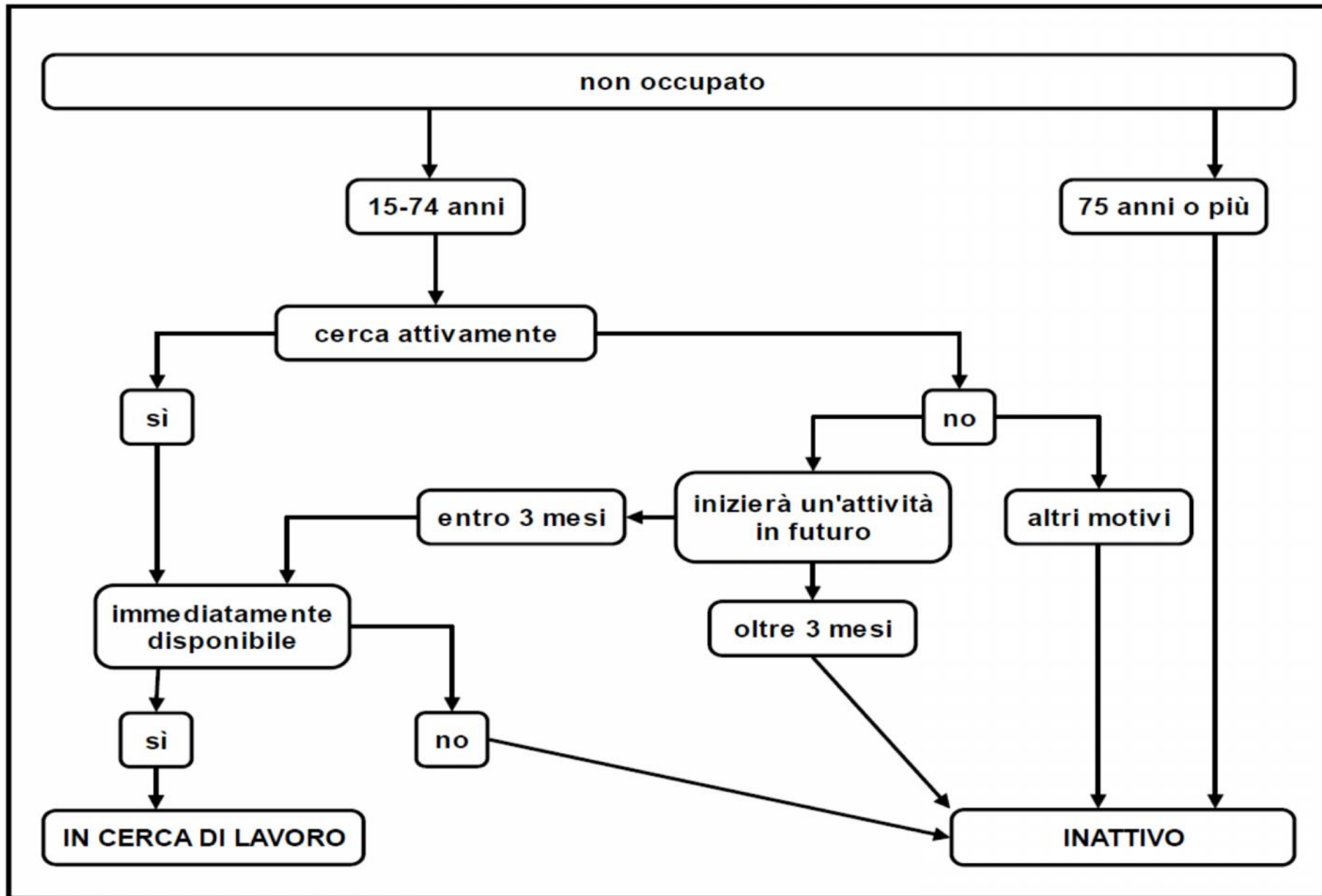
- hanno effettuato almeno un'azione attiva di ricerca di lavoro nei trenta giorni che precedono l'intervista e sono disponibili a lavorare (o ad avviare un'attività autonoma) entro le due settimane successive all'intervista;
- oppure inizieranno un lavoro entro tre mesi dalla data dell'intervista e sono disponibili a lavorare (o ad avviare un'attività autonoma) entro le due settimane successive all'intervista, qualora fosse possibile anticipare l'inizio del lavoro.

Inattivi: comprendono le persone che non fanno parte delle forze di lavoro, ovvero quelle non classificate come occupate o in cerca di occupazione.

Individuazione degli occupati



Individuazione delle persone in cerca di occupazione



I TASSI

- ***Tasso di attività:*** rapporto tra le persone appartenenti alle forze di lavoro e la corrispondente popolazione di riferimento.
- ***Tasso di occupazione:*** rapporto tra gli occupati e la corrispondente popolazione di riferimento.
- ***Tasso di disoccupazione:*** rapporto tra le persone in cerca di occupazione e le corrispondenti forze di lavoro.
- ***Tasso di disoccupazione di lunga durata:*** rapporto tra le persone in cerca di occupazione da dodici mesi e oltre e le forze di lavoro.

Indagine sulle forze di lavoro: popolazione e intervista

La popolazione di riferimento è costituita da tutti i componenti delle famiglie residenti in Italia, anche se temporaneamente all'estero. La popolazione residente comprende le persone, di cittadinanza italiana e straniera, che risultano iscritte alle anagrafi comunali.

L'unità di rilevazione è la **famiglia di fatto**, definita come “insieme di persone coabitanti, legate da vincoli di matrimonio, parentela, affinità, adozione, tutela o da vincoli affettivi”.

L'intervista alla famiglia viene effettuata utilizzando una rete di rilevazione controllata direttamente dall'Istat mediante tecniche **CATI** (*Computer assisted personal interview*) e **CAWI** (*Computer assisted telephone interview*). Se tali rilevazioni si rivelano inattuabili, la famiglia viene sostituita oppure (come *extrema ratio*) interviene un rilevatore per svolgere un'intervista “in presenza”.

In generale le informazioni vengono raccolte con riferimento alla settimana che precede l'intervista.

Indagine sulle forze di lavoro: il campione

Il campione utilizzato è **a due stadi**, rispettivamente **comuni** e **famiglie**, con stratificazione delle unità di primo stadio.

In ciascun trimestre vengono intervistati circa 77 mila famiglie residenti in oltre 1.000 comuni di tutte le province del territorio nazionale (il numero di comuni può cambiare negli anni).

I comuni (unità' di 1° stadio) sono raggruppati in strati per **dominio territoriale minimo** (provincia) e per **dimensione demografica**.

Tutti i comuni capoluogo di provincia o con popolazione superiore ad una soglia predeterminata per ciascuna provincia, detti "**AR**" (**autorappresentativi**), sono sempre presenti nel campione, ma con frequenza differente: i primi e i comuni AR di maggiore dimensione sono inclusi in ogni rilevazione settimanale, mentre i comuni AR di minore dimensione sono inclusi una sola settimana al mese.

I comuni la cui popolazione è al di sotto delle soglie prefissate, detti “**NAR**” (**non autorappresentativi**), invece entrano nel campione a rotazione, ciascuno per una sola settimana al mese, attraverso uno schema di associazione casuale delle settimane ai comuni NAR del campione provinciale.

Per ciascun comune (AR e NAR) viene estratto dalla lista anagrafica un campione casuale semplice di famiglie, con dimensione proporzionale alla dimensione dello strato nella popolazione (in realtà, si tratta di 4 campioni distinti, perché per ogni “famiglia base” vengono estratte altre 3 “famiglie supplenti” nel caso questa non possa o non voglia partecipare all’indagine).

Ogni famiglia selezionata viene intervistata per due trimestri consecutivi; segue un’interruzione per i due successivi trimestri, dopo di che essa viene nuovamente intervistata per altri due trimestri (schema di rotazione “**2_T, 2_T, 2_T**” delle unità di 2° stadio).

In tutto, ogni famiglia campionata rimane nel campione per un periodo di 18 mesi, dopo di che viene sostituita da un’altra famiglia.

Con il campionamento ruotato si vuol stimare non solo il livello del fenomeno, ma anche quante persone in un dato intervallo di tempo hanno cambiato stato nei riguardi dell'occupazione.

Per chiarire il meccanismo, si tenga presente che ciascun gruppo di 4 famiglie (la “famiglia base” e le 3 supplenti) estratto nel trimestre x dell'anno t , definito “gruppo di rotazione”, nello schema seguente è contraddistinto con lettere maiuscole.

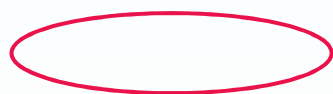
Ogni trimestre viene dato l'avvio a un nuovo gruppo di rotazione.

I diversi gruppi si alternano nella sequenza temporale in modo che ciascuno di essi si sovrapponga a gruppi della stessa coorte solo e sempre per il 50% del periodo in cui permangono nell'indagine. Ciò avviene sempre negli stessi trimestri, a distanza di un anno.

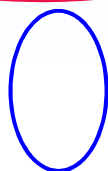
Inoltre, la sovrapposizione fra due determinati gruppi di rotazione (ad esempio E e F) può avvenire solo e sempre per il 25% del loro periodo di permanenza, sempre negli stessi trimestri e a distanza di un anno.

Schema di rotazione indagine sulle Forze di Lavoro

Gruppo di Rotazione	Anni, Trimestri					
	t, I	t, II	t, III	t, IV	t+1, I	t+1, II
A	X	X				
B		X	X			
C			X	X		
D				X	X	
E	X	X			X	X
F		X	X			X
G			X	X		
H				X	X	
I					X	X
L						X



sovrapposizione 50% in due trimestri successivi



sovrapposizione 25% nello stesso trimestre a distanza di un anno

Indagine sulle forze di lavoro: riporto dei risultati campionari alla popolazione

Il riporto dei dati campionari all'universo, che nel caso delle Forze di Lavoro è costituito dalla popolazione residente (al netto dei membri permanenti delle convivenze), si effettua per tutti gli s strati e distintamente per ciascun sesso.

La frequenza campionaria di ciascuna modalità riguardante i maschi o le femmine (ad es., la frequenza degli occupati, disoccupati, ecc.) rilevata in un dato strato i del campione viene moltiplicata per un coefficiente “di espansione”, MK_i o FK_i (rispettivamente, per i maschi e le femmine), calcolato ad ogni rilevazione mediante il rapporto tra numero di maschi (o femmine) residenti nell' i -mo strato alla data di riferimento dell'indagine e numero di maschi (o femmine) campionati nello stesso strato..

Ad esempio, presso 60 famiglie residenti in un comune-campione, si sono rilevati 118 maschi e 129 femmine, di cui 70 M e 25 F si sono dichiarati occupati nella settimana di riferimento. Se la popolazione residente di tale comune è, alla stessa data, pari a 48.254 persone (23.131 M e 25.123 F), si avranno i seguenti coefficienti di espansione:

$${}_M K_i = 23.131 / 118 = 196,0 \quad ; \quad {}_F K_i = 25.123 / 129 = 194,8 .$$

Ora si può stimare facilmente il numero di occupati nel comune:

numero di occupati	$196,0 \times 70 =$	13.720
numero di occupate	$194,8 \times 25 =$	<u>4.870</u>
Totale occupati		18.590

Tale procedimento, esteso a tutti i comuni-campione (AR o NAR) di tutti gli strati, permette di conoscere, per somma, l'ammontare degli occupati nelle singole regioni e quindi nell'intero Paese.

Quanto detto per l'occupazione vale anche per le altre modalità rilevate nell'indagine (disoccupazione, ore di lavoro, ecc.).

In generale, il riporto dei risultati campionari alla popolazione può essere formalizzato come segue:

Per stimare l'ammontare del carattere Y di una popolazione finita di N unità, a partire dalle n determinazioni y_j di un campione estratto da detta popolazione, se si conosce la probabilità p_j che la j -ma unità aveva di essere estratta al momento del campionamento, si può utilizzare lo *stimatore* (ossia lo strumento matematico più adeguato a fornire la stima del parametro):

$$\hat{Y} = \frac{1}{n} \sum_{j=1}^n \frac{y_j}{p_j} .$$

Se le probabilità sono tutte uguali, la formula si semplifica:

$$\hat{Y} = \frac{1}{n} \sum_{j=1}^n \frac{y_j}{1/N} = N \sum_{j=1}^n \frac{y_j}{n} = N\bar{y} \quad ,$$

ossia Y può essere stimato moltiplicando la media rilevata nel campione (stima corretta della media della popolazione) per la numerosità della medesima popolazione .

Tale formulazione, valida anche per il campionamento senza reinserimento, prende il nome di *stimatore per espansione*, poiché stima il totale del carattere nella popolazione come risultato della moltiplicazione tra il coefficiente di espansione N/n e il **totale** misurato tramite le unità campionate.

Campionamenti non probabilistici: il campionamento per quote

- *Sostanzialmente, questa tecnica è una specie di campionamento stratificato con scelta rimessa all'intervistatore (senza uso di liste e dunque senza criteri oggettivi), con un risparmio di costo ma minor fedeltà alla struttura reale della popolazione.*
- 1. Si suddivide la popolazione in sottogruppi sulla base di variabili di cui si conosce la distribuzione.
- 2. Si determinano le quote del campione in modo proporzionale ai sottogruppi della popolazione.
- 3. L'intervistatore può scegliere a propria discrezione i soggetti da intervistare all'interno di ciascuna quota.

Campionamenti non probabilistici: il campionamento a scelta ragionata

- E' costituito da unità scelte in modo da somigliare nell'insieme alla popolazione da cui sono tratte, in riferimento all'oggetto dell'indagine..
- E' tanto più rappresentativo quanto più sono vere le informazioni su cui si basa la scelta.
- Non sono applicabili gli schemi di calcolo delle probabilità.
- Dipende molto dalle scelte degli operatori .

Campionamenti non probabilistici: il campionamento bilanciato

Spesso il campionamento ragionato viene bilanciato, ossia si considerano una serie di *variabili di controllo*, attinenti all'indagine, per le quali sono note le sintesi (per esempio, le medie nella popolazione) e si cerca di campionare unità le cui variabili di controllo abbiano media uguale o molto prossima a quella della popolazione.

In alternativa, tale selezione può essere fatta “a posteriori” su un campione ragionato di numerosità maggiore del necessario, eliminando le unità le cui variabili di controllo divergono troppo in una direzione (o nell'altra), in modo che la loro media nel campione si approssimi sempre più alla media rilevata nella popolazione.

Campionamenti non probabilistici: il campionamento a valanga (*snowball*)

E' un campionamento di soggetti da collettivi inizialmente ignoti al rilevatore, nel quale si selezionano alcune prime unità appartenenti alla sub-popolazione in esame, le quali sono chiamate non solo a partecipare all'indagine (per es. rispondendo a un questionario), ma anche a indicare altre unità di propria conoscenza (in genere, da 1 a 3).

Esso è spesso utile per:

1. fini di marketing (reti di consumatori)
2. ricerche tramite “esperti” o “stakeholder”
3. ricerche statistiche su fenomeni illegali o su temi delicati

Tale campionamento si dice “a valanga” perché ad ogni passo il campione già rilevato coinvolge nuove unità.

Il processo di selezione e rilevazione, dunque, si svolge con la collaborazione delle unità selezionate.

Un semplice schema di campionamento di questo tipo basato sulla tecnica dei "tre amici più prossimi", ossia sulla richiesta rivolta al campione di individui di indicare altre tre persone nella stessa situazione.

In altri disegni campionari, questa tecnica chiede di identificare, fra tre amici prossimi, se uno o più di essi rientra nel *dominio* dell'indagine (ossia nella sub-popolazione studiata).

La selezione iniziale delle unità può essere **casuale** oppure orientata all'obiettivo dell'indagine.

Un *buon campione*

In conclusione, per ottenere un *buon campione*:

- è meglio intervistare 300 soggetti scelti casualmente dalla lista della popolazione che 1000 raggiunti in modo arbitrario;
- è auspicabile un confronto fra le caratteristiche del campione e quelle note della popolazione;
- Il numero ottimale di unità che dovrebbe far parte del campione si ottiene con procedure (in genere legate a concetti probabilistici) che tengano conto delle caratteristiche della popolazione.