

CAMPIONAMENTO

Per una buona **inferenza o induzione statistica**, bisogna affrontare il problema del campionamento, ovvero del come raccogliere un campione della popolazione affinché la si possa studiare induttivamente.

Per fare questo, intanto il Campione deve essere **RAPPRESENTATIVO**.

La rappresentatività è garantita da una **campionamento casuale** o aleatorio, in cui cioè ogni elemento della popolazione ha la medesima probabilità di entrare a far parte del campione.

Per esempio, avendo una popolazione (finita) di 500 elementi, ognuno di essi avrà una probabilità di essere estratto pari a $1/500$, e nessuno di essi ne avrà una differente.

Casuale pertanto non significa *preso a caso* ma **equiprobabile**.

Solo così si può applicare la Teoria della probabilità per *inferire* correttamente i dati.

IL CAMPIONE ALEATORIO

Un campione è detto aleatorio quando gli elementi disponibili della popolazione hanno la stessa probabilità di farne parte.

Perché ci si riferisce ad elementi disponibili e non semplicemente agli elementi della popolazione?

Per rispondere, bisogna tener conto che il campionamento può essere

CON O SENZA REIMMISSIONE.

Per estrarre un campione casuale si procede così:

- * Si numerano gli elementi (per rendere appunto equiprobabile - come lo sono i numeri - l'estrazione)
- * Si estraggono a caso i numeri associati agli elementi.

Ora non resta che scegliere se effettuare reimmissioni o meno.

Quando la popolazione è teoricamente infinita e il campione è piccolo, CON O SENZA è ininfluente.

Quando la popolazione è finita, la differenza fra i due metodi c'è.

Per esempio, con una popolazione di 500 elementi, ogni elemento che viene estratto **con** reimmissione ha $P=1/500$ e si dice che i relativi risultati delle estrazioni **sono indipendenti**. Ciascuna delle unità potrà essere estratta più volte. (SI DICE "ESTRAZIONE BERNOULLIANA" e il numero dei campioni ottenuti è pari a $D'_{N,n}=N^n$)

Con una estrazione **senza** reimmissione, invece, ad ogni estrazione la probabilità cambia, nel senso che aumenta man mano per gli elementi rimasti (il numero dei campioni è in

questo caso pari a $\binom{N}{n} = \frac{N!}{n!(N-n)!}$ **Attenzione:** ogni elemento disponibile, comunque,

presenta, ad ogni stadio di campionamento, l'equiprobabilità di essere estratto!

Nell'esempio precedente, avremmo:

I estrazione: $1/500$

II estrazione: $1/499$

III estrazione: $1/488...$ Ecc.

I risultati di queste estrazioni **NON sono indipendenti** giacché ciascuna dipende dalla precedente.

AFFIDABILITA' ED EFFICENZA

Lo studio di un campione non è fine a se stesso, ma serve per risalire alle caratteristiche della popolazione alla quale si riferisce, per esempio attraverso stime dei parametri riassunti nella seguente tabella:

	Popolazione (PARAMETRI)	Campione (STATISTICHE o STIME)
Indicatore		
Media	μ	$\bar{x}$
Deviazione	σ	s
Varianza	σ^2	s^2
Ampiezza	N	n
	SONO TUTTE COSTANTI	SONO TUTTE VARIABILI ALEATORIE. che formano una Distribuzione Campionaria

Guardiamo per esempio la media.

Da un lato, abbiamo la media μ della popolazione, dall'altro la media stimata $\bar{x}$ proveniente da campioni aleatori. La differenza tra queste due medie (supposte non uguali) prende il nome di **errore di campionamento**, che è una misura di **affidabilità** del campionamento.

Per quanto riguarda l'**efficienza**, essa è legata al costo: un campionamento è più efficiente di un altro se, a parità di affidabilità, è meno costoso.

TIPI DI CAMPIONAMENTO

CAMPIONAMENTO CASUALE SEMPLICE A UNO O PIU' STADI

Ha il proprio principio nella casualità delle estrazioni.

Per l'estrazione casuale si adottano i numeri casuali. Ogni numero casuale è composto da 5 cifre estratte casualmente dall'insieme delle 10 cifre (0,1,2...9). Queste 5 cifre hanno le seguenti proprietà:

- * Sono indipendenti, ovvero ottenute mediante estrazioni con reimmissione;
- * Ogni cifra ha la stessa probabilità di essere una delle 10 possibili;
- * Ogni gruppo di cifre ha la stessa probabilità di essere 00, 01,...98,99 se coppia, 000, 001,...098,099 se terna, ecc.

Il campionamento ad 1 stadio si adotta quando le unità del campione sono facilmente individuabili e l'insieme di partenza è limitato.

Per esempio: da un insieme di 215 professori, se ne devono scegliere 20 per una Commissione.

I numeri aleatori (215) che si attribuiscono sono in questo caso terne: 001, 002,...215. Quindi si può usare una Tavola di numeri casuali scegliendo, nel nostro caso, gruppi da tre e partendo la lettura dall'incrocio casuale di una riga e di una colonna. Dei numeri estratti, dobbiamo prendere solo quelli tra 1 e 215 e ciascuno una volta sola, perché l'estrazione è logicamente senza reimmissione (si tratta di persone).

Ma se la popolazione di partenza è costituita da molti elementi, questo metodo non è più idoneo.

Nell'esempio di un industriale che volesse controllare il grado batterico dei formaggi che produce (e che non può ovviamente farlo per tutta la produzione), come campionerà?

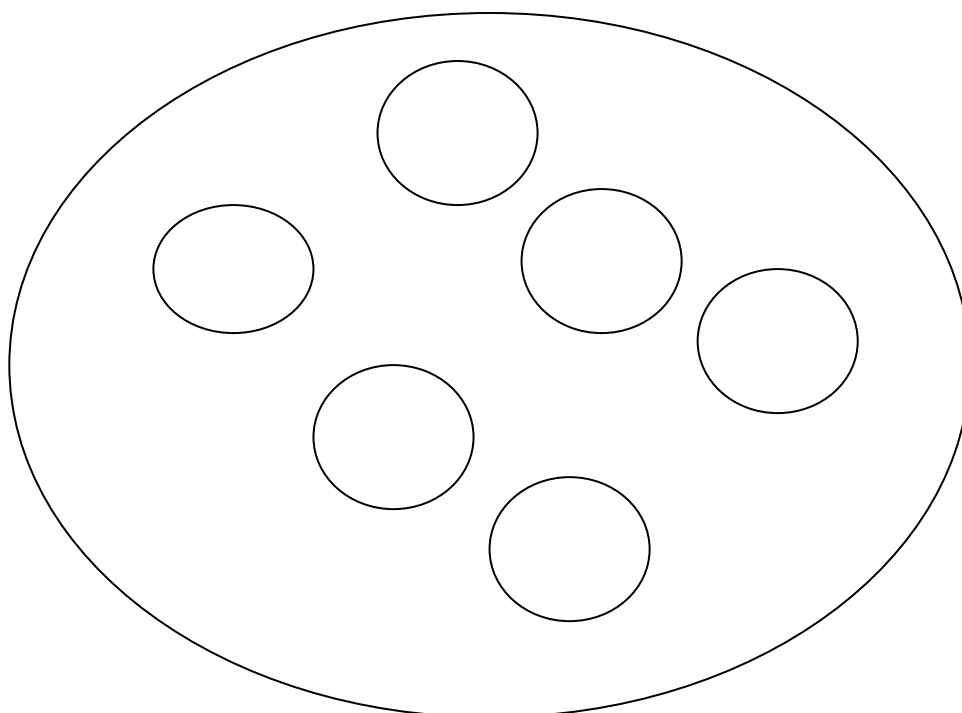
Supponendo che l'industria lavori 200 giorni all'anno per otto ore al giorno e che produca migliaia di pezzi di formaggio, sarà necessario un **campionamento multiplo**, che suddivida la popolazione in sottogruppi via via sempre meno complessi. Per esempio:

- * 1° stadio: scelta (casuale) di alcuni giorni tra i duecento; poniamo 50;
- * 2° stadio: scelta (casuale) di 1 delle 8 ore lavorative, numerando le ore da 0 a 7;
- * 3° stadio: all'interno di quell'ora, scelta delle porzioni di formaggio da esaminare.

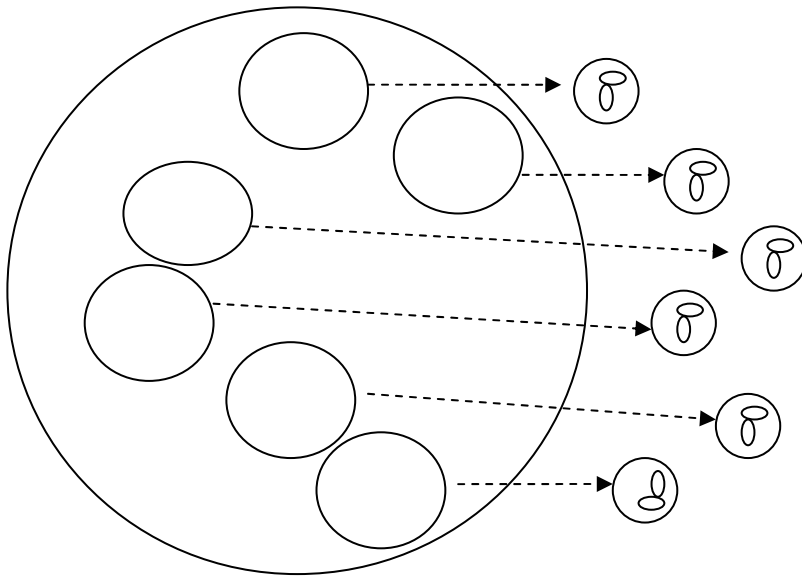
CAMPIONAMENTO STRATIFICATO

La stratificazione è la suddivisione dell'insieme di partenza in gruppi (**strati**); ciascuno strato riunisce tutti gli elementi della popolazione che hanno un carattere in comune (**fattore di stratificazione**).

Per esempio, il carattere "professione" come fattore di stratificazione potrebbe essere suddiviso in 6 strati: *professionisti, impiegati privati, impiegati pubblici, artigiani, commercianti, disoccupati*.



Una volta stratificata la popolazione, bisogna estrarre unità campionarie (dette **campi di indagine**) da ogni strato. Il numero degli elementi per ogni campione deve essere proporzionale alla dimensione dello strato rispetto alla popolazione (**criterio di proporzionalità**); per esempio, se lo strato 1 riunisce il 5% della popolazione, bisogna estrarre da questo strato un numero di elementi pari al 5% del campione.



Esempio: Si fa un'indagine su studenti così ripartiti per corsi:

1° anno	420	30%
2° anno	280	20%
3° anno	210	15%
4° anno	210	15%
5° anno	182	13%
fuori corso	98	7%
totale	1400	100%

Si decide per un campione di 200 elementi.

In questi 200 studenti, ce ne dovranno essere:

il 30% del 1° anno	= 60
il 20% del 2° anno	= 40
il 15% del 3° anno	= 30
il 15% del 4° anno	= 30
il 13% del 5° anno	= 26
il 7% dei fuori corso	= 14

I vantaggi della stratificazione sono evidenti per **popolazioni eterogenee**, nel senso che si assicura la partecipazione omogenea di ogni strato al campione finale. Quindi, possiamo dire che **la stratificazione riduce gli errori dovuti allo sbilanciamento del campione** (sbilanciamento = partecipazione insufficiente o preponderante dei suoi elementi).

CAMPIONAMENTO A GRAPPOLI

Questo campionamento si utilizza di solito per popolazioni molto vaste.

In questo campionamento, gli elementi non vengono estratti uno alla volta ma per **sottogruppi** (grappoli) legati da criteri comuni: negozi di una certa zona, dati raccolti in uno stesso periodo, ecc).

DIFFERENZA TRA STRATIFICATO E GRAPPOLO

Gli strati sono differenti tra loro, e ciascuno contiene elementi simili (lo strato degli impiegati - tutti impiegati - è diverso dallo strato dei commercianti - tutti commercianti -).

I grappoli, invece, sono simili tra loro (tanto che ciascuno dovrebbe tendere a rappresentare da solo l'intera popolazione) e ciascuno contiene elementi con caratteristiche diverse: una famiglia è un grappolo di individui (diversi tra loro), un condominio è un grappolo di famiglie (diverse tra loro).

Procedura:

- si sceglie a caso un certo numero di grappoli i cui elementi entrano nel campione (campionamento a 1 stadio);
- da ogni **unità primaria** (grappolo) si estraggono **unità secondarie** che entrano nel campione (campionamento a 2 stadi);

si può continuare...

Esempio:

Si vuole stimare quanti litri di spumante si vende in Italia in un anno.

Si possono raggruppare i punti di vendita per province;

poi si possono estrarre alcune province (poniamo 8) e si limita lo studio nei punti vendita di queste soltanto (primo campionamento).

Per ognuna delle province, si possono estrarre tot punti vendita (poniamo 5- secondo campionamento).

In totale, si avranno così $8 \cdot 5 = 40$ elementi che costituiscono il campione.

In genere, il campionamento a grappoli produce informazioni meno affidabili di quello casuale semplice. Nel nostro esempio, se l'unità primaria "provincia" fosse unica e fosse l'Emilia, si otterrebbero dati sovradimensionati per la vendita dello spumante perché questa provincia è ad alto tenore economico rispetto alla media nazionale.

Tuttavia, i costi di questo campionamento sono ridotti: le unità del grappolo sono in genere molto vicine (i punti vendita della provincia o le province stesse) e ciò riduce costi di spostamento e durata del sondaggio.

CAMPIONAMENTO SISTEMATICO

Consiste nell'ordinare e numerare gli elementi dell'insieme di partenza, e poi estrarne gli elementi per il campione seguendo intervalli regolari.

Esempio:

da un gruppo di 10.000 elementi, si desidera un campione di 500.

Si calcola il rapporto $10000/500 = 20$,

si numerano i singoli elementi, e se ne preleva uno ogni 20 fino a ottenere il campione desiderato di 500.

E' importante che il punto di inizio dei prelievi sia scelto a caso (non necessariamente dal primo elemento) giacché così si garantirà l'equiprobabilità delle estrazioni.

Questo campionamento diventa poco affidabile quando nella popolazione esistono fluttuazioni periodiche. Per esempio, nella stima degli introiti medi di un supermercato non si dovrà scegliere il ciclo di 1 settimana giacché nel campione potrebbero capitare gli incassi di tutti i Sabato o dei lunedì, così come nei periodi pasquali o natalizi.