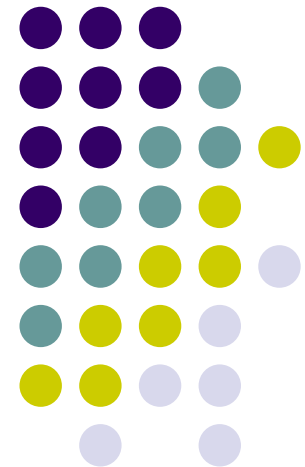


Il campionamento statistico



prof. C.Guida



- Per determinare le caratteristiche fondamentali di una popolazione statistica non è sempre necessario analizzare *tutta* la popolazione, ma risulta sufficiente esaminare un

CAMPIONE STATISTICO



Dopo aver analizzato le
informazioni relative ad un
campione....

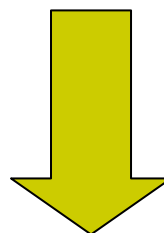


....i risultati ottenuti sulle statistiche
campionarie saranno utilizzati per risolvere
problemi di

INFERENZA STATISTICA



DALLA STATISTICA DESCRITTIVA



ALL'INFERENZA STATISTICA



L'INFERENZA STATISTICA

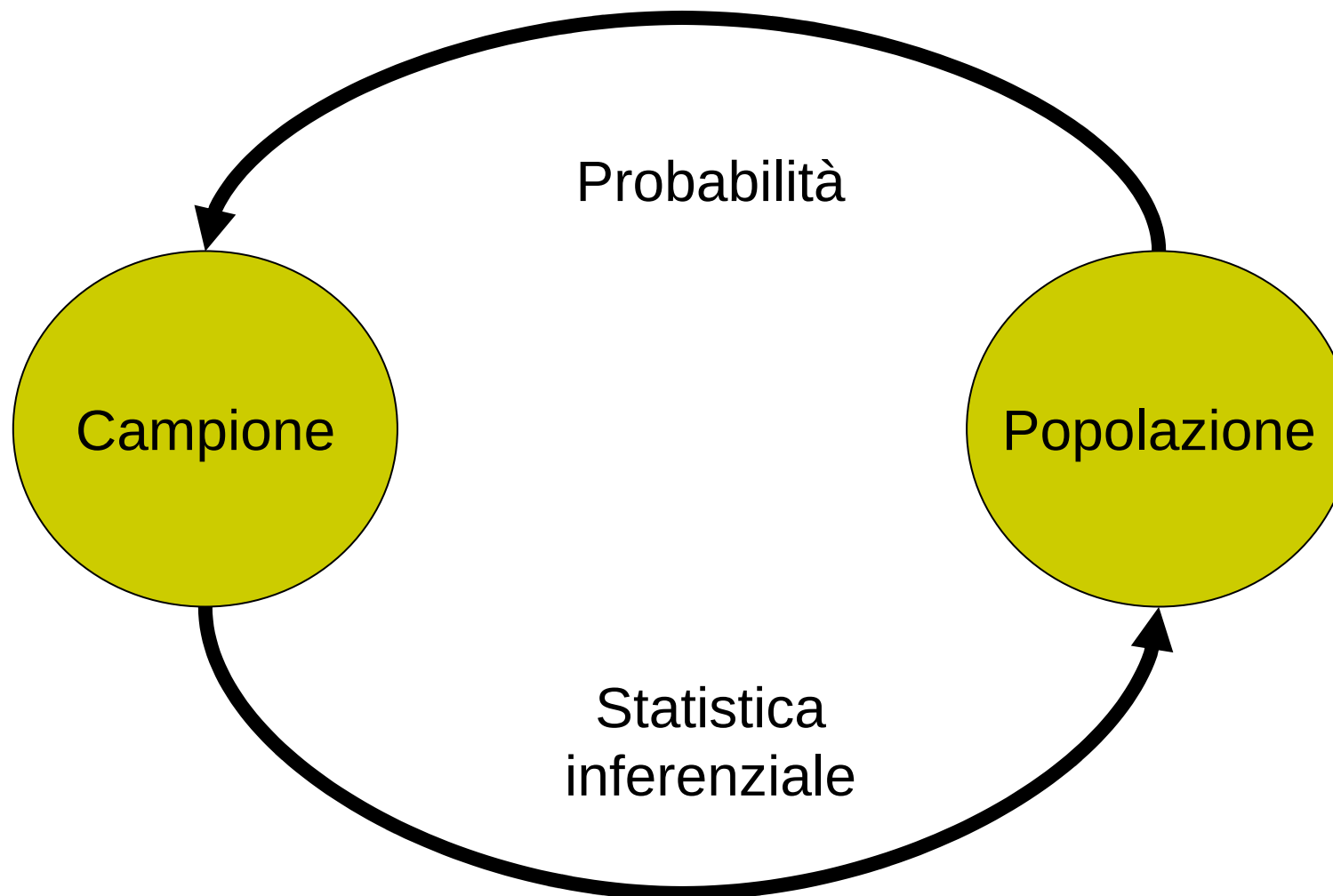
SI OCCUPA DI ESTENDERE

AD UNA INTERA POPOLAZIONE

LE INFORMAZIONI CHE SI RICAVALANO DALLE

STATISTICHE CAMPIONARIE

Statistica inferenziale





Lo studio di un campione, pertanto, 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:

	CAMPIONE		POPOLAZIONE
• Media	$\bar{x}$	STIME →	μ
• Deviazione	s		σ
• Varianza	s^2		σ^2
• Ampiezza	n		N

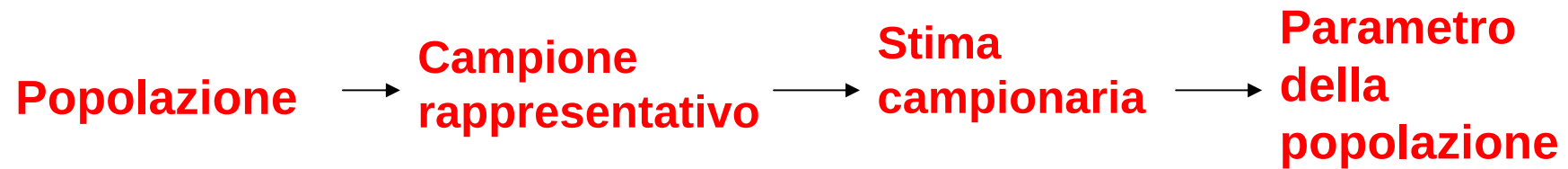
AFFIDABILITA' ED EFFICIENZA



.....nella precedente slide,

- 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.

RIEPILOGANDO.....



Inferenza statistica

Cos'è un campione statistico





DEFINIZIONE

Un campione statistico è un sottoinsieme
opportunamente scelto dall'intera popolazione



Per **campione statistico** s'intende dunque quel gruppo di unità elementari (non necessariamente persone), sottoinsieme particolare della popolazione o universo, individuato in essa in modo da consentire, con margini di errori contenuti, la generalizzazione all'intera popolazione.



Per quali ragioni può essere preferibile
analizzare i dati di un campione piuttosto di
quelli dell'intera popolazione ?



LE RAGIONI POSSO ESSERE VARIE



Si ricorre ad un campione perché.....

- ***La popolazione può essere molto vasta***
(risulta allora troppo costoso analizzare tutte le N unità statistiche)

Un sondaggio effettuato su un insieme di N unità statistiche è certamente più lungo e costoso di un sondaggio effettuato su un insieme ristretto di n individui !!!!

$$n < N$$



Si ricorre ad un campione perché.....

- ***Le misure possono essere distruttive***

(ad esempio nel caso di misure affidabilità o durata di un dato prodotto o componente elettronico)

Se volessimo misurare la durata di un componente elettronico, è necessario far funzionare tale componente fino a quando si guasta. Pertanto è conveniente sottoporre a tali verifiche un numero limitato di componenti !!!!!

.....i costi sarebbero elevati.

Si ricorre ad un campione perché.....



- *Le unità statistiche non presentano variabilità*

In questo caso risulta **INUTILE**
effettuare molte misure

I parametri della popolazione risultano determinati con poche misure.

***Se volessimo ad esempio determinare la media di 500
unità statistiche TUTTE dello stesso valore, tale media
sarà uguale a quella ottenuta se considerassimo
soltanto 10 unità!!!!***

Si ricorre ad un campione perché.....



- ***Non tutti gli elementi della popolazione sono disponibili***

(ad esempio, indagini sul passato possono essere effettuate solamente sui dati storici a disposizione, che spesso sono incompleti)

Si ricorre ad un campione perché.....



- *Vi è un notevole risparmio di tempo nel raggiungimento dei risultati*

in quanto, anche se si utilizzano gli elaboratori elettronici, la fase di DATA-ENTRY è notevolmente ridotta se si limita l'indagine a pochi elementi della popolazione complessiva!!!

IL PIANO DI CAMPIONAMENTO



E' L'INSIEME DI OPERAZIONI CHE PORTANO A:

1. DEFINIRE LA POPOLAZIONE OBIETTIVO NELLA RILEVAZIONE STATISTICA;
2. DEFINIRE LE UNITA' CAMPIONARIE;
3. STABILIRE LA DIMENSIONE DEL CAMPIONE
4. SCEGLERE IL CAMPIONE O I CAMPIONI SUI QUALI AVVERRA' LA RILEVAZIONE STATISTICA SECONDO UN METODO DI CAMPIONAMENTO;
5. FORMULARE, INFINE, UN GIUDIZIO SULLA "BONTA'" DEL CAMPIONE.

Metodi di campionamento



- Campione casuale semplice
- Campione sistematico
- Campione stratificato
- A due o più stadi
- Campione a grappoli



Campione casuale semplice

Il campione, per riprodurre in miniatura la popolazione, deve essere rappresentativo



Popolazione → Procedimento casuale → Campione rappresentativo

La rappresentatività è garantita da un campionamento casuale

Il campione **casuale** è estratto con procedimento tale che tutte le unità della popolazione hanno la **stessa probabilità** di essere estratte

ESTRAZIONE

Con ripetizione

**ESTRAZIONE
di tipo bernoulliano**

ESTRAZIONE

Senza ripetizione

**ESTRAZIONE
di un elemento
alla volta**

**ESTRAZIONE
in blocco**



- Quando la popolazione è teoricamente infinita e il campione è piccolo, CON O SENZA reintroduzione è influente.
(la probabilità che in un campione figurino più volte lo stesso elemento è del tutto trascurabile)
- Quando la popolazione è finita, la differenza fra i due metodi c'è.

ESTRAZIONE CON RIPETIZIONE (ESTRAZIONE BERNOULLIANA)



Per esempio, con una popolazione di 500 elementi, ogni elemento che viene estratto ha $P=1/500$ e si dice che i relativi risultati delle estrazioni **sono indipendenti**. Ciascuna delle unità potrà essere estratta più volte.

Il numero dei campioni ottenuti è pari a

$$D_{N,n}^r = N^n$$

Disposizioni con ripetizione

ESTRAZIONE SENZA RIPETIZIONE



Stesso esempio di una popolazione di 500 elementi:

$P_1=1/500$ (prob. che ha il 1° elemento di essere estratto)

$P_2=1/499$ (prob. che ha il 2° elemento di essere estratto)

$P_3=1/498$ (prob. che ha il 3° elemento di essere estratto)

e così via.....

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

ALLORA:

**SE L'ESTRAZIONE
E' EFFETTUATA
UN ELEMENTO ALLA VOLTA
(schema esaustivo)**



Il numero dei campioni ottenuti è pari a:

$$D_{N,n} = N(N-1)(N-2) \times (N-n+1) = \frac{N!}{(N-n)!}$$

Disposizioni semplici

SE L'ESTRAZIONE E' EFFETTUATA IN BLOCCO



Il numero dei campioni ottenuti è pari a:

$$C_{N,n} = \binom{N}{n} = \frac{N!}{n!(N-n)!}$$

Combinazioni semplici

Procedura per l'estrazione di un campione casuale



- Si numerano gli elementi (per rendere appunto equiprobabile - come lo sono i numeri - l'estrazione)
- Si estraggono a caso i numeri associati agli elementi.
- Non resta che scegliere se effettuare reimmissioni o meno.

TASSO DI CAMPIONAMENTO



Scelto il criterio con cui si vuole costruire lo spazio campionario, note le dimensioni N della popolazione ed n del campione

Definiamo

Tasso di campionamento

Il rapporto

$$\mathbf{n/N}$$

(si esprime in %)

Problema

Da una popolazione composta da 5 unità statistiche (A, B, C, D, E)
si voglia estrarre un campione casuale di numerosità 2.



Schema con ripetizione (bernoulliano)

- Ciascuna unità campionata viene riposta nella popolazione e può essere estratta nuovamente
- **Lo spazio campionario, ovvero l'insieme di tutti i possibili campioni è:**
AA AB AC AD AE
BA BB BC BD BE
CA CB CC CD CE
DA DB DC DD DE
EA EB EC ED EE
- Vi sono in tutto $5^2 = 25$ campioni, dati dalle disposizioni con ripetizione di 5 elementi a 2 a 2.

Schema senza ripetizione con estrazione di un elemento alla volta

- Ciascuna unità campionata viene tolta dalla popolazione e non può essere estratta nuovamente
- **Lo spazio campionario, ovvero l'insieme di tutti i possibili campioni è:**
~~AA~~ AB AC AD AE
BA ~~BB~~ BC BD BE
CA CB ~~CC~~ CD CE
DA DB DC ~~DD~~ DE
EA EB EC ~~ED~~ ~~EE~~
- Vi sono in tutto $= 20$ campioni, dati dalle disposizioni semplici

Stesso problema

Da una popolazione composta da 5 unità statistiche (A, B, C, D, E)
si voglia estrarre un campione casuale di numerosità 2.



Schema di estrazione in blocco

- Ciascuna unità campionata viene tolta dalla popolazione e non può essere estratta nuovamente
- Non sarà possibile distinguere l'elemento AB da BA e così via.....
- **Lo spazio campionario, ovvero l'insieme di tutti i possibili campioni è:**

~~AA~~ **AB** ~~AC~~ ~~AD~~ ~~AE~~
BA ~~BB~~ ~~BC~~ ~~BD~~ ~~BE~~
~~CA~~ ~~CB~~ ~~CC~~ ~~CD~~ ~~CE~~
~~DA~~ ~~DB~~ ~~DC~~ ~~DD~~ ~~DE~~
~~EA~~ ~~EB~~ ~~EC~~ ~~ED~~ ~~EE~~

$$C_{N,n} = \binom{N}{n} = \binom{5}{2} = \frac{5 \times 4}{2!} = 10$$

Vi sono in tutto = 10 campioni, dati dalle combinazioni **senza**
ripetizione di 5 elementi a 2 a 2.



NELLA PRATICA SI DIMOSTRANO DI
GRANDE UTILITA' PER LA FORMAZIONE
DEI **CAMPIONI DI TIPO BERNOULLIANO**

(ESTRAZIONE CON RIPETIZIONE)

LE TAVOLE DEI NUMERI CASUALI

Tavola dei numeri casuali



46072 97718 73189 59878 04588 69266 38434 40272 89771 00686
50496 39692 24085 71779 15687 66207 90331 35776 85413 15800
12600 63927 50359 37540 41359 11011 41613 65568 75066 27455
48616 57382 85689 27861 59135 38360 34711 60409 86253 71462
46642 76507 83543 17067 24598 99305 53754 26068 94233 65309
66476 03290 50341 38126 466**27 58122** 05830 88610 37392 49322
10911 58689 72799 85227 44187 07286 36816 96732 72213 56158
58369 19998 13711 36703 61578 71593 63712 44037 66954 50747
91136 89653 40289 30930 80642 32392 14276 79849 44117 93915
66983 44986 21176 13889 25913 95601 99590 38581 36491 29925
84659 22963 46657 93568 45189 15907 53529 32914 31407 74481
16370 15896 26540 63456 77955 19472 90187 18836 38374 50823
51962 18844 85961 66062 38212 55951 72211 39045 67240 30163

Se dobbiamo selezionare un campione di 100 unità da una popolazione di 1231 unità si sceglie a caso il primo numero di 4 cifre e si includono nel campione tutte le cifre $\leq$ 1231

2758 è il numero scelto a caso, ma non fa parte del campione.
Invece 1220 fa parte del campione

E' possibile utilizzare generatori di sequenze di nr. casuali

ESERCIZI PER CASA



SVOLGERE GLI ESERCIZI PROPOSTI
SUL SITO DEL PROF.



<http://www.profguida.it>

link: informazioni → news



Campione sistematico

Campionamento sistematico

Qualora si disponga di elenchi degli elementi di una popolazione da campionare per individuare le n unità del campione tra gli N della Popolazione si sceglie dagli elenchi una unità ogni K (N/n) elementi, a partire da una qualunque unità scelta a caso fra le prime K .



Si voglia estrarre un campione di 100 unità da una popolazione di 1000. Si sceglie un passo di $1000/10=100$. Supponiamo che il numero scelto a caso tra 1 e 100 sia 77. Le unità campionate sono le seguenti: 77; 177; 277; 377; 477; 577; 677; 777; 877; 977



Campione stratificato

Campionamento stratificato

La popolazione è divisa in strati internamente omogenei e quindi si procede con campionamento casuale da ciascuno strato.



Supponiamo di avere una popolazione di 7448 unità stratificate per sesso e classe di età

numerosità degli strati	classi di età			
	tra 14 e 34	tra 35 e 64	maggiore o uguale a 65	
Sesso				
Maschi	1879	1046	789	3714
Femmine	1756	976	1002	3734
	3635	2022	1791	7448



numerosità degli strati	classi di età			
Sesso	tra 14 e 34	tra 35 e 64	maggiore o uguale a 65	
Maschi	1879	1046	789	3714
Femmine	1756	976	1002	3734
	3635	2022	1791	7448
pesi degli strati	tra 14 e 34	tra 35 e 64	maggiore o uguale a 65	
Maschi	0.25	0.14	0.11	0.50
Femmine	0.24	0.13	0.13	0.50
	0.49	0.27	0.24	1.00

0,25 cifra approssimata

$0,25 = 1879 / 7448$
(frequenza relativa)



pesi degli strati	tra 14 e 34	tra 35 e 64	maggiore o uguale a 65	
Maschi	0.25	0.14	0.11	0.50
Femmine	0.24	0.13	0.13	0.50
	0.49	0.27	0.24	1.00

composizione di un campione stratificato di 500 unità	tra 14 e 34	tra 35 e 64	maggiore o uguale a 65	
Maschi	126	70	53	249
Femmine	118	66	67	251
	244	136	120	500

$$0,25:1=x:500$$

$$x = 0,25 \cdot 500 = 126$$

$$0,11:1=x:500$$

$$x = 0,11 \cdot 500 = 53$$

nei calcoli tenere conto delle cifre precedentemente approssimate

Campionamento a grappoli

La popolazione è divisa in grappoli (gruppi di unità statistiche elementari) si estrae un campione casuale di grappoli



Campionamento a più stadi

E' una tecnica di campionamento che risulta molto vantaggiosa quando la popolazione da studiare è molto numerosa e gli elementi possono essere raggruppati in diversi sottoinsiemi. Essa consiste in una prima fase caratterizzata da uno o più campionamenti a grappolo, seguita da una seconda fase in cui, all'interno dei "clusters" selezionati, si procede al campionamento delle unità seguendo una delle metodiche precedentemente enunciate (es. campionamento casuale semplice e campionamento stratificato)

Le indagini sulle famiglie dell'Istat seguono un piano di campionamento in cui i comuni (unità primarie) vengono stratificati e scelti a caso (1° stadio) e da ogni comune campionato viene scelto un campione di famiglie (2° stadio)



- Per approfondimenti fare riferimento al libro di testo e/o al file “Campionamento.pdf” reperibile sul sito del prof.:

[*http://www.profguida.it*](http://www.profguida.it)

link: servizi on line → materiale didattico

Buon lavoro!!!!