

LEZIONI DI STATISTICA MEDICA

Prof. Roberto de Marco

Lezione n.10

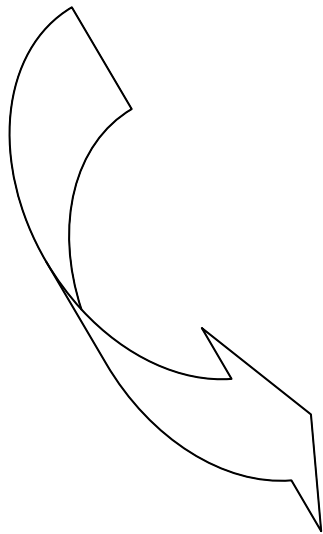
- L'inferenza statistica I



*Sezione di Epidemiologia & Statistica Medica
Università degli Studi di Verona*

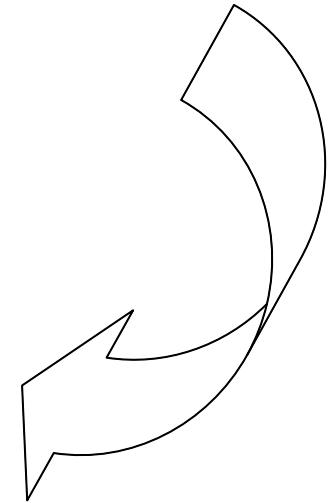
STATISTICA DESCRITTIVA

Metodi per la descrizione e
sintesi di un insieme di
osservazioni su un campione



METODI E MODELLI PROBABILISTICI

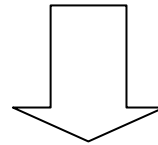
Modelli che permettono di
descrivere mediante pochi
parametri la distribuzione di una
variabile casuale nella popolazione



INFERENZA STATISTICA

INFERENZA STATISTICA

STUDIO DELLE RELAZIONI TRA CAMPIONE E POPOLAZIONE

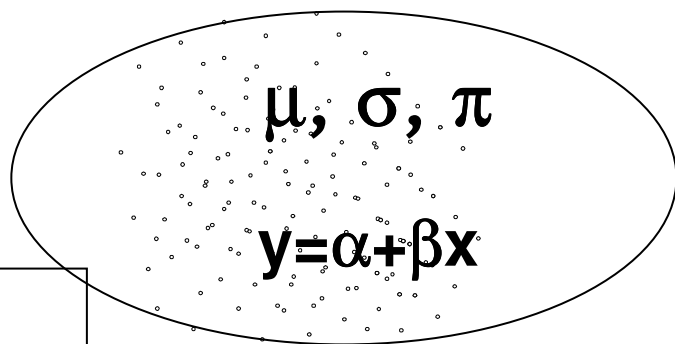


possibilità, sulla base dei risultati ottenuti su un campione, di fare delle affermazioni sulla popolazione

Nella ricerca il CAMPIONE (l'esperienza particolare che viene considerata in uno studio) è un mezzo per apprendere e/o approfondire una relazione o un fenomeno che si vuole generalizzare a una POPOLAZIONE

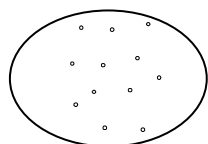
La popolazione il più delle volte è puramente astratta, non limitata nè nello spazio nè nel tempo (universo)

POPOLAZIONE o UNIVERSO



TEORIA DEL CAMPIONAMENTO
Quali soggetti selezionare?

STATISTICA INFERENZIALE
Cosa possiamo dire dei veri parametri della popolazione?
Qual è il margine d'incertezza?

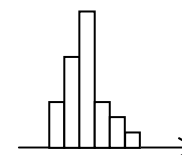


CAMPIONE

x_1, x_2, \dots, x_n

y_1, y_2, \dots, y_n

STATISTICA DESCRITTIVA



\bar{x}, s, p
STATISTICHE

$Y = a + bx$

CENNI di TEORIA del CAMPIONAMENTO

Molte ricerche vengono programmate con lo scopo di pervenire a **conclusioni generali**, valide per tutte le unità statistiche della popolazione, sfruttando i risultati ottenuti da un numero ridotto di osservazioni

La teoria del campionamento concerne le modalità di selezione del **CAMPIONE** dalla popolazione, al fine di rendere possibile la generalizzazione dei risultati.

UTILIZZO del CAMPIONE



VANTAGGI:

1. risparmio di lavoro e di costi dell'indagine perché vengono ridotte le unità di osservazione
2. la raccolta dell'informazione può essere più attendibile e più accurata
3. unica possibilità quando la popolazione su cui si vogliono fare inferenze è infinita.

SVANTAGGI:

1. imprecisione delle stime; le misure calcolate sono solo una approssimazione delle vere misure della popolazione e variano da campione a campione.

L'utilizzo del campione introduce delle fonti di errore nella stima dei parametri incogniti della popolazione:

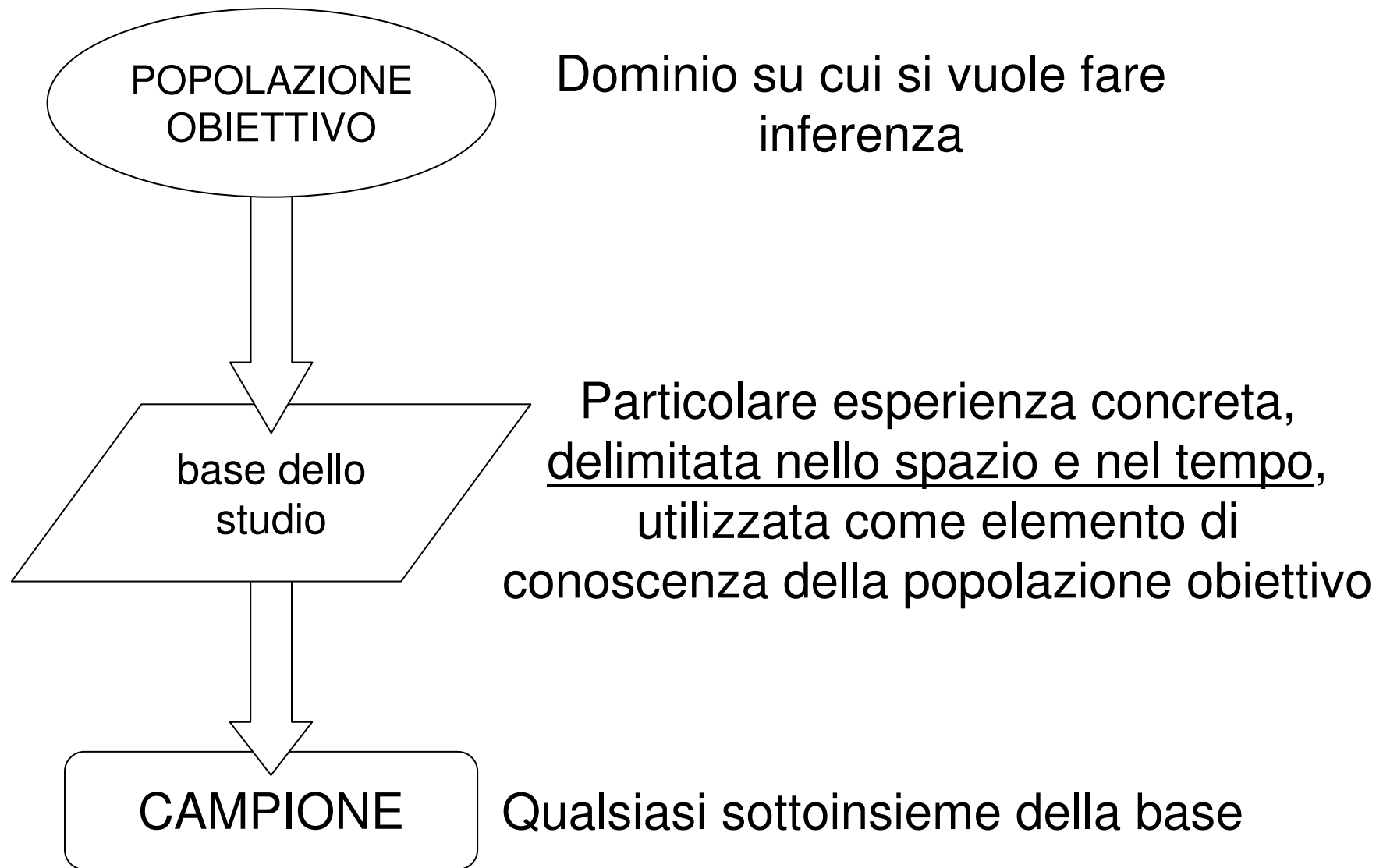
errori sistematici

vizi o bias legati alla **non rappresentatività** del campione prodotto dalla procedura di campionamento: le stime si allontanano in modo sistematico dal parametro della popolazione

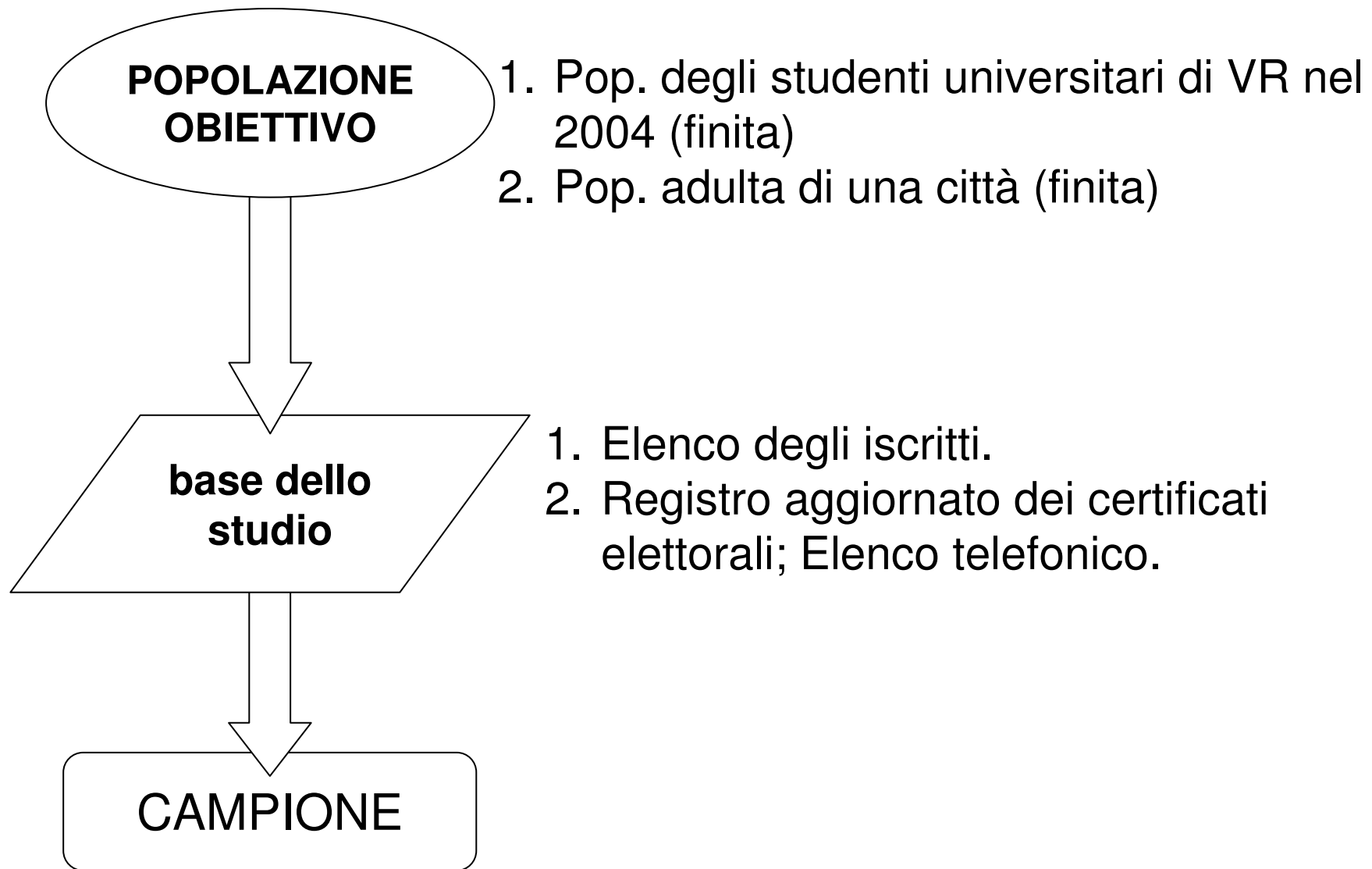
errori campionari

intrinseci alla procedura di campionamento; **influenzano la precisione della stima**. La dimensione dell'errore può essere predetta in base alla teoria della probabilità

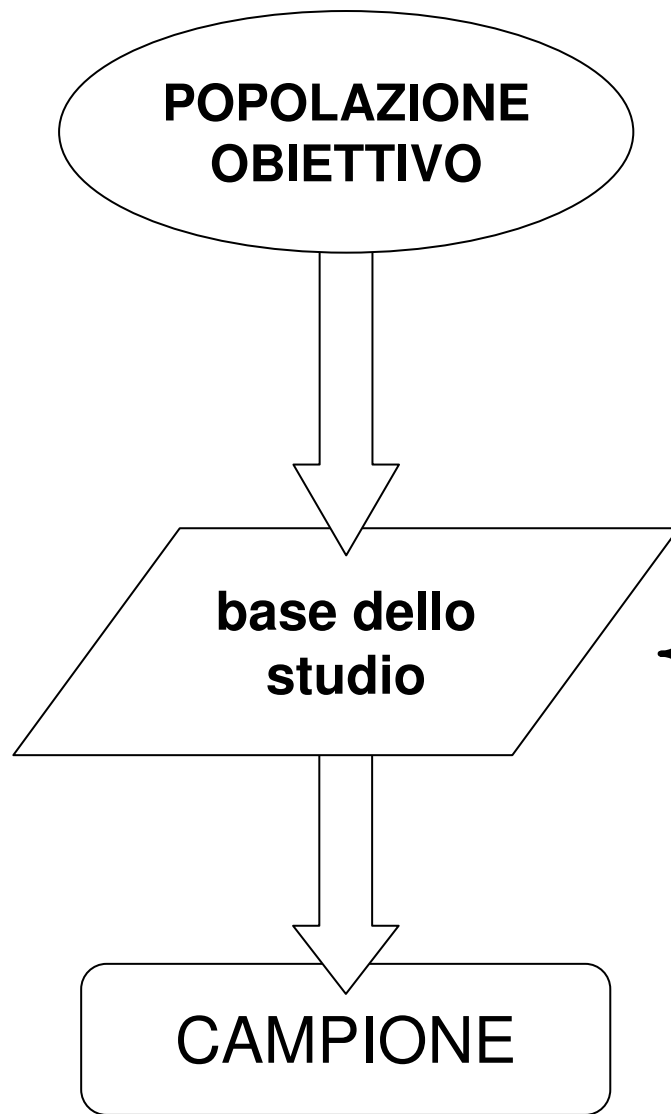
SCHEMA della PROCEDURA di CAMPIONAMENTO



esempi:



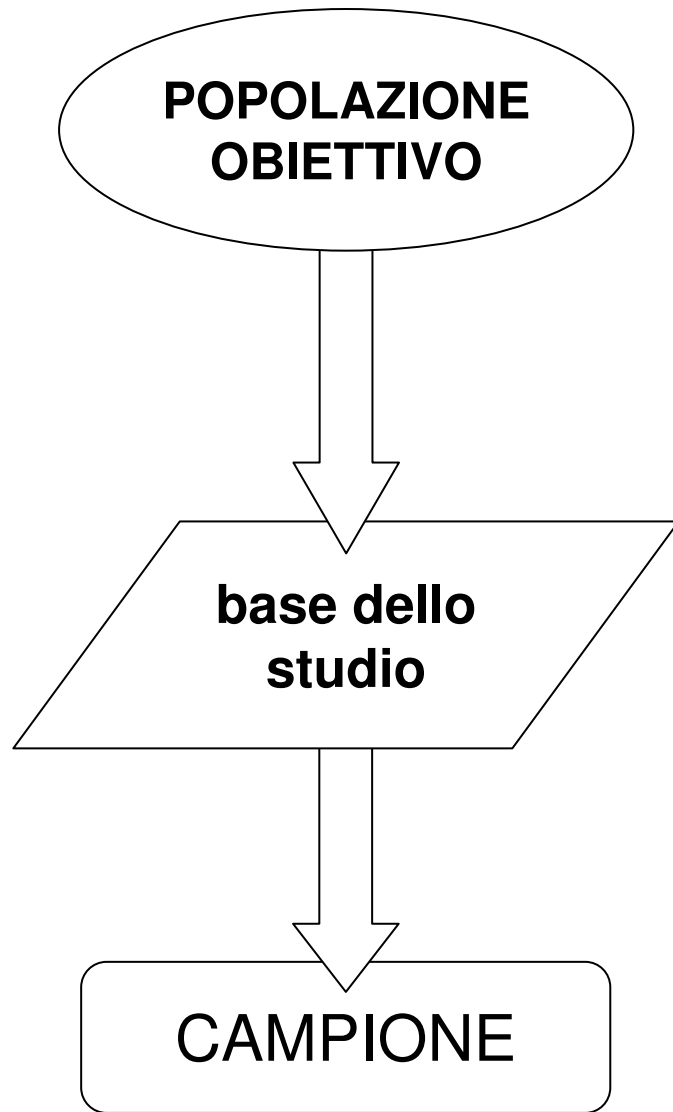
esempi:



Pop. degli ipertesi (infinita)

- Tutti i soggetti ipertesi che si rivolgono al loro medico a seguito di disturbi legati alla malattia in una data area e tempo;
- Tutti i soggetti di una data area che hanno valori di pressione arteriosa superiori a un certo cut-off;
- Tutti i soggetti ricoverati con diagnosi di ipertensione.

esempi:



Dominio della relaz. tra fumo e carcinoma polmonare

Opportuno insieme in cui sono rappresentate alcune categorie di fumo (alto, medio, basso, nullo) simili tra di loro eccetto che per l'abitudine al fumo

SCELTA della BASE dello STUDIO

La **base di uno studio** è scelta con criteri logici (e di efficienza) in funzione della sua idoneità rispetto alla popolazione obiettivo.

In genere:

1. nelle **indagini campionarie o di prevalenza**, mirate alla stima dei parametri della popolazione, deve essere garantita la rappresentatività della popolazione obiettivo.
2. nelle **indagini etiologiche**, mirate allo studio dei fattori responsabili dell'insorgenza di specifiche patologie, deve essere garantita la confrontabilità dei gruppi che hanno esposizioni (fattori di rischio) differenti.

indagini campionarie: esempio 1

In un'indagine mirata a valutare i consumi alimentari di una provincia del meridione le possibili basi potrebbero essere:

- elenco telefonico provinciale
- elenco dei certificati elettorali
- elenco dei certificati di residenza



indagini campionarie: esempio 2

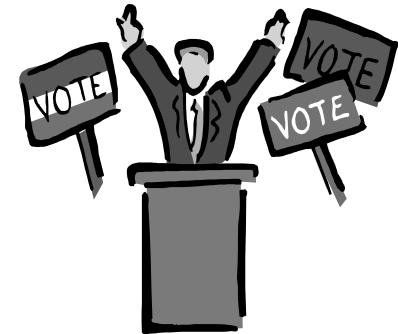
- In un'indagine mirata a valutare le complicazioni in bambini tra 3 e 5 anni affetti da morbillo :



- tutti i bambini ricoverati in ospedale con diagnosi di morbillo
- tutti i bambini iscritti alle scuole materne della zona
- elenco dei pediatri della zona

N.B.: Se la base non è rappresentativa della popolazione obiettivo le stime ottenute sono sistematicamente errate!

Anno 1936, U.S.A.: ELEZIONI PRESIDENZIALI



Candidati: **Roosevelt** e **Landon**

Literary Digest condusse un'indagine campionaria per predire i risultati delle elezioni.

Popolazione obiettivo: tutti i votanti degli Stati Uniti.

Base: liste riportate negli elenchi telefonici.

L'indagine predisse una vittoria globale di **Landon**

Roosevelt vinse con il più largo margine mai raggiunto in un'elezione presidenziale fino a quel tempo.



Nel 1936 vi era un gran numero di persone, per la maggior parte sostenitrici di Roosevelt, che non possedeva un telefono.

La base da cui è stato scelto il campione era “biased” (viziata). Il ceto sociale più elevato era sovrarappresentato nella base scelta.

N.B.: la grandezza del campione è relativamente priva di rilevanza nel compensare gli effetti dei vizi di selezione avvenuti nel campionamento **(il campione era costituito da oltre un milione di risposte!)**

SCELTA DEL CAMPIONE

Il campione è un qualsiasi sottoinsieme della base.

La scelta del campione deve essere pertinente agli obiettivi della ricerca e, per permettere la generalizzazione dei risultati, dovrebbe **rappresentare la base** e non essere frutto di una particolare **selezione** cosciente o incosciente (equazione personale d'errore)

Es.: scelta di un campione di studenti per la misura del Q.I.



Tipi di campioni

Non-Probability Sample (campione non casuale o non

probabilistico): “ Un campione selezionato in modo tale da *non garantire* a ogni membro della popolazione *una probabilità nota* di essere selezionato.”

Probability Sample (campione casuale o probabilistico): “ Un

campione selezionato in modo tale *da garantire* a ogni membro della popolazione una *probabilità nota e diversa da 0* di essere selezionato.”

Campioni non probabilistici

Campione di convenienza: scelto per la facilità con cui i soggetti possono essere selezionati (es: studenti di una classe, residenti in una data via).

Campione per scelta ragionata: consiste di persone che si ritiene importante studiare poiché possiedono specifiche caratteristiche e non esiste una base di campionamento opportuna (es: pazienti affetti da una data patologia o severità,...vagabondi...)

Campione per quote la popolazione viene preliminarmente suddivisa in quote in base ad alcune caratteristiche (es. 100 maschi e 100 femmine), lasciando agli intervistatori la possibilità di scegliere le persone da intervistare (nel rispetto delle quote assegnate).

Campioni non probabilistici- Quando usarli?

- * Quando è già noto un gruppo che si sa rappresentare la popolazione (es. alcune circoscrizioni elettorali).
- * Quando è impossibile disporre di una base di campionamento (es. homeless, malattie rare.)
- * Nelle nuove ricerche in cui i casi di interesse non sono noti o mai stati studiati prima (es. Malattie nuove).

N.B I risultati ottenuti da campioni non casuali non permettono generalizzazioni con un definito e noto grado di accuratezza. In altre parole, i dati rappresentano solo le unità che sono state studiate.

Campioni probabilistici o casuali

I campioni probabilistici garantiscono a ciascuna unità della popolazione (base) una definita e positiva probabilità di far parte del campione.

La scelta dei soggetti è affidata al caso (randomizzazione).

Sono basati sulla teoria della probabilità e permettono la possibilità di applicare le tecniche dell'inferenza statistica ai dati ottenuti.

Permettono la generalizzazione dei risultati con un dato e noto margine d'errore

Tipi di campioni probabilistici

Campione casuale semplice(ccs) Consiste nell'estrazione casuale di un certo numero di elementi dall'insieme della base (ogni membro ha la stessa probabilità di essere selezionato)

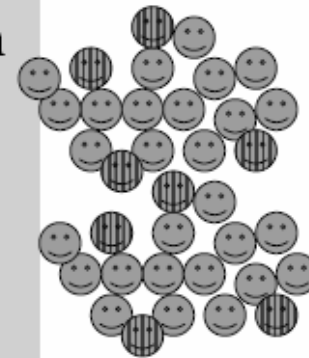
Campione stratificato La popolazione o la base di studio viene precedentemente divisa in sottogruppi (strati) da ognuno dei quali viene estratto un ccs.

Campione a cluster: Aggregati di unità (cluster) vengono selezionati mediante ccs da una popolazione. Vengono poi selezionati tutti gli elementi di ogni cluster.

Campionamento a più stadi: Vengono effettuati ccs che coinvolgono più stadi. Per esempio, nel I° stadio vengono randomizzati i cluster e nel II° gli individui.

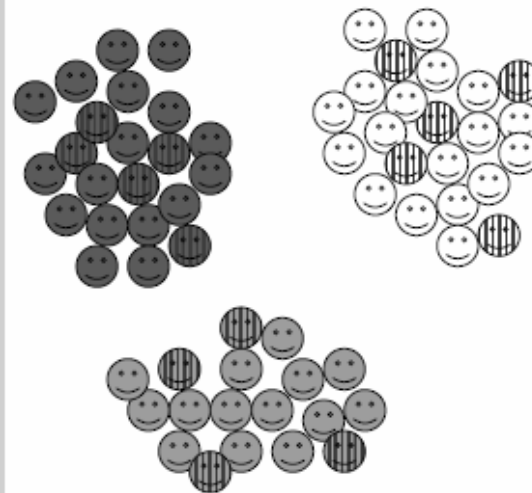
Simple Random Sampling

- Each element in the population has an equal probability of selection AND each combination of elements has an equal probability of selection
- Names drawn out of a hat
- Random numbers to select elements from an ordered list

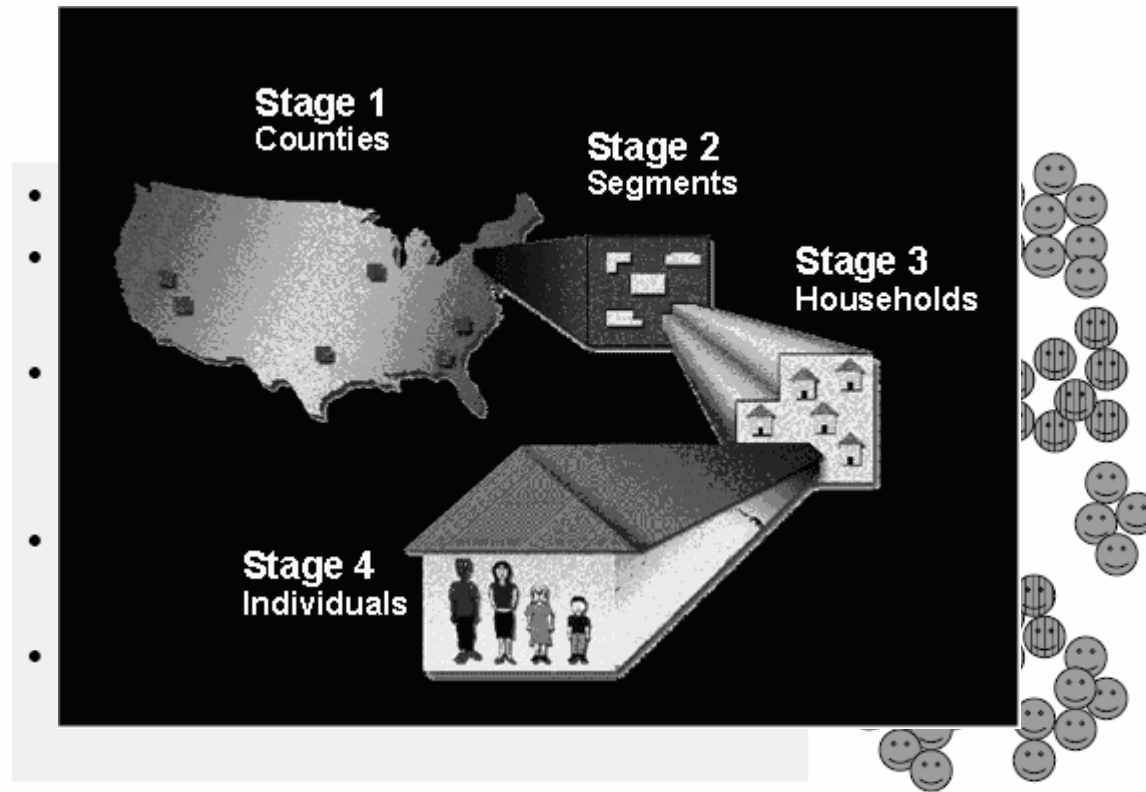


Stratified Random Sampling-1

- Divide population into groups that differ in important ways
- Basis for grouping must be known before sampling
- Select random sample from within each group



Multistage sampling in the NHANHES STUDY



GENERAZIONE di un CAMPIONE CASUALE SEMPLICE

(mediante utilizzo delle tavole dei numeri casuali)

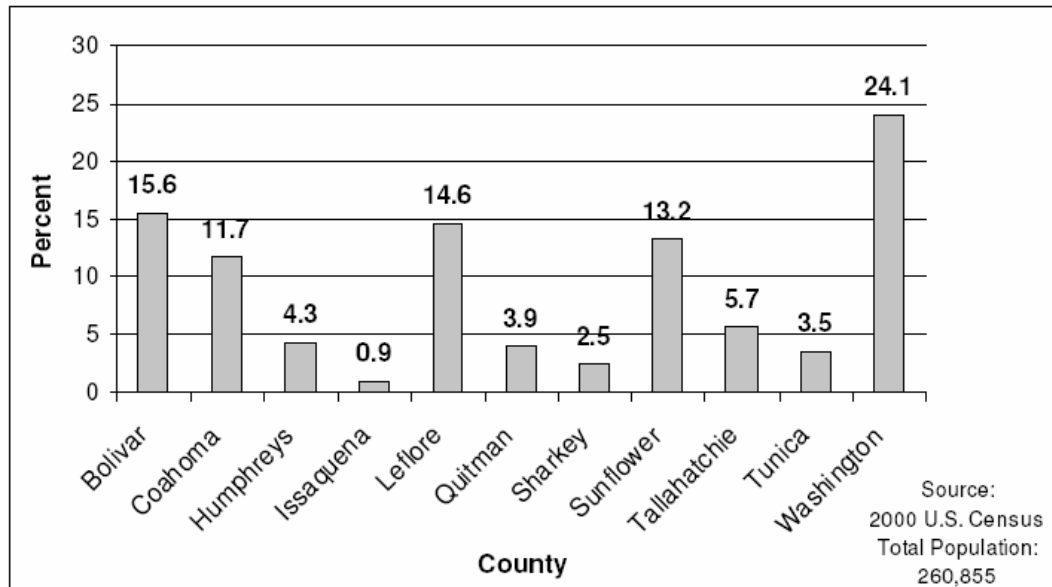
Sia definita una base composta da N individui, e sia n la dimensione prescelta del campione:

1. ad ogni individuo della base, da 1 a N , è assegnato un numero progressivo.
2. si leggono le tavole a gruppi di tante cifre quante sone le cifre di N .
3. i soggetti corrispondenti ai primi n numeri letti (non ripetuti e tali per cui $n < N$) saranno selezionati a far parte del campione.

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	46627	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
55289	18202	49671	58143	44912	68204	31391	52319	10633	12146
48023	09540	85628	06281	57154	59317	21691	75261	90525	52591
14798	38093	82618	59617	60396	47863	20462	41026	93710	34491
93722	00368	76492	09912	49084	36237	53900	95067	60697	70823
92215	91249	67934	96279	14949	51163	64807	75102	55473	41117
84980	02710	86329	14559	60912	25090	88126	38326	14818	73918
15786	94713	47704	50971	83736	22432	45223	26815	88408	03901
14510	54630	46853	80698	21089	06777	48303	46988	73035	78366
34300	02413	05981	60951	61023	86349	81066	78329	85623	00742
76204	75471	82181	35454	93086	28888	46459	92850	75617	94005
15161	14287	07003	47560	82243	87339	50047	72822	49783	88463
55376	47645	65456	81222	16625	60816	50228	06328	90450	95496
85252	82677	47132	19295	49170	70543	02715	94290	66588	78247
59173	65864	11685	29750	96562	44030	95630	70747	03598	57338
75755	50084	24656	35590	50263	36487	47167	11955	83545	27189
66704	35388	18479	14849	50527	39960	07032	28009	08718	88005
06825	50142	34241	39373	14363	74470	74684	68779	24274	01953
15205	30117	78412	25253	76498	49688	24256	29259	30967	84152
26513	16234	90329	39268	52305	13938	84976	83552	09715	96443
31216	78892	86748	22159	68678	30969	73137	19875	90005	13434
12738	57513	29679	07678	69125	47869	27664	31192	96509	45292
30264	28122	12021	08704	03626	48134	44675	72004	94533	15530
29137	62351	71997	43811	97560	91770	71973	96267	34254	51623
28166	65149	99684	63318	13423	30061	16421	02000	38288	39696
51309	25148	81989	26320	70166	80126	19348	50146	18261	33036
23212	17788	15066	28825	82313	95166	17890	51844	42524	85515
18162	81375	11775	93908	27783	51448	31913	42555	48292	36972

Tavola dei numeri casuali

Comparing a random sample to the population (Delta Rural Poll).



← Total population distribution, by county.

Figure 8b: Percent of Total Population by County

Total sample distribution, by county. →

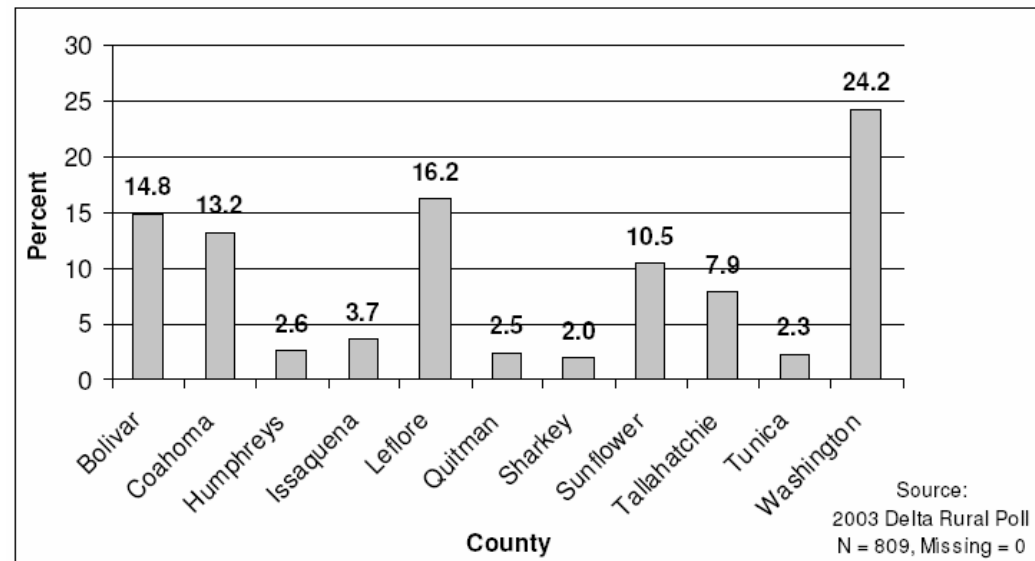


Figure 8a: Percent of Respondents by County